

PERSPECTIVE

Data-Centric Engineering in modern science from the perspective of a statistician, an engineer, and a software developer

Christophe Ley^{1*} , Mike Tibolt² and Dirk Fromme³

¹Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Gent, Belgium

²ArcelorMittal, ArcelorMittal Global R&D, Esch-sur-Alzette, Luxembourg

³3plusplus GmbH, Sonsbeck, Germany

*Corresponding author. E-mail: christophe.ley@ugent.be

All authors contributed equally to this study.

(Received 30 January 2020; revised 22 February 2020; accepted 24 February 2020)

Keywords: Big Data; data science; data security; education; structural engineering

Abstract

Data-Centric Engineering is an emerging branch of science that certainly will take on a leading role in data-driven research. We live in the Big Data era with huge amounts of available data and unseen computing power, and therefore a crafty combination of Statistics (or, in more modern terms, Data Science), Computer Science and Engineering is required to filter out the most important information, master the ever more difficult challenges of a changing world and open new paths. In this paper, we will highlight some of these aspects from a combined perspective of a statistician, an engineer and a software developer. In particular, we will focus on sound data handling and analysis, computational science in Structural Engineering, data care, security and monitoring, and conclude with an outlook on future developments.

Impact Statement

This position paper discusses the importance of the recent yet important branch called Data-Centric Engineering. It is meant to be accessible for a wide readership and to shed light on this perhaps still abstract expression. Especially the fact that the paper is jointly written by a statistician, an engineer, and a software developer guarantees a multi-faceted view. We will explain the ubiquity of data in our everyday lives, discuss the need for engineering to be re-thought and re-taught from a data-centric perspective, and warn against potential misuses of data. In particular, we will give recommendations regarding data security and monitoring. It is essential that we come to realize the chances that the current Big Data era offers us, but it is at least as important that we are aware of the accompanying challenges and dangers. Data alone, without a clever analysis, interpretation, and usage, are not worth much!

1. Introduction

In 2012, IBM stated that 90% of the data available today have been generated over the past 2 years (Sagiroglu and Sinanc, 2013). Technological advances combined with the creation of new powerful devices able to collect, store, and transfer huge amounts of data have re-shaped the landscape of our every-

day lives and, in particular, of science. We live in the Big Data era, and data have replaced oil as the most valuable resource on earth (The Economist, 2017). It is therefore no surprise that statisticians, data analysts, or data scientists are regularly cited among the best jobs in recent years, and that Data Science (a modernized version of Statistics, in some sense, see the next section) has emerged as a modern pillar of scientific research or, quoting Jim Gray, as the “fourth paradigm of science” (besides empirical, theoretical, and computational science). Data-driven research has become a cornerstone of several domains such as structural engineering, medicine, health science, biology, genetics, finance, or environmental sciences, to cite but these. In recent years, the interest in data-driven research has been particularly growing in the vast field of engineering (Chiaia et al., 2019; Madi and Radovanovi, 2012; Marrongelli et al., 2019), especially in combination with the advent of computational sciences. This has led to the recently created new field of Data-Centric Engineering.

One of the ultimate goals being the creation of digital twins, that is, a digital replica sharing the same physical properties as a real system, Data-Centric Engineering is, by nature, inter-disciplinary. This poses a big challenge: one needs to combine knowledge from Engineering and Data Science in order to enter this domain, or alternatively seek for fruitful collaborations between researchers from each field, who by nature speak distinct scientific languages. Therefore, in order to discuss in the best possible way about the opportunities, dangers, and challenges of Data-Centric Engineering, the authors of this paper have three very distinct areas of expertise, hereby reflecting the inter-disciplinary character of the domain: the first author is a statistician, the second author an engineer, and the third author a software developer.

