Energy Policy 81 (2015) 117–121

Contents lists available at ScienceDirect

Energy Policy

journal homepage: www.elsevier.com/locate/enpol

Viewpoint

Data as an asset: What the oil and gas sector can learn from other industries about "Big Data"

Robert K. Perrons [a,b,*], Jesse W. Jensen [c]

[a] Queensland University of Technology, GPO Box 2434, Brisbane, Qld 4001, Australia
[b] Centre for Strategy and Performance, University of Cambridge, United Kingdom
[c] Intel Corporation, 2200 Mission College Blvd., Santa Clara, CA 95054, USA

HIGHLIGHTS

- Upstream oil and gas industry frequently discards or ignores the data it collects
- The sector tends to view data as descriptive information about the state of assets
- Leaders in Big Data, by stark contrast, regard data as an asset in and of itself
- Industry should use Big Data tools to extract more value from digital information

ARTICLE INFO

Article history:
Received 9 October 2014
Received in revised form
15 January 2015
Accepted 22 February 2015

Keywords:
Big data
Oil and gas
Information technologies
Data

ABSTRACT

The upstream oil and gas industry has been contending with massive data sets and monolithic files for many years, but "Big Data" is a relatively new concept that has the potential to significantly re-shape the industry. Despite the impressive amount of value that is being realized by Big Data technologies in other parts of the marketplace, however, much of the data collected within the oil and gas sector tends to be discarded, ignored, or analyzed in a very cursory way. This viewpoint examines existing data management practices in the upstream oil and gas industry, and compares them to practices and philosophies that have emerged in organizations that are leading the way in Big Data. The comparison shows that, in companies that are widely considered to be leaders in Big Data analytics, data is regarded as a valuable asset—but this is usually not true within the oil and gas industry insofar as data is frequently regarded there as descriptive information about a physical asset rather than something that is valuable in and of itself. The paper then discusses how the industry could potentially extract more value from data, and concludes with a series of policy-related questions to this end.

1. Introduction

"Big Data"[1] is a rather vague term that describes the application of new tools and techniques to digital information on a size and scale well beyond what was possible with traditional approaches (Lohr, 2012b), typically involving data sets that are so large and complex that they require advanced data storage, management, analysis, and visualization technologies (Chen, et al., 2012). Like many industries, the upstream oil and gas sector has seen a flurry of initiatives and high-profile publications (e.g., Anand, 2013; Beckwith, 2011; Holdaway, 2014) about this topic, which have in turn translated into significant discussion within industry conferences (e.g., Feblowitz, 2013) and among practitioners. Critics of Big Data caution that the transformational potential of these analytical capabilities may be somewhat oversold and misunderstood (Harford, 2014; Lohr, 2012b),[2] but the oil and gas sector has already been noticeably impacted by several of the technologies underpinning these changes.

* Corresponding author at: Queensland University of Technology, GPO Box 2434, Brisbane, Qld 4001, Australia.
E-mail address: perrons@alum.mit.edu (R.K. Perrons).

[1] The term "Big Data" is believed to have been coined by the astronomy and genomics communities in the 2000s (Mayer-Schönberger and Cukier, 2013, p. 6), but the concept has been used more widely since then.

[2] Gary King, the director of Harvard University's Institute for Quantitative Social Science, goes as far as predicting that there "is no area that is going to be untouched" by Big Data (Lohr, 2012a).

http://dx.doi.org/10.1016/j.enpol.2015.02.020
0301-4215/

By most accounts, the oil and gas industry's data is already "big." Modern oil and gas seismic data centers can easily contain as much as 20 petabytes[3] of information, which is roughly equivalent to 926 times the size of the U.S. Library of Congress (Beckwith, 2011). If this amount of information was copied into books and put on a single continuous bookshelf, it would go around the Earth's equator approximately six times (Beckwith, 2011). And while seismic data sets are notoriously large and cumbersome, many other aspects of the oil and gas industry are also generating significantly more data than they used to (Perrons, 2010b). What is more, there is every reason to believe that this trend towards more digital information is just getting warmed up. Current estimates suggest that the total amount of digital data in the world—including things like books, images, e-mails, music, and video—is doubling every 2–3 years (Lohr, 2012a; Mayer-Schönberger and Cukier, 2013).

