

## Chapter 5

# Data Science is Here: Are We Ready to Benefit From the Opportunities It Provides?

**Dimitar Grozdanov Christozov**

 <https://orcid.org/0000-0002-5780-3117>

*American University in Bulgaria, Bulgaria*

**Katia Rasheva-Yordanova**

*University of Library Studies and Information Technologies, Bulgaria*

**Stefka Toleva-Stoimenova**

*University of Library Studies and Information Technologies, Bulgaria*

### ABSTRACT

*With the advent of big data, the search for respective data experts has become more intensive. This study aims to discuss data scientist skills and some topical issues that are related to data specialist profiles. A complex competence model has been deployed, dividing the skills into three groups: hard, soft, and analytical skills. The primary focus is on analytical thinking as one of the key competences of the successful data scientist taking into account the trans-discipline nature of data science. The chapter considers a new digital divide between the society and this small group of people that make sense out of the vast data and help the organization in informed decision making. As data science training needs to be business-oriented, the curricula of the Master's degree in Data Science is compared with the required knowledge and skills for recruitment.*

DOI: 10.4018/978-1-7998-2104-5.ch005

*“Heads or tails, gentlemen?” said Swift.*

—Isaac Asimov, *The Machine that Won the War*, 1961

## INTRODUCTION

The recent evolution of information technologies reaches the point allowing different entities to accumulate “Big Data” as well as to offer tools to process such data in a manageable way. This offers great opportunities for deeper understanding of causes and effects, driving forces, influencing factors, and circumstances. But there are many challenges hindering the real benefit of exploring Big Data. About 60 years ago, the famous scientist and science-fiction novelist Isaac Asimov wrote a short story “The Machine that Won the War” pointing out the challenges we are facing now: the volume, variety, and velocity of data and the need of a huge computer – Multivac – to put all those data together, to optimize processing, and to predict outcomes. But the story pointed out many other problems as reliability of the supplied data, data processing algorithms, interpretation of results, and trust toward the results in making the final decisions. Now this science fiction vision is technological reality, but the challenges to benefit of those achievements are still the same: the many V’s (Laney, 2012; Normandeau, 2013) associated with Big Data, the reliability of data, the bias in developing algorithms, the bias in interpreting results, and the most important – the trust in making decisions based on results acquired via this technology.

In the last decade we are witnessing an unprecedented explosion of organizations’ attention toward data and how to benefit from data. Understanding of the value of data, together with availability of technologies allowing to process huge amounts of data objects in a meaningful time-frame, resulted in appearing of a new scientific field – Data Science. Today, Data Science is one of the most-discussed in research and practices as many organizations are striving to use the data they possess or control in a way to improve effectiveness and efficiency of their operations (Kowalczyk and Buxmann, 2014), and to gain competitive advantages. In a nut shell the issue is whether society as whole, different entities – commercial or public, and individual citizens are ready to benefit of the opportunities provided. What are the obstacles, success factors, and how to address the Big Data challenges?

Currently Big Data is defined by expanding Gartner’s original **3V** definition “*BD is high **volume**, high **velocity**, and/or high **variety** information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization (Laney, 2012).*” by adding more Vs as **Veracity** – the biases, noise and abnormality in data; **Validity** – whether the data are correct and accurate for the intended use; **Volatility** – how long is data valid and how long should it be stored; etc. (Normandeau, 2013) BD definition must also address the ability to explore data of a given amount and complexity by given IT – “a set of data objects represents BD if it is close to the upper bound of the volume and complexity of data that a human can manage to manipulate for purpose with the aid of available information technology” (Christozov and Toleva-Stoimenova, 2015). This definition addresses also human ability to learn via use of technology and also allows expanding further with appearing new dimensions of what represent “complexity”.

The information technologies nowadays allow registering and storing practically all facts associated to any event happen within the scope of an entity in searchable repositories. This naturally leads to accumulation of a huge volume of heterogenic data, known as “Big Data”. In many cases, this kind of storages are designed in a way to support efficient retrieval, but they do not support extracting useful patterns and relations – learning – in the needed level of convenience. In order to gain knowledge from

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/data-science-is-here/251310](http://www.igi-global.com/chapter/data-science-is-here/251310)

## Related Content

---

### The Effectiveness of Self-Directed English Learning through SNS: Adopting Facebook based on Gamification

Eun-Sok Won and Jeong-Ryeol Kim (2018). *International Journal of Mobile and Blended Learning* (pp. 1-10).

[www.irma-international.org/article/the-effectiveness-of-self-directed-english-learning-through-sns/205559](http://www.irma-international.org/article/the-effectiveness-of-self-directed-english-learning-through-sns/205559)

### Learning about Blended Learning through Students' Experiences: An Exploratory Study in Postgraduate Guidance and Counselling Programs in a University with Campuses in Australia and Singapore

Margaret Anne Carter, Paul Pagliano, Cecily Knight and Donna Goldie (2016). *Mobile and Blended Learning Innovations for Improved Learning Outcomes* (pp. 60-83).

[www.irma-international.org/chapter/learning-about-blended-learning-through-students-experiences/151855](http://www.irma-international.org/chapter/learning-about-blended-learning-through-students-experiences/151855)

### The Future of ICT/LKT in the Teaching and Learning of Social Sciences, Geography, and History: Adaptive Learning and Artificial Intelligence

(2020). *ICTs and Innovation for Didactics of Social Sciences* (pp. 268-291).

[www.irma-international.org/chapter/the-future-of-ictlkt-in-the-teaching-and-learning-of-social-sciences-geography-and-history/250925](http://www.irma-international.org/chapter/the-future-of-ictlkt-in-the-teaching-and-learning-of-social-sciences-geography-and-history/250925)

### Literacies and Learning in Motion: Meaning Making and Transformation in a Community Mobile Storytelling Project

Jessica Katherine Frawley and Laurel Evelyn Dyson (2018). *International Journal of Mobile and Blended Learning* (pp. 52-72).

[www.irma-international.org/article/literacies-and-learning-in-motion/210086](http://www.irma-international.org/article/literacies-and-learning-in-motion/210086)

### The Educational Research Flipped Inclusion Between Social Metamorphosis and Technocratic Hybridizations

Tonia De Giuseppe, Annalisa Ianniello, Eva Podovšovnik and Felice Corona (2020). *Examining the Roles of Teachers and Students in Mastering New Technologies* (pp. 27-57).

[www.irma-international.org/chapter/the-educational-research-flipped-inclusion-between-social-metamorphosis-and-technocratic-hybridizations/251306](http://www.irma-international.org/chapter/the-educational-research-flipped-inclusion-between-social-metamorphosis-and-technocratic-hybridizations/251306)