The paper is organized as follows. In section “Data Science in a Nutshell,” we briefly touch upon Data Science in order to be fully equipped to delve into a real-life example of Data-Centric Engineering in section “Data Science in Structural Engineering: An Interdisciplinary Approach.” Section “Data Science in Structural Engineering: Data Care and Safety” deals with the important aspect of data care and security. We complete this paper in section “Conclusion” with a conclusion about the main chances and dangers of this new domain, and with our main message, namely that, in order for the field to stand a chance to flourish, a new type of targeted education is indispensable.

2. Data Science in a Nutshell

2.1. What does Data Science actually mean?

Data Science is a multi-disciplinary field whose aim is to extract knowledge from data. In some sense, it can be perceived as a modernized version of Statistics which follows the same aim. Quoting Nate Silver at the Joint Statistical Meetings of 2013, “Data Scientist is just a sexed up word for Statistician.” According to Diggle (2015), Statistics is the Data Science of our modern times. In the same vein, Ley and Bordas (2018) claim that Data Science is actually Statistics 2.0. What may set Data Science apart from the more “classical” Statistics (probability model building, data description, inference, and prediction) is the incorporation of machine learning (supervised, unsupervised, and reinforcement learning) and computer science besides these classical statistics. It thus consists of both top-down (a model is built that represents the information contained in the data) and bottom-up (starting point are the data and the model is then built by a computer) approaches.¹ It is this crafty combination that makes Data Science so attractive for theoreticians and practitioners, and has caused the hype around this “recent” field. It is definitely the right discipline at the right moment bearing the right name, and therefore rightfully at the core of Data-Centric Engineering.

2.2. (Big) Data alone are not worth much!

One of the biggest misconceptions related to the Big Data phenomenon is that the sole fact of having massive data sets will solve all problems.² Nothing is less true, because data do not speak by themselves (though this sentence is often incorrectly used). Various steps are required to obtain the best out of the data

¹ We refer the reader to Ley and Bordas (2018) for a thorough discussion on Data Science and for several relevant references.

² The first author was once asked at a conference if Big Data would also replace computer simulations.

amounts: a smart data collection strategy (smart devices and wearables have opened new doors here), a thorough data cleaning (checking for outliers, handling missing data, etc.), a flawless data analysis (via statistical or machine learning methods), and ultimately a correct interpretation of the results. Besides the development of new methods to analyse data, the data analysis as such is also part of Data Science.

3. Data Science in Structural Engineering: An Interdisciplinary Approach

3.1. *The assets of the application of Data Science in Structural Engineering*

Over a long time, Data Science and Structural Engineering were considered as two completely separated disciplines and developed independently. In the coming years, a clear change of mentality is expected to take place in Structural Engineering. The present section summarizes the benefits and challenges of the assessment of data in Structural Engineering, more specifically for the monitoring of bridge structures. A case-study about digital inspection of an Italian bridge will highlight the advantages of Data-Centric Engineering for safety aspects of bridges in general.

Regular inspections of structures, for instance of bridges, are recognized as an appropriate tool to ensure their safety throughout their lifetime. Inspections of sensitive structural details are carried out by engineers and technicians who access the structures by cranes and/or enter the structures by special entrances to control the details from inside. Inspections consist in visual inspection and measurements with specific techniques like X-rays and ultrasonic to detect cracks inside the parent steel material and welds. Depending on the observations, engineers decide for specific maintenance actions on the structure. Maintenance actions are for example the repair of cracked welds in steel bridges, the replacement of cracked concrete elements and corroded reinforcements. The regular inspections and maintenance costs usually represent the dominant part of the user costs of the structures. An economical structural solution stands out by low inspection and maintenance costs. Many details are in areas of the structure which are difficult to access and consequently the inspection is expensive, time consuming, and error prone. Therefore, the trend in civil engineering will develop toward digital inspections of structures, whereupon bridges are predestinated for digital inspection.