Big Data tools and techniques have most famously delivered value in the social media sector and retail industries (Harford, 2014; Mayer-Schönberger and Cukier, 2013), but have also led to major breakthroughs in a diverse range of other contexts, including scientific research (Frankel and Reid, 2008; Kluger, 2014; Lynch, 2008), healthcare (Chen, et al., 2012), heavy equipment manufacturers (Mehta, 2013), and professional sports (Leahey, 2013).[4]

2. What has made Big Data possible?

Big Data is not a result of a single silver-bullet technology, but rather the coming together of several innovations and novel ideas in a highly complementary way. Four of these technological developments are particularly noteworthy:

(1) A precipitous decline in data storage costs. As shown in Fig. 1, the cost of storing digital information has been falling at an exponential rate for a long time. Several years ago, it was standard practice in many industries to discard significant collections of data when their initial use had passed, as there was a real economic expense associated with archiving the data afterward (e.g., Feblowitz, 2013). This is much less true today (Komorowski, 2014).

(2) Continued growth in the processing speeds of computing devices. Moore's Law, which states that the number of transistors on integrated circuits doubles approximately every two years, has been continuing unabated since the 1970s (Fig. 2).[5] It therefore follows that the amount of computing power offered in commercially available devices has been increasing at a similarly impressive rate (Ball, 2000).

(3) Breakthroughs in relevant areas of mathematics. Whereas traditional data sets have historically needed to be fairly structured, orderly, and static, digital information in the era of Big Data is frequently noisy, messy, raw, unstructured, and dynamic (Ouellette, 2013). Recent developments in mathematics—most notably, geometry—have significantly helped Big Data practitioners see through the messiness of these new data sets to find useful information and relationships. Considerable progress has been made in representing massive data sets as networks of geometrical nodes and edges so that the data can be rationalized using a suite of mathematical tools known as topological data analysis (TDA). Simply put, TDA is a way of getting structured information out of unstructured data so that machine-learning algorithms can be applied to it (Carlsson, 2009; Ouellette, 2013).

(4) The development of software platforms such as Google's MapReduce or its open-source rival Apache™ Hadoop®.[6] These tools make it possible to break large data sets into smaller chunks that can be delegated to several computing devices. The results of the calculations arising from each of the smaller chunks can then be re-integrated at the end of the process. This approach frequently uses cloud computing infrastructure as a platform for transferring these data chunks to different computing devices, and then bringing back the results.

3. How is Big Data different from what was done previously?

Despite the hype that has been generated around the topic of Big Data, the overarching objective is something that organizations in the energy sector and everywhere else have been aspiring to for a long time: to make better decisions (Regalado, 2014). What is changing, however, are the specific mechanisms by which these decisions are made. First, Big Data differs from traditional approaches on account of the "three Vs": volume, velocity, and variety (McAfee and Brynjolfsson, 2012). The steadily decreasing costs associated with collecting and storing data have resulted in a fundamental shift in thinking about data quality and volume. Historically, data collection was predicated on sampling from a subset of an overall population, and trying to make the collected data from that sample as accurate as possible. By contrast, the move towards Big Data has led to a much greater tolerance for messiness and imprecision. This more relaxed approach to vagueness has been compensated, however, by much larger volumes of data. Underlying this change of philosophy is the belief that "more trumps better" (Mayer-Schönberger and Cukier, 2013, p. 33).

These much larger volumes of data are now able to move with ever-increasing velocities such that a bewildering number of system variables can be monitored in nearly real-time. Moreover, this data is coming from a wider variety of sources and in an increasingly broad array of formats. Some data is "born digital," meaning that it was created specifically for digital use by a computer or data processing system (e.g., e-mail, web browsing, GPS locations); other data is "born analog," meaning that it comes from the physical world, but can increasingly be converted into a digital format (e.g., voice or visual information captured by phones, cameras, or video recorders, or data collected from wearable devices). The rising capability of "data fusion" makes it more and more possible to bring together disparate sources of data to glean fresh insights that nobody predicted (White House, 2014).

4. How Big Data is unfolding differently in the oil and gas sector

In light of this sweeping global trend, it is hard to imagine a future in which the oil and gas industry is not collecting

[3] This statistic is less impressive when you consider that Walmart, the U.S. retail giant, collects more than 2.5 petabytes of data every hour from customer transactions (McAfee and Brynjolfsson, 2012).

