Journal of
Organization Design

Big Data and Organization Design

Vol. 3, No. 1 (2014)

Special Issue Editors:
Richard M. Burton
Dolly Mastrangelo
Fabrizio Salvador

ISSN 2245-408X

TABLE OF CONTENTS

SPECIAL ISSUE: BIG DATA AND ORGANIZATION DESIGN

RESEARCH ARTICLES

POINT OF VIEW ARTICLES

CASE STUDY

Journal of Organization Design
ISSN: 2245-408X

INTRODUCTION

This Special Issue on Big Data and Organization Design addresses the challenge of big data for the design of organizations. Big data refers to the availability to organizations of massive amounts of heterogeneous and continuously updated information. Practitioners agree that the availability of such information creates challenges and opportunities for organizations that have never been seen before. The articles presented here take up this challenge and discuss avenues for future research and practice on organization design in the era of big data.

The genesis of the special issue stems from papers presented at the World Summit on Big Data and Organization Design, which was held at the Université Panthéon-Sorbonne (Paris) on May 16-17, 2013. Initiated by the Organizational Design Community (ODC) and co-sponsored by IBM, the Interdisciplinary Center for Organizational Architecture (ICOA) at Aarhus University (Denmark), and the Université Panthéon-Sorbonne, the conference featured 11 keynote and distinguished speakers and included 130 participants from 26 countries. The speakers and participants represented academia, business, and government. Seventy-four papers co-authored by the 130 participants were accepted by the Organizing Committee. The articles in this special issue offer a selection of the issues and opportunities posed by big data and their implications for organizations.

The special issue begins with three research articles. Galbraith builds on his earlier ideas on big data and organization design (*Journal of Organization Design*, Vol. 1, Issue 2) to discuss shifts in the internal distribution of power likely to be brought on by a strategic emphasis on big data. He also discusses how seizing the opportunity of big data implies increasing the speed of decision making within the organization, enabling the creation of entirely new businesses. Berner, Graupner, and Maedche develop the provocative proposition that big data requires a transformation from command and control hierarchies to post-bureaucratic organizational structures and processes wherein employees at all levels are empowered while simultaneously being controlled. Grossman and Siegel address the issue of how analytics capability is distributed within an organization, stressing the importance of building a critical mass of analytics staff, centralizing or distributing the analytics staff to support critical big data processes, and establishing an analytics governance structure to ensure that critical analytics processes are supported by the organization as a whole.

Next, two point of view articles discuss the implications of big data for higher education and for organizational structure. Miller argues that realizing the potential of big data requires a new mind-set that is not yet reflected in the academic curricula of universities scrambling to develop degree programs in data science. Korhonen speculates about how a strategic emphasis on big data will be manifested in an organization's structure. Noting that historically strategy and structure co-evolve, he predicts the shape of an organization that embraces the big data phenomenon.

Finally, the case study article by Gabel and Tokarski discusses how big data is affecting their own organization. Their account of RTI International, a non-profit survey research firm, shows how an organization with plenty of quantitative analysis talent and expertise nevertheless needs to engage in a major transformation in order to deal successfully with big data issues.

We, the special issue editors, wish to thank the authors for their contributions, the reviewers for their advice, and the editors of *Journal of Organization Design* for publishing these interesting and valuable articles. We hope the special issue sparks research and debate on the topic of big data and organization design.

Richard M. Burton
Dolly Mastrangelo
Fabrizio Salvador

JOD, 3(1): 1 (2014)
DOI: 10.7146/jod.16855

ORGANIZATION DESIGN CHALLENGES RESULTING FROM BIG DATA

JAY R. GALBRAITH

Abstract: Business firms and other types of organizations are feverishly exploring ways of taking advantage of the big data phenomenon. This article discusses firms that are at the leading edge of developing a big data analytics capability. Firms that are currently enjoying the most success in this area are able to use big data not only to improve their existing businesses but to create new businesses as well. Putting a strategic emphasis on big data requires adding an analytics capability to the existing organization. This transformation process results in power shifting to analytics experts and in decisions being made in real time.