For digital inspection, sensors are applied on sensitive details of the bridge and continuously measure their loading (e.g., stresses, strains, etc.). The data of the loading are stored on computers and regularly assessed by civil engineers. The engineers determine the damage of the details and identify the required maintenance actions. The latest development in computational engineering allows for the application of sensors which continuously measure the loading of the details during the whole lifetime of the bridge. The lifetime of a bridge is defined with 100 years and one can easily imagine the amount of data which is collected over this timespan. The advantage of the method consists in the remote inspection of the bridge. The engineers no longer need to physically access the details of the bridge, which they can inspect safely from the office. This reduces the inspection costs and the measurements can be more precise.

3.2. *The challenge of an interdisciplinary approach*

The challenge consists in the correct and effective assessment of the Big Data to conclude on the safety of the bridge and to launch maintenance actions only if required. Since civil engineers are not experts in analyzing such data, the method is often ineffective. A general strategy to assess the measured data does currently not exist in civil engineering and the engineers apply a case-by-case approach. This leads to a situation where not all the information is extracted from the data. The risk is high that the damage of the bridge is wrongly calculated from the data (too high or too low). As a result, the safety of the bridge is wrongly assessed, or needless and uneconomic maintenance actions are undertaken. Another challenge is the hardware to store the data. Many small engineering offices are currently inadequately equipped to evaluate the data and to apply for digital inspections. A standard gear for the evaluation of data for small and medium design offices needs to be defined.

3.3. *Case study: the Marchetti bridge*

Let us now become more concrete and illustrate our statement via a case-study for digital inspection, the so-called Active Monitoring. This is a perfect example of Data-Centric Engineering as an interdisciplinary



Figure 1. Main beam of the Marchetti bridge (Chiaia et al., 2019) Springer Nature Switzerland AG 2020. This figure has been reproduced with the permission of the copyright holder, and is not included in the Creative Commons licence applied to this article. For other re-use, please contact the copyright holder.

approach between Data Science and Structural Engineering and is considered to be a driver for the future developments of digitalization in engineering.

The Marchetti bridge is a steel arch high way bridge between Torino and Milano, Italy. The structural layout is a single arch connected to the inferior way with a span of 250 m. The deck consists of a chain beam and transverse cantilevers with a spacing of 8 m. The bridge carries two carriageways of a total width of 17.55 m per way (Chiaia et al., 2019). Figures 1 and 2 show the superstructure of the bridge.

Active Monitoring has been developed by G. Ventura for ARCOS Engineering. It is a new paradigm in bridge monitoring. It can be considered as a novel approach to Structural Health Monitoring (SHM). Several examples of SHM can be found in the literature, see Chiaia et al. (2019), Marrongelli et al. (2019), as well as Cunha et al. (2013), Koo et al. (2013), and Magalhães et al. (2009). Active Monitoring consists in a real-time assessment of the measured data through a computational model. Sensors are placed on specific details and locations of the bridge. They measure loads, displacements, and other important parameters. The acquired data are automatically elaborated by an algorithm and sent to a



Figure 2. Arch of the Marchetti bridge (Chiaia et al., 2019) Springer Nature Switzerland AG 2020. This figure has been reproduced with the permission of the copyright holder, and is not included in the Creative Commons licence applied to this article. For other re-use, please contact the copyright holder.

