

FEATURE

Are Data Scientists Doing the Job They Were Hired For?

Kon Leong

JAN 6, 2016 6:30am ET

Data scientists have it tough.

Really! It's an incredulous claim if one is to believe the media buzz regarding the desirability of the data scientist role. Data scientists currently have the privilege of having one of the hottest jobs and (and hottest pay scales) in modern business, but their reality is rarely the same as their job description. The problem is that the average data scientist isn't spending much time analyzing data – he or she is spending time managing it.

In a perfect world, data scientists would be free to access and manipulate all enterprise content quickly and fluidly without impairment. But the reality of the data environment for most businesses is a scattered and messy ecosystem of multiple systems and software; each used for different content and management functions.

Data is duplicated, disconnected, and disjointed. There is no single portal or platform for search. When data scientists are forced to gather content from IT systems that are sprawled across innumerable platforms and departments, they are left grasping at straws and with little more than a flawed convenience sample. Garbage in, garbage out.

The data scientist won't likely answer all of our problems, but data management just might – if given enough time and planning. As 2016 starts to dawn upon us, business leadership is starting to realize that analytics skills alone will do little to make sense of enterprise-scale content. Data is not useful until it has been managed and ordered, at which point it can be analyzed. The two can no longer be separate; as management practices form the foundation for all downstream handling of

content.

Supporting Data Scientists

To manage information defensibly and continuously isn't just an IT problem, it's a legal, compliance, business intelligence, risk management and productivity problem. It's a holistic business challenge that no single data scientist can tackle alone or be expected to fix; there needs to be buy-in and support from top level leadership. Data is the lifeblood of the business, and its smooth and efficient flow is necessary for knowledge and ideas to circulate.

The management of unstructured content has gone from being a security and legal concern to being a business-wide profitability concern. Data that is generated in the daily course of business is not simply a byproduct of human activity, but rather a massive trove of insight into workflow patterns, social connections, bottlenecks, and hubs of communication. In theory, this content should be massively valuable to a data scientist, yet lack of centralized management makes the data difficult to aggregate and manipulate. In order to execute their role correctly, data scientists need help with their data.

The Privacy Conversation

The concern over consumer data privacy has become a common household conversation, and there is reasonable consumer suspicion of aggressive data collection and use. However, a much less-discussed topic is the analysis of business data that employees generate while at work. Business emails, calendar entries, documents, and IMs all can reveal massive amounts of information regarding how people work, act, and think.

Traditionally, this content was considered "free game" – perfectly acceptable to analyze, because it was rightful property of the business. But as the privacy conversation becomes more nuanced, firms may need to give a second look at their internal data use.

Outcomes of Data Analysis

Not all insight derived from analysis is "good" insight; sometimes adverse – or even illegal – patterns can come to light during an attempt to analyze a seemingly benign set of data. The increasing ease-of-use of enterprise analytics tools means that potential employee harassment, product defects, or consumer complaints might be uncovered by an individual that does not know how to address or report these issues.

Public social media analysis by a marketing department may unexpectedly

reveal mass outcry over perceived adverse effects of a new product or device. Data scientists, in particular, are analysis experts and not legal experts. A business that is truly supportive of their data scientists will take ongoing steps to ensure that all individuals involved in data manipulation are given ongoing training and direct lines of communication to compliance, risk management, and legal teams.

Your data scientist – no matter how brilliant – cannot squeeze blood from a stone. The outcome of data analysis depends almost entirely on the underlying infrastructure of enterprise information governance, and 2016 will only further bring difficulty for organizations that have not invested time and effort into their data management practices. The more scattered and overlapping systems are, the less valuable the data becomes for large-scale analysis.

It's time to give data scientists a hand and make a New Year's resolution to start the governance discussion with all stakeholders at the table. Analytics insight won't be the only thing to improve – the entire health and profitability of the business will.

(About the author: Kon Leong is president, CEO and co-founder of ZL Technologies)

© 2016 [SourceMedia](#). All rights reserved.