

The Evolution of Data Science: A New Mode of Knowledge Production

Jennifer Priestley, Kennesaw State University, USA

Robert J. McGrath, University of New Hampshire, USA

ABSTRACT

Is data science a new field of study or simply an extension or specialization of a discipline that already exists, such as statistics, computer science, or mathematics? This article explores the evolution of data science as a potentially new academic discipline, which has evolved as a function of new problem sets that established disciplines have been ill-prepared to address. The authors find that this newly-evolved discipline can be viewed through the lens of a new mode of knowledge production and is characterized by transdisciplinarity collaboration with the private sector and increased accountability. Lessons from this evolution can inform knowledge production in other traditional academic disciplines as well as inform established knowledge management practices grappling with the emerging challenges of Big Data.

KEYWORDS

Big Data, Computer Science, Data Science, Knowledge Management, Knowledge Production, Statistics, Transdisciplinarity

INTRODUCTION

The terms “big data”, “data science” and “analytics” have pervaded the global common speak over the past decade. While populist in many cases, these terms are rooted in the real practice of being able to measure and analyze phenomena in larger amounts, faster and with a longer and more robust historical perspective, all facilitated by technological advances and the lower cost of data storage. Data, once defined by a numerical representation of some measurement, has today evolved into an atomic unit that can be captured – that is measured, seen or heard – and thus extracted, analyzed and converted into information and ultimately into new knowledge. What began only a few years ago as a growing swell of the data ocean has become a tsunami of impacts into everyday life, or the “datafication” of the economy (Dumont, 2016).

This datafication has resulted in many organizations sprinting to better leverage the data they collect and capture the data they do not. The argument that knowledge, as a summation of data through the knowledge management pyramid (Ackoff, 1989), is the only sustainable source of competitive advantage is arguably more relevant today than when it was first posited (Drucker, 1995). It has also led many companies to declare that they are, in fact, data and information organizations more so than

DOI: 10.4018/IJKM.2019040106

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

they are purveyors of the products they sell (e.g. Capital One (Dee, 2016), Alibaba (Liyakasa, 2015) and Ford (Blanco, 2016)). Cities too are becoming “smarter” with data-driven innovations geared at efficient energy consumption, optimized traffic and parking, and the promotion of green and healthy practices. And individuals are becoming more data driven, with many exploring opportunities by an ever increasing “quantified self”; a concept related to the self-tracking of any number of physical, behavioral, social and many other phenomena by individuals (Swan, 2013). A revolution, or perhaps evolution, to be sure.

An unexpected consequence of these rapid (r)evolutionary changes has been the emergence of the ubiquitous and pervasive “talent gap” – the term used to describe the challenge of organizations to find people with the necessary skills to extract and analyze massive amounts of data (structured and unstructured) to generate meaningful information. Simply put, the demand for these skills has materialized so rapidly, traditional sources of supply for new talent (i.e., colleges and universities) have been ill-equipped to develop and train talent at the scale and pace demanded.

The issues related to the emergence of data science and the associated talent gap have implications for larger conversations related to organizational knowledge management. Jennex (2017) recognized the role of Big Data in the revised knowledge management pyramid. The traditional pyramid first presented by Ackoff (1989) established the framework that organizational wisdom derives from knowledge, information, and finally from data. In the revised pyramid, Jennex places a finer lens on the lowest level of the pyramid by calling out incremental layers between information and reality. These new layers include “Data”, defined as “discrete facts...that can be stored in a database” (Jennex & Bartczak, 2013), “Big Data”, defined as data that is “too big, too fast or too hard for existing tools to process” (Madden, 2012), and “IoT”, defined as a sensor network of networks with devices continually generating vast amounts of data and facilitating the evolving definition of what data even is. This evolution in thinking from a simplistic single layer at the base of the pyramid to a more detailed treatment of data within the knowledge management pyramid increases the resolution of the lens through which reality can be detected.

It is the concepts, tools, and algorithms around “data science” that will enable a sustainable organizational approach to the translation of the layers of data into information, knowledge and ultimately to organizational wisdom/intelligence. However, where those organizational knowledge activities meet societal ones and who addresses those “fault” lines become an issue as data sources become more democratized and real time (Spender, 2007; Money & Cohen, 2018).

These types of issues have led many in academia to consider the conversations around “data science” more formally. Is this truly a new field of study, or is data science simply an extension or specialization of a discipline that already exists, such as statistics or computer science, or mathematics? The answers to these questions are not trivial and have implications for both academics as well as practitioners engaged in addressing the challenges in knowledge production and management related to the emergence of Jennex’s more detailed treatment of data within the knowledge management pyramid.

A Brief History of Data Science

The term “data science” has been traced back to computer scientist Peter Naur in 1960 (Naur, 1992), but “data science” also has evolutionary seeds in statistics. In 1962, the famed statistician John W. Tukey wrote:

For a long time I thought I was a statistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have ... come to feel that my central interest is in data analysis... data analysis is intrinsically an empirical science. (Tukey, 1962)

The fields of data manipulation have grown largely through methods in mathematics, statistics and computer science during this period, with research from Peter Naur, who published “Concise Survey of Computer Methods” in 1974; Gregory Piatetsky-Shapiro who organized and chaired the

11 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/article/the-evolution-of-data-science/217372

Related Content

Advancing Knowledge About Sexual Harassment Is a Critical Aspect of Organizational Development for All Employees

Quatavia McLester, Darrell Norman Burrell, Calvin Nobles and Ileana Castillo (2021). *International Journal of Knowledge-Based Organizations* (pp. 48-60).

www.irma-international.org/article/advancing-knowledge-about-sexual-harassment-is-a-critical-aspect-of-organizational-development-for-all-employees/287776

A Revised Knowledge Pyramid

Murray E. Jennex and Summer E. Bartczak (2013). *International Journal of Knowledge Management* (pp. 19-30).

www.irma-international.org/article/a-revised-knowledge-pyramid/99641

Managing Legal and Negotiation Knowledge

John Zeleznikow (2011). *Encyclopedia of Knowledge Management, Second Edition* (pp. 1065-1071).

www.irma-international.org/chapter/managing-legal-negotiation-knowledge/49052

Transformation Based XML Query Optimization

Dunren Che (2005). *Transformation of Knowledge, Information and Data: Theory and Applications* (pp. 75-94).

www.irma-international.org/chapter/transformation-based-xml-query-optimization/30441

Groupware: Enabling Knowledge Sharing Across Time and Space

Meliha Handzic (2007). *Socio-Technical Knowledge Management: Studies and Initiatives* (pp. 58-68).

www.irma-international.org/chapter/groupware-enabling-knowledge-sharing-across/29336