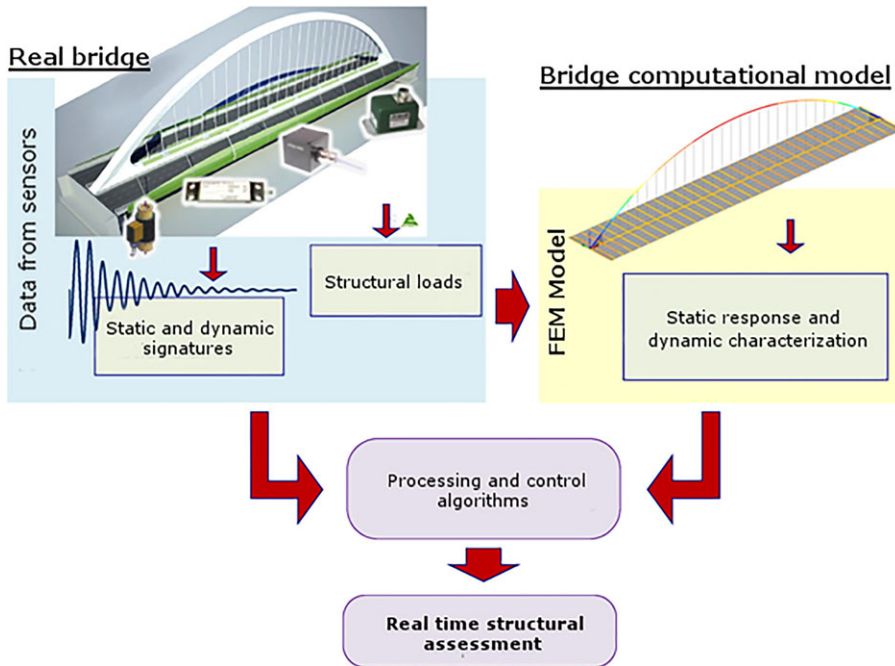


Figure 3. Methodology of Active Monitoring (Chiaia et al., 2019). Springer Nature Switzerland AG 2020. This figure has been reproduced with the permission of the copyright holder, and is not included in the Creative Commons licence applied to this article. For other re-use, please contact the copyright holder.

finite-element model of the bridge for the real-time evaluation of its health status. A certain deviation between the measured values and the values from the numerical model is an indication of a safety degradation of the bridge. Figure 3 summarizes the philosophy of Active Monitoring.

For the Marchetti bridge, different sensors are applied on 12 monitoring points. The range of sensors covers high resolution servoinclinometers, temperature sensors, humidity sensors, triaxial accelerometers, wind pressure transducer, strain gages, and load cells. The locations are depicted in Figure 4.

For the processing of the data, a custom software has been developed by ARCOS Engineering. The software acquires the data, calls the finite element engine and processes back the data, automatically compares the measured and calculated data and generates warning if the difference exceeds a threshold

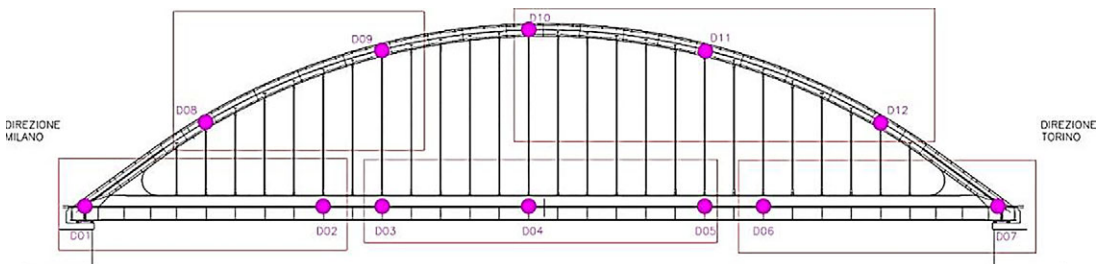


Figure 4. Position of the sensors on the Marchetti bridge (Chiaia et al., 2019). Springer Nature Switzerland AG 2020. This figure has been reproduced with the permission of the copyright holder, and is not included in the Creative Commons licence applied to this article. For other re-use, please contact the copyright holder.