[4] The 2011 movie *Moneyball*, starring Brad Pitt, chronicles how the low-budget Oakland A's started using historical performance data and arcane baseball statistics to spot undervalued players. Intensive data analysis has since become commonplace not only in baseball but also in other sports, including English soccer (Lohr, 2012a).

[5] Note that even though this figure's timeline stops at the year 2000, the microprocessor industry has more or less continued to deliver new products that have kept pace with Moore's Law until the present day.

[6] Using Hadoop, Visa was able to reduce the processing time for two years' worth of test records—which translates to approximately 73 billion transactions—from one month to only 13 min (Mayer-Schönberger and Cukier, 2013, p. 46).

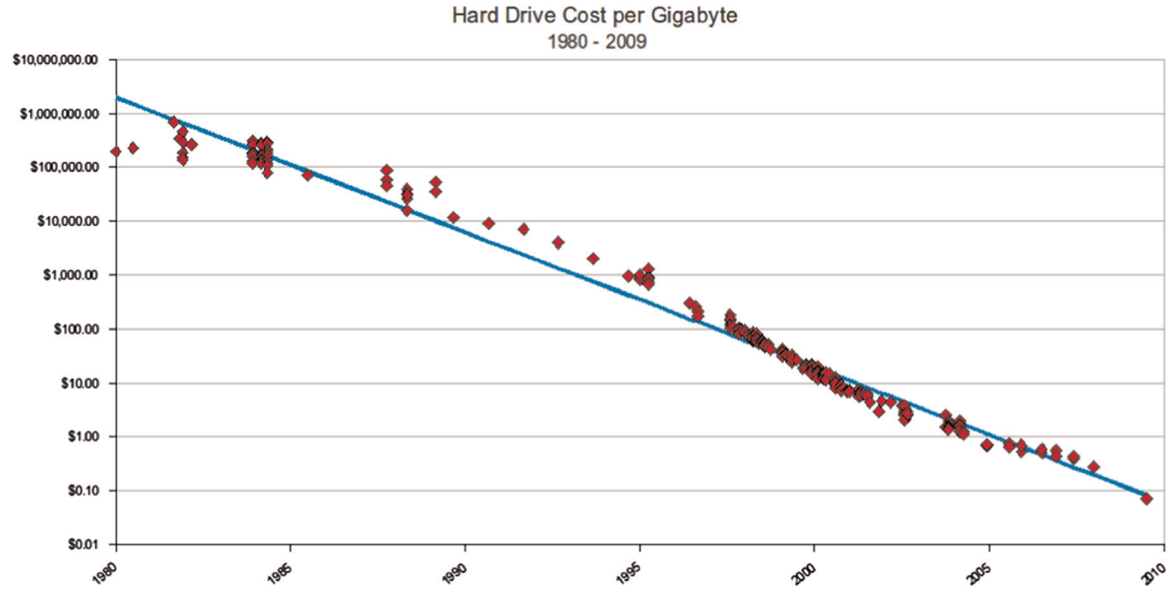

Fig. 1. Data storage costs, 1980–2009 (Komorowski, 2014)

significantly more data than it does at the moment. And with such large amounts of digital information accumulating around them, it is easy to understand why many industry insiders believe that they are already solidly on track to reap the benefits of Big Data.

But the oil and gas sector seems to be approaching these rapidly growing data sets with the same attitudes and analytical techniques that have been with the industry for years. As Feblowitz (2013) suggests, a lot of potentially valuable digital information harvested from upstream oil and gas assets is barely given a cursory glance, and much of it is simply thrown away. Moreover, in those instances where data is stored, it is often kept by the service companies responsible for generating it rather than the operator in charge of managing the long-term welfare of the asset.

This is not how the Big Data revolution is unfolding in many other industries. Several other sectors—most notably, the healthcare, financial, retail, and media industries—have come to realize that new and valuable insights are frequently gleaned from using new techniques to analyze massive data sets in ways that were never possible with smaller ones. These insights tend not to be discovered by testing hypotheses between variables whose relationships are well understood; rather, they are found by applying advanced analytical techniques to massive numbers of variables that, at first blush, might seem to be unrelated. FICO, an American analytics company, has discovered a surprisingly tight relationship between aspects of a person's car ownership records and their propensity to take prescribed medication. Aviva, a large UK-based insurance firm, has developed predictive models that use credit reports and consumer-marketing data to identify health risks among prospective applicants. And by carefully examining what its customers are purchasing from different departments, Target, the U.S. retailer, "knows when a woman is pregnant without the mother-to-be explicitly telling it so" (Mayer-Schönberger and Cukier, 2013, p. 57).