Keywords: Big data, organization design, analytics capability, strategy and structure, new organizational forms

The subject of big data has burst on the scene in the last few years. Today, it is receiving more than its fair share of media hype. If we look beyond the hype, however, we can see some real substance underlying this phenomenon. Big data has been given legitimacy by reports from the World Economic Forum, the McKinsey Global Institute, and *The Economist* Intelligence Unit. In their recent book, *The New Digital Age*, Eric Schmidt and Jared Cowen (2013), executives and thought leaders at Google, have given us a balanced discussion of the entire digital landscape. Nonetheless, the digital age has its critics who warn of the dangers involved (Morozov, 2013). At the end of the day, I believe that privacy issues and other dangers will be resolved and big data – or whatever we eventually call it – will be a capability that is designed into all of our organizations. However, organizations will vary in the difficulties they experience in building their big data analytics capability.

As organizations attempt to develop a big data analytics capability, they will encounter obstacles as well as opportunities. In this article, I discuss how big data presents a company with the opportunity to start an entirely new business. To take advantage of this opportunity requires analytics capability that shifts power in the organization and dramatically increases the speed of decision making. (For a complete discussion, see Galbraith, 2014.) After first discussing the twin design challenges of power shifts and real-time decision making, I describe how Nike's creation of a Digital Sports Division has taken advantage of big data to build an entirely new business. At the end of the article, I summarize the impact of big data on organization design using the Star Model™ framework.

NEW ATTRIBUTES OF BIG DATA

Firms and other organizations have been using large databases and analytics for the last couple of decades. Transactions are stored in data warehouses and analyzed with data-mining algorithms to extract insights. What is new about big data today? First, there are more – and different kinds – of data. In the past, it was stored, structured data. This data was largely from transactions and was stored as rows and columns. Today, we store unstructured data from a variety of sources. The data could be photos from a mobile phone, maps from a GPS device, video from a surveillance camera, audio from a call center, e-mails, tweets, and text messages. All of this data can be digitized, analyzed, and stored.

Journal of Organization Design
JOD, 3(1): 2-13 (2014)
DOI: 10.7146/jod.8856

Second, this new data is accessible in real time. Before, data in the data warehouse was historical and described outcomes that had already occurred. Now, we can receive data about events as they are happening and perhaps influence their outcomes. Historically, credit card companies stored all of their transactions in a database and analyzed them with fraud-detecting algorithms. Fraudulent transactions were then turned over to the police to investigate. The companies could distinguish chronically late payers from people who had lost their jobs. Customer service could then take the appropriate actions with each group. Today, a fraudulent transaction can be detected while the fraudster is still at the checkout counter. An algorithm operating in real time can determine that the transaction is a charge on a stolen credit card. The clerk at the checkout counter can be advised to delay the suspect. Store security can be informed to apprehend the person and confiscate the credit card. Thus, real-time data allow us to influence the outcome and prevent bad outcomes before they happen. This capability is new. However, this new capability is only possible if we have an organization that is designed to operate in real time. We need to design a decision process that uses real-time data, analyzes it to produce instant insights, and processes those insights to arrive at real-time decisions. Using real-time decisions, organizations can take quick action. We need much faster-acting companies in order to profit from big data.

POWER SHIFTS

Before an organization can make real-time decisions, it must get data scientists and analytics experts embedded into decision processes. This will require a shift in power from experienced and judgmental decision makers to digital decision makers. Every organization has an establishment, a power structure with a vested interest in the status quo. The establishment is currently making investment decisions, setting customer priorities, and deciding on new product features. These are the same decisions that new insights from big data can improve. But will the current leadership adopt or reject new insights? In order to be successful, the organization needs to execute a shift in power to the digital experts who generate new insights from big data. A shift in power is necessary to accomplish the changes that are needed to fully embed the big data analytics capability.

Competence-Enhancing

One factor that will determine the magnitude of the power shift is the amount of resistance that the big data proponents will encounter from the establishment. The amount of resistance will depend on whether this new capability is competence-enhancing or competence-destroying (Tushman & Anderson, 1996). For example, when e-commerce came along, Dell adopted it immediately. Taking orders over the Internet was competence-enhancing for Dell; it enhanced the company's direct sales to end-users business model. Hewlett-Packard's strength was its relationship with resellers and retailers. E-commerce was a competence-destroying innovation; it would disintermediate HP's resellers. HP was slower to adopt e-commerce, keeping the firm's resellers in the distribution chain at a higher cost. So, with respect to any particular big data initiative, companies need to determine where they are on the competence enhancing-destroying continuum.