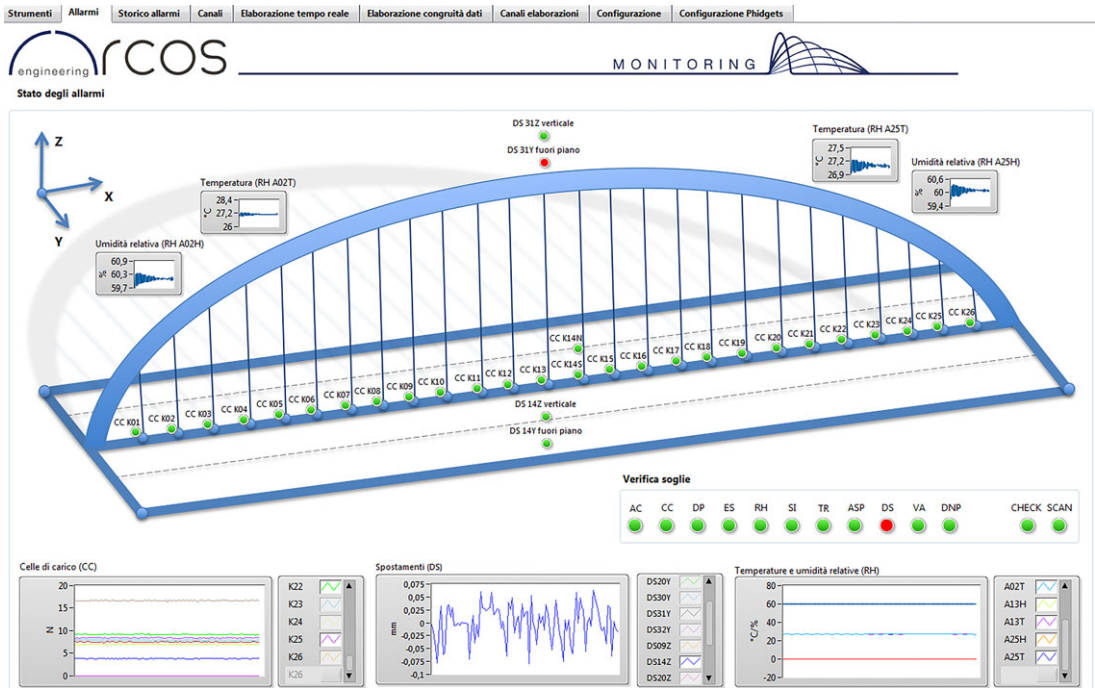


Figure 5. Semaphore in the Active Monitoring System (Chiaia et al., 2019). Springer Nature Switzerland AG 2020. This figure has been reproduced with the permission of the copyright holder, and is not included in the Creative Commons licence applied to this article. For other re-use, please contact the copyright holder.

value. The finite element software runs on a high-end Microsoft windows workstation. Figure 5 shows an example of the output of the numerical model and the comparison to the measured data from the sensors installed on the bridge.

Further information on Active Monitoring, which was first applied on the Marchetti bridge in 2015, can be found in Chiaia et al. (2019).

The example of the Marchetti bridge in Italy highlights the benefits of Data-Centric Engineering. The automatic assessment and comparison of the data eliminates the “human factor” and allows for a quick, economical and reliable monitoring of bridges. It should be noted that Active Monitoring can still not replace the analytical capacity of an expert in structural data processing and interpretation (Chiaia et al., 2019). Moreover, the data to be measured on the bridge should be clearly defined and limited to the needs of the safety assessment. Too many parasitic data would exceed the capacity of the data requirement system and would falsify the results of the finite element software. Finally, the data itself is not worth anything without a proper processing and evaluation of it. The data for themselves are not important in Structural Engineering but rather the information one can gain from them; see section “(Big) Data alone are not worth much!”