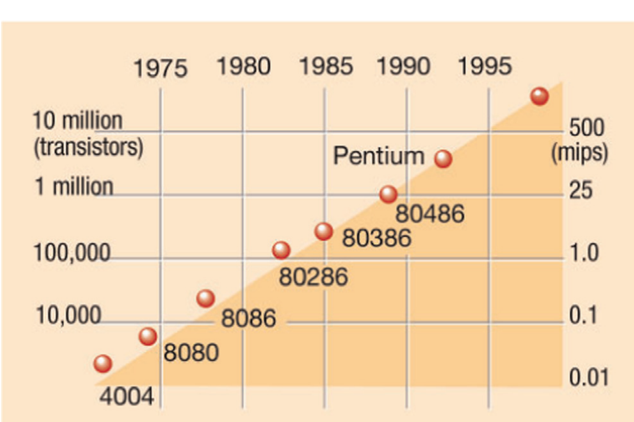

Fig. 2. Increase in availability of computing power, 1970–2000 (Ball, 2000).

Companies that are leading the charge in Big Data are not merely creating value by monitoring relationships that they already knew about, but by finding patterns and making predictions based on complex relationships that were previously unknown. And whereas scientists have traditionally sought to understand the causality and mechanisms underlying these kinds of relationships, cutting-edge users of Big Data frequently care about the "what" far more than the "why." In other words, "When we let the data speak, we can make connections that we had never thought existed" (Mayer-Schönberger and Cukier, 2013, p. 14). This fundamentally different approach to data analysis carries with it a valuable lesson about the kinds of information that should be collected: in the age of Big Data, potential value is lurking within *all* digital information, no matter how inconsequential and disconnected it might seem at the time it is collected.

Therein lies a fundamental difference of opinion between the oil and gas sector and other industries that are widely considered to be leaders in Big Data such as Facebook and Amazon. While there is no denying that the upstream oil and gas industry is awash in digital information—and, indeed, several Big Data technologies have been used for much longer in that industry than in many others—the way that the industry manages data does not actually bear much resemblance to how it gets used by Big Data leaders in Silicon Valley or in the retail sector. The oil and gas industry tends to regard data as information that describes the state of an asset; leaders in Big Data, by stark contrast, realize that *data is a valuable asset in and of itself*.

It goes without saying that the business models behind social networks and online retailing are profoundly different from that of the oil and gas sector, and that data delivers value within the upstream oil and gas industry in a way that is unlike many other sectors. But at the same time, it is no less true that the competitive landscape of the oil and gas industry is growing ever more reliant on information technologies and computing power. The internet began as a curious plaything at the beginning of that revolution but, over the past 20 years, it has clearly become an indispensable part of how business is conducted within the sector. And so it goes with Big Data: many of today's oil and gas companies clearly do not consider most of their digital information to be mission-critical to their profitability (Feblowitz, 2013). But they probably will start to think of data that way in the years ahead.

In the future, the issue of data management—including how data will be collected, formatted, stored, and owned—will be an important part of service contract negotiations and agreements between collaborating firms. Companies within the industry will jealously guard the data that they hold and, as suggested by Anand (2013), they will strategically attempt to augment in-house data assets with digital information procured from outside sources, thereby giving them the kinds of massive data sets in which the value of Big Data is often lurking.

5. Conclusions and key questions

This viewpoint outlined the high-level capabilities and new technologies that have given rise to Big Data, and briefly examined the potential relevance of these changes to the oil and gas sector. The amount and kinds of value that Big Data can deliver will clearly vary considerably from one industry to the next, but a few important themes are emerging across the entire business landscape. Among the most important of these is an evolutionary change in the market's perceptions of data—in particular, the notion of data as an asset (Perrons and Jensen, 2014b) described earlier. This fundamental shift raises several important policy-related questions for the oil and gas industry. Specifically:

(1) In the future, significant strategic advantage will arise from an organization's ability to be the integrator and high-level analyzer of data within its ecosystem. How should oil and gas companies re-shape their contracting and collaboration strategies in light of this new reality?

(2) Big Data is not merely about organizations doing what they have always been doing, but with more data. Instead, these forces are resulting in fundamentally new business models in many industries (Mayer-Schönberger and Cukier, 2013). So how could Big Data change the oil and gas industry's business model?

