

Let the Data Speak for Themselves: Opportunities and Caveats

Koenraad Debackere[†]

Faculty of Economics & Business, Katholieke University, Naamsestraat 6,
B-3000 Leuven, Belgium



Koenraad Debackere is a professor in innovation management and policy at University of Leuven, KU Leuven, Belgium. He has a specific interest in the development and use of research and development (R&D) and innovation indicators to measure the impact and outcome of R&D and innovation policy. He is the promotor-coordinator of ECOOM — the Flemish Centre on R&D monitoring.

Citation: Koenraad Debackere (2016). Let the Data Speak for Themselves: Opportunities and Caveats.
Received: Jan. 11, 2016
Revised: Jan. 14, 2016
Accepted: Jan. 15, 2016

“Let the data speak for themselves.” This is the strapline an academic, high-tech spin-off added to its logo. The spin-off originated from a computer science research group at a leading European university. This research group is heavily involved in novel, exciting scientific projects such as the development of neural network algorithms, advanced text mining methods and techniques, and machine learning. The group has been recognized for its various and many contributions to developments in the field of data analytics.

The broad field of data analytics is a rather new one. It operates at many intersections, such as computing and medicine or computing and business. One of the pioneers at the intersection of management and computing is Prof. Michael A. Rappa, at North Carolina State University, who already in the late 1990s started work on advanced analytics. This led to the creation of the North Carolina State University (NCSU) Institute for Advanced Analytics in 2006 and the design of a Master of Science in Analytics (MSA), see <http://analytics.ncsu.edu>. In the world of business economics, where an MBA education is a premier career step-stone, bringing an academic MSA degree to the forefront signals an important evolution. Business decisions are indeed increasingly based on data intelligence, the capability



JDIS
Journal of Data and
Information Science
Vol. 1 No. 1, 2016
pp 3–5

DOI: 10.20309/jdis.201602

<http://www.jdis.org>

[†] Corresponding author: Koenraad Debackere (E-mail: koenraad.debackere@kuleuven.be).

Perspective

and ability to efficiently and effectively bring together, analyze, and interpret huge dataset from various sources and with a wide variety of data types.

This brings us back to the spin-off company. The message the founders of the spin-off wanted to convey is clear. Starting from the humongous data availability, enabled by the omnipresent trend of increasing digitization, see also (Brynjolfsson & McAfee, 2014), in turn enabled by the still relevant Moore's law (Moore, 1965), data analytics is here to lead the way in many fields of science, technology, and decision making. We just need to think of modern medical research and its increasing reliance on molecular computing, molecular modeling, all the 'omics' research that has become mainstream in biomedical sciences, the availability of huge amounts of patient and clinical data, etc. In short, the continuous and significant progress in computing power, sensor technology and data capture, process and processing technology, mathematical engineering, and algorithm development are enabling the capture, the analysis, and the use of data as never before.

It thus is logical that across the globe scientific communities are now organizing around data analytics as a "field." New publication outlets emerge that support this journey. The *Journal of data and Information Science (JDIS)* (www.jdis.org) is one of them. It devotes itself to the converging area of data science, information science, and computer science, focusing on the study and application of theories, methods, techniques, services, etc., using big data to support knowledge discovery for decision and policy making, where the big data may include metadata or full content, text or non-textual, structured or non-structured, domain specific or cross-domain, dynamic or interactive data.

The consequences of this evolution are far-reaching. It creates exciting opportunities. Take for instance the seminal work by Charles P. Snow (Snow, 1959). In this book, Charles P. Snow develops and describes the concept of the Two Cultures. The two cultures he identified were those of the literary intellectuals and of the natural scientists. Between them, a wide and insurmountable gap exists, he claims. This profound incomprehension and mutual suspicion impede the application of technology to alleviate the world's grand challenges.

It is obvious that today's digital and data revolution challenges the very concept of two cultures. The advent of fields of inquiry like digital humanities points to the blurring of boundaries between the two cultures. Digital technology and information science are at the origin of a novel "merger" of the two cultures, creating new opportunities for research and theory development. And this brings us back to the spin-off and its strapline.

"Let the data speak for themselves" thus highlights the tremendous opportunities for scientific inquiry that emerge from the digital and data science revolution. However, it also brings an important caveat to the research desk or bench. Data that



merely speak for themselves may lead the way to an overly inductive scientific approach. Such inductive approach, though, can never substitute for diligent, accurate and rigorous theory development and building.

As a consequence data and information science should become strongly linked to the theory dimension of any scientific inquiry. It is obvious that the sophistication of novel data and information science methods offers tremendous opportunities for theory testing, theory advancement, and theory development. In order to fully capture those opportunities, it is important to avoid the slippery slope of pure induction. Hence, this perspective is a plea for the technical developments in data and information science to be continuously focused and fine-tuned by the selective lens of theory building and advancement. So, instead of letting the data speak for themselves, we should let them speak for theory.

References

- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. New York: W.W. Norton & Company.
- Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics*, 38(8), 114–117.
- Snow, C.P. (1959). *The two cultures and the scientific revolution*. London: Cambridge University Press.



This license allows readers to copy, distribute and transmit the Contribution as long as it is attributed back to the author. Readers are permitted to alter, transform or build upon the Contribution as long as the resulting work is then distributed under this or a similar license. Readers are not permitted to use the Contribution for commercial purposes. Please read the full license for further details at <http://creativecommons.org/licenses/by/4.0/>