4. Data Science in Structural Engineering: Data Care and Safety

As we have seen, Data-Centric Engineering may result in many benefits. On the other side, one has to be aware that there are also some risks involved. Taking a structure like the Marchetti bridge as an example, it is easy to understand that such a work is an expensive and critical element of the infrastructure. As such, it has to be protected. Common steps to do so are for instance regular manual surveillance and to install road

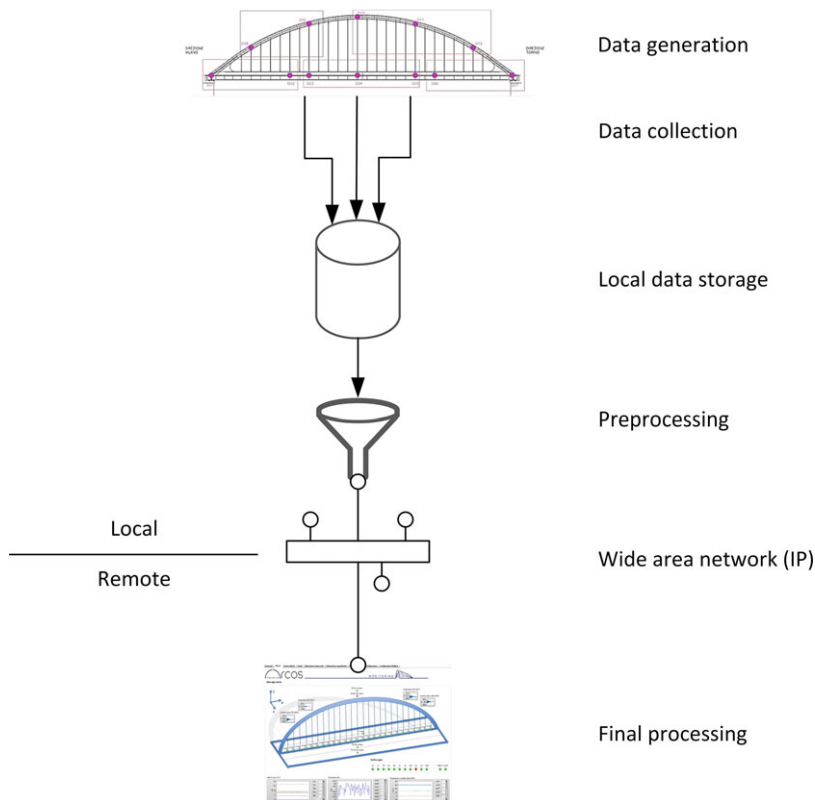


Figure 6. Data flow. 3plusplus GmbH, 2020. Incorporates a figure from Chiaia et al. (2019) (Figure 4). Springer Nature Switzerland AG 2020. This figure has been reproduced with the permission of the copyright holder; and is not included in the Creative Commons licence applied to this article. For other re-use, please contact the copyright holder.

signs to limit the speed allowed of cars crossing the bridge or to limit their weight. However, collecting and transmitting data as described before rises new dangers which can affect the security of the bridge. The data itself and any processing of it has to be protected against misbehavior or even manipulation. This begins right at the sensor's site and ends with the transmission of the data via any type of channel.

The graph in Figure 6 is designed to give an impression of the data flow and its hardware and software involved.

Figure 7 shows some hardware design and implementation concepts, used to fulfil required safety standards.

Each item, every software and all processes have to be included in the conceptual design of the structural health monitoring. Therefore, the development of safety and reliability standards is a premise for this type of implementations. Such standards should also be developed for the detection of software malfunction and its reporting, including all actions taken to overcome the issues. Hopefully, the widespread use of electronics and software in our daily life will fuel appropriate research, development, and implementation.

There is already a long history of developing concepts for the assurance of software reliability (Myers 1976). Known methods and proceedings include Fault Tree Analysis, Failure Mode and Effects Analysis, Markov Chains, and Redundancies (Lehrstuhl SRS—Sicherheit und Zuverlässigkeit, 2019). Nevertheless, the increasing complexity of modern open and distributed systems requires even more sophisticated approaches to ensure reliable, safe, and durable operation of such systems.