(3) Finally, Big Data is resulting in noticeable changes in job descriptions and the internal architecture of organizations. A new breed of professionals known as “data scientists” has emerged that understands analytics, information technologies (IT), and mathematics while also having the ability to communicate effectively with decision-makers (Davenport, et al., 2012). This represents a marked departure from the data analysts that have been residing in many organizations for years. EMC Corporation, a large provider of data storage technologies, puts such a premium on the value of these multi-disciplinary skill sets that it acquired Greenplum, a Big Data technology company, so that it could expand its capabilities in this domain and develop these kinds of data scientists (Davenport, et al., 2012). Should oil and gas companies be cultivating teams of data scientists in order to address the rising importance and increasingly multi-disciplinary nature of IT and data management? And how should these new roles be connected to existing parts of the company? At Merck, the pharmaceuticals giant, data scientists are members of the of the drug discovery and development teams. A few universities and organizations within the upstream oil and gas industry are making modest inroads in this area (Mahdavi, 2008; Perrons, 2010a) but, for the most part, the sector's discipline silos have remained largely unchanged.

The industry has succeeded handsomely in learning how to generate a staggering amount of data, but it is still wrestling with the question of what to do with it. Early indicators of change are indeed starting to appear, however. Recent discussions in more practitioner-focused venues hint at some emerging firm-level policies that are clearly intended to create data capture strategies with Big Data types of capabilities in mind (Coles, et al., 2014). Also, new policies and organizational structures are being experimented with in individual operating units to support the application of Big Data types of tools and techniques in a broad range of topics, including well and reservoir management (Crockett and Kurrey, 2014), health and safety (Pettinger, 2014), and subsurface exploration (Binotto, et al., 2014). The number and boldness of these kinds of policy changes will only gather pace as the years roll on and Big Data types of analytical methods gradually become more accepted within the industry.

The oil and gas sector's digital revolution is unfinished. The case for moving towards digital oilfield technologies was largely based on the ability of those tools to help the industry make better decisions—and when you peel back all the hype, that is ultimately what Big Data is about, too (Regalado, 2014). In this way, Big Data is not the dawn of a new age for the oil and gas sector, but rather the next phase of a digital transformation that started a long time ago. The industry's digital revolution will be complete when the sector figures out how to monetize the data that it is now capable of collecting, and then uses it to create all the value that it can.

Acknowledgments

A different version of this viewpoint (Perrons and Jensen, 2014a) was presented at the Society of Petroleum Engineers' 2014 Annual Technical Conference and Exhibition (ATCE) in Amsterdam, the Netherlands. Several of the points discussed here were also put forward in a short guest editorial (Perrons and Jensen, 2014b) principally aimed at oil and gas industry practitioners. Special thanks to Parviz Peiravi and Ron Kasabian from Intel for their insights and ideas throughout the preparation of this paper, and to three anonymous reviewers for their feedback.

References

Anand, P., 2013. Guest editorial: Big Data is a big deal. J. Pet. Technol. 65 (4), 18–21.
Ball, P., 2000. Chemistry meets computing. Nature 406 (6792), 118–120.
Beckwith, R., 2011. Managing Big Data: cloud computing and co-location centers. J. Pet. Technol. 63 (10), 42–45.
Binotto, A.P.D., Sultanum, N., Cerqueira, R., September 15–17, 2014. Cloud-based remote visualization of Big Data to subsurface exploration. In: Proceedings of the Society of Petroleum Engineers (SPE) Large Scale Computing and Big Data Challenges in Reservoir Simulation Conference and Exhibition. Paper no. SPE-172983-MS. Istanbul, Turkey.
Carlsson, G., 2009. Topology and data. Bull. Am. Math. Soc. 46 (2), 255–308.
Chen, H., Chiang, R.H.L., Storey, V.C., 2012. Business intelligence and analytics: from Big Data to big impact. MIS Q. 36 (4), 1165–1188.
Coles, D., Prange, M., Djikpesse, M., October 26–31, 2014. Optimal survey design for big geodata. In: Proceedings of the Society of Exploration Geophysicists Conference. Paper no. SEG-2014-0224. Denver, Colorado, USA.
Crockett, B., Kurrey, K., April 1–3, 2014. Smart decision making needs automated analysis: making sense out of Big Data in real-time. In: Proceedings of the Society of Petroleum Engineers (SPE) Intelligent Energy Conference and Exhibition. Paper no. SPE-167879-MS. Utrecht, the Netherlands.
Davenport, T.H., Barth, P., Bean, R., 2012. How "Big Data" is different. MIT Sloan Manag. Rev., 54; pp. 22–24.
Feblowitz, J., March 5–7, 2013. Analytics in oil and gas: the big deal about Big Data. In: Proceedings of the SPE Digital Energy Conference and Exhibition. Paper no. SPE 163717. The Woodlands, Texas, USA.

Frankel, F., Reid, R., 2008. Distilling meaning from data. Nature 455 (7209), 30.
Harford, T., 2014. Big Data: Are We Making a Big Mistake? Financial Times, March 28 issue. Retrieved on 31 March 2014 from ⟨http://www.ft.com/intl/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz2xV4TdTmn⟩.
Holdaway, K.R., 2014. Harness Oil and Gas Big Data with Analytics. John Wiley and Sons, Hoboken, New Jersey.
Kluger, J., 2014. Finding a second Earth. Time (Asia-Pacific edition) 183 (1), 30–32.
Komorowski, M., 2014. A History of Storage Costs. Retrieved on 19 May 2014 from ⟨http://www.mkomo.com/cost-per-gigabyte⟩.
Leahey, C., 2013. The NFL's Plan to Tackle Big Data. Fortune (Asia-Pacific edition) 168 (9), 9.
Lohr, S., 2012a. The age of Big Data. New York Times, February 11 issue. Printed on p. SR1 of the New York edition of the newspaper. Retrieved on 19 May 2014 from ⟨http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html⟩.
Lohr, S., 2012b. How Big Data became so big. New York Times, August 11 issue. Also published on p. BU3 of the August 12 issue of the New York Edition with the headline "Amid the Flood, a Catchphrase is Born." Retrieved on 19 May 2014 from ⟨http://www.nytimes.com/2012/08/12/business/how-big-data-became-so-big-unboxed.html⟩.
Lynch, C., 2008. How does your data grow? Nature 455 (7209), 28–29.
Mahdavi, M., 2008. The digital petroleum engineer: Carpe Diem! (Guest Editorial). J. Pet. Technol. 60 (10), 16–17.
Mayer-Schönberger, V., Cukier, K., 2013. Big Data: A Revolution That Will Transform How We Live, Work and Think. John Murray Publishers, London, United Kingdom.
McAfee, A., Brynjolfsson, E., 2012. Big Data: the management revolution. Harv. Bus. Rev. 90 (10), 60–68.
Mehta, S.N., 2013. Where brains meet brawn. Fortune (Asia-Pacific edition) 168 (7), 34.
Ouellette, J., 2013. The Mathematical Shape of Things to Come. Quanta Magazine (online), October 4 issue. Retrived on 20 May 2014 from ⟨http://www.simonsfoundation.org/quanta/20131004-the-mathematical-shape-of-things-to-come/⟩.
Perrons, R.K., 2010a. Organizational flexibility in the digital age: the stakes are higher than you think. J. Pet. Technol. 62 (6), 16–18.
Perrons, R.K., 2010b. Perdido ties together shell digital oilfield technologies. World Oil 231 (5), 43–49.
Perrons, R.K., Jensen, J., October 27–29, 2014a. Data as an asset: what the upstream oil and gas industry can learn about "Big Data" from companies like Facebook. In: Proceedings of the Society of Petroleum Engineers Annual Technical Conference and Exhibition (ATCE). Paper no. SPE-170633-MS. Amsterdam, the Netherlands.
Perrons, R.K., Jensen, J., 2014b. The unfinished revolution: what is missing from the E and P industry's move to "Big Data". J. Pet. Technol. 66 (5), 20–22.
Pettinger, C.B., June 8–11, 2014. Leading indicators, culture and Big Data: using your data to eliminate death. In: Proceedings of the American Society of Safety Engineers (ASSE) Professional Development Conference and Exposition. Paper no. ASSE-14-689. Orlando, Florida, USA.
Regalado, A., 2014. Data and decision making. MIT Technol. Rev. 117 (2), 61–63.
White, House, 2014. Big Data: Seizing Opportunities, Executive Officer of the President. Preserving Values, Washington, DC.