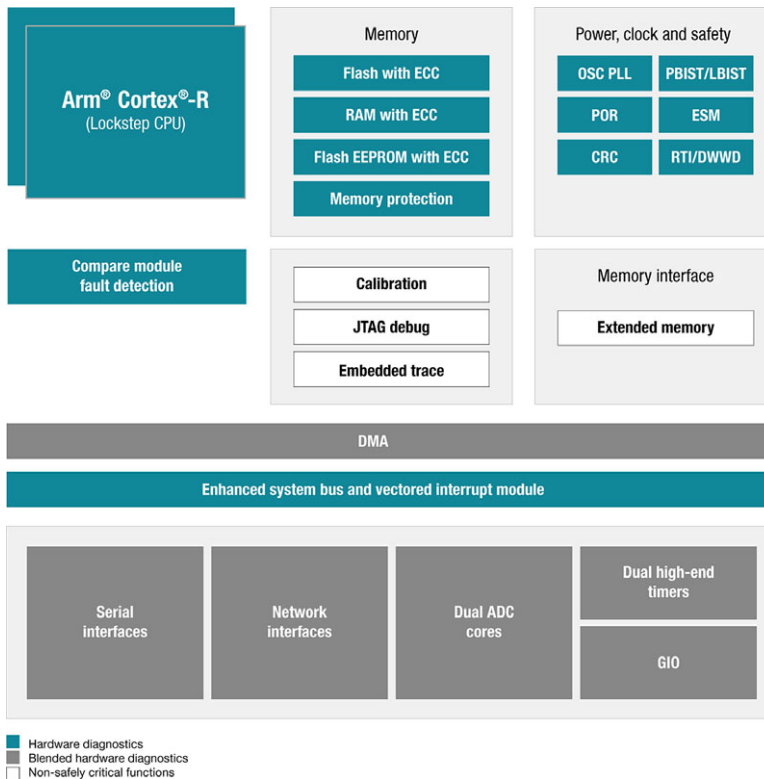


Figure 7. On-chip diagnostics for functional safety.³ Courtesy of Texas Instruments—Microcontrollers (MCU). This figure has been reproduced with the permission of the copyright holder, and is not included in the Creative Commons licence applied to this article. For other re-use, please contact the copyright holder.

This is presently an essential part of the related research (Sicherheit und Zuverlässigkeit 2019). Consequently, the task for engineers these days is to adopt the latest concepts of software or even system safety. A close cooperation of research and engineering is required as well as an interdisciplinary attitude for all parties involved. Additionally, the implementers' work has to include ways to dynamically test and improve the safety and intended usage of the systems and to protect them against failures and unintended usage.

For software developers, the classical first paradigm has changed. Programming does not start any longer with:³

```
int main()
{
    printf("Hello, World!\n");
    return 0;
}
```

This was true for 70 years and implemented in at least 50 programming languages.⁴

³ Courtesy of Texas Instruments—Microcontrollers (MCU). Available at <http://www.ti.com/microcontrollers/hercules-safety-mcus/overview.html#simplify-certification> (accessed 9 December 2019).

Now, it should read:

```
int main()
{
    printf("Hello, Safety!\n");
    return 0;
}
```

One can say, there is no security without safety.

5. Conclusion

We hope to have conveyed through this position paper our view on the chances and challenges of the exciting new field which is Data-Centric Engineering. Our final take-home message is threefold:

1. *Data-centric thinking has become a necessity* in various domains, and in particular of course in Data-Centric Engineering. The wealth of available data has created this necessity, which will drive forward the research in engineering in a probably not yet seen speed.
2. *Pay attention with the data!* Data alone are not worth much and may convey wrong messages. It is of utmost importance to analyse and use the large amount of data in the best possible way, and to keep an eye on data security.

The third message is a natural consequence of the former two:

3. *Data-based education will become an essential skill in engineering studies.* Advanced data analysis methods are absent from nearly all classical engineering study programs, which is the main reason for the “researchers speaking different languages” phenomenon mentioned in the Introduction. Data-Centric Engineering is highly inter-disciplinary, and scholars therefore should begin to train the new generation of engineers and scientists in a mind-open and domain-overlapping way. A combined knowledge of Engineering, Computational Science and Data Science will provide a powerful skillset to master the future challenges in a fast-changing, data-oriented world, be it at academic, commercial or entrepreneurial level. Only then will Data-Centric Engineering come to entire fulfilment.

Acknowledgments. The authors and Cambridge University Press thank Springer Nature for permission to reproduce Figures 1–5 (Chiaia et al., 2019). These figures are not included in the Creative Commons licence applied to this article. For other re-use, please contact the copyright holder.

Funding Statement. This work received no specific grant from any funding agency, commercial or not-for-profit sectors.

Competing Interests. The authors declare no competing interests exist.

Authorship Contributions. Conceptualization: C.L., M.T., and D.F.; Methodology: C.L., M.T., and D.F.; Writing original draft: C.L. (sections “Introduction,” “Data Science in a Nutshell,” and “Conclusion”), M.T. (section “Data Science in Structural Engineering: An Interdisciplinary Approach”), and D.F. (section “Data Science in Structural Engineering: Data Care and Safety”). All authors approved the final submitted draft.

Data Availability Statement. Data availability is not applicable to this article as no new data were created or analyzed in this study.

⁴70 Years Of “Hello, World!” With 50 Programming Languages. Available at <https://medium.com/javarevisited/70-years-of-hello-world-with-50-programming-languages-2400de893a97> (accessed 9 December 2019).

References

- Chiaia B, Ventura G, Zannini Quirini C and Marasco G** (2019) The Marchetti Bridge: active control for maintenance and structural safety. In *9th International Conference on Arch Bridges*. University of Porto, Porto, Portugal: Springer, pp. 866–873.
- Cunha A, Caetano E, Magalhães F and Moutinho C** (2013) Recent perspectives in dynamic testing and monitoring of bridges. *Structural Control Health Monitoring*, 20, 853–877.
- Diggle PJ** (2015) Statistics: a data science for the 21st century. *Journal of the Royal Statistical Society Series A*, 178, 793–813.
- Koo KY, Brownjohn JMW, List DI and Cole R** (2013) Structural health monitoring of the Tamar suspension bridge. *Structural Control Health Monitoring*, 20, 609–625.
- Lehrstuhl SRS—Sicherheit und Zuverlässigkeit** (2019) Available at https://www.unidue.de/srs/Forschung3_de.php (accessed 9 December 2019).
- Ley C and Bordas SPA** (2018) What makes data science different? A discussion involving statistics 2.0 and computational sciences. *International Journal of Data Science and Analytics*, 6, 167–175.
- Madi M and Radovanovi M** (2012) Comparative modeling of CO₂ laser cutting using multiple regression analysis and artificial neural network. *International Journal of Physical Sciences*, 7, 2422–2430.
- Magalhães F, Cunha Á and Caetano E** (2009) Online automatic identification of the modal parameters of a long span arch bridge. *Mechanical Systems and Signal Processing* 23, 316–329.
- Marrongelli G, Gentile C and Saisi A** (2019) Anomaly detection based on automated OMA and mode shape changes: application on a historic arch bridge. In Arêde A and Costa C (eds), *Proceedings of ARCH 2019. Structural Integrity, Vol. 11*. Cham: Springer.
- Myers GJ** (1976) *Software Reliability*. New York: John Wiley & Sons.
- Sagiroglu S and Sinanc D** (2013) Big data: a review. In *International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, San Diego, CA, USA pp. 42–47.
- Sicherheit und Zuverlässigkeit** (2019) Available at <https://www.softwaresysteme.pt-dlr.de/de/sicherheit-und-zuverlaessigkeit.php> (accessed 9 December 2019).
- The Economist** (2017) The world's most valuable resource is no longer oil, but data. *Leader's Section*, May 6 2017.