Procter & Gamble is an example of a company for whom big data is competence-enhancing. P&G is a very analytical company and has had an analytics group since 1992. Plus, it will try anything that might increase its understanding of consumer behavior. As a result, P&G is adopting big data practices ahead of most other companies. The big data initiative is led by the CIO and supported by the CEO. P&G has adapted its hiring practices to bring in more data scientists. For the past five years, P&G and Google have exchanged teams of people annually. Google wants to learn about advertising, and P&G wants to learn from Google's digital acumen. At P&G, all managers are upgrading their digital skills. Moreover, every manager's digital and analytical performance gets assessed in the performance management process. The CIO and the business leaders have identified 88 business processes that are being redesigned and accelerated to operate in real time. So, P&G is a good example of a company that has embraced big data.

In contrast, a good example of a competence-destroying situation is the arrival of big data in sports. Most of us have seen the baseball movie Moneyball. Billy Beane, the general manager of the Oakland Athletics, brings in "sabre-metrics" expert Peter Brand. Billy wants Peter to advise him on putting together the best team possible, but on Oakland's very low budget. Peter is 25 years old and an economics graduate from Yale who has never played the game. Of course, they run into the chief scout, Grady, and his grey-haired scouting team. Grady first tries to keep Peter out of the meeting. "Does Pete really need to be here?" Then he shifts to, "You can't put a team together with a computer." The meeting is a clash between Peter's data and Grady's opinions. Billy makes his decisions based on Peter's data, but then they run into the manager, Art Howe, who will not play a data-chosen player. He is insubordinate when Billy commands him to play the player. Billy then makes a trade so that Art has to play Billy's choice. Thus, the arrival of data and analytics at the Oakland Athletics destroyed or diminished the experience-based competence of Grady and his fellow talent scouts. Today, almost all American baseball teams and European soccer teams use big data and analytics to some degree.

Such scenes will play out in many companies where an analog establishment is making product and marketing decisions based on years of experience and historical data. The digital newcomers will clash with these old pros and lose if the leadership, like Billy Beane, does not support them.

Chief Digital Officer

Another example can be found in risk management in financial services. Risk management departments were created along with a Chief Risk Officer (CRO). Those departments had very sophisticated, quantitative risk models. Just prior to the financial crisis, risk managers were waving red flags and trying to be heard. The bankers saw them as the "revenue reduction department." When the CEOs backed the bankers, the CROs had little impact. Today, nearly all banks have CROs reporting directly to the CEO and the role is staffed with a talented person. The effectiveness of the CROs and supporting regulation are still being tested. Where the leadership actively considers risk data and recommendations, the CRO and risk experts will be integrated into the decision process. At that point, power will have been shifted.

The question now is whether big data needs its equivalent of the CRO. There are quite a few proponents for a Chief Digital Officer (CDO). However, the role can take several forms. At P&G, it is the CIO who has played that role. The IT function has always worked with the businesses to introduce new information systems and analytical approaches. Given P&G's analytical orientation, a separate CDO role is probably not needed. The CIO can wear two hats: CIO and CDO.

Intel is using a partnership between the CIO and the CMO to take the lead in implementing big data analytics. Intel has a history of "two managers in a box." Starbucks is reported to have a full-time CDO. IBM has made the biggest change with an Enterprise Transformation Head. She has been in that role for 10 years, and her task has been to transform IBM to grid computing and now to the cloud and big data. IBM wants all of its processes to be converted so that it can be a model for its customers. The transformation head was a star line manager who ran the server and storage businesses before taking the new role. The CIO and process design activities all report to her. She, in turn, reports to the CEO. So, IBM has put a lot of power and authority behind its equivalent of a CDO.

It seems that the amount of power and authority of a CDO should be matched with the relative amount of difficulty and priority of implementing big data. If big data is a competence-enhancing innovation, a CIO wearing a double hat like P&G could be sufficient. If a company is at the other end of the continuum and big data is competence-destroying, more power and authority will be needed. At the destroying end, a role like IBM's Enterprise Transformation Head will be required to adopt big data.

A CDO role of some type is needed even in competence-enhancing companies. There are a number of corporate-led initiatives that are needed to embed the new analytics capability in the company's decision-making processes. First, companies need a strategy and plan. Where are the best opportunities for investment in big data? Companies also need to link applications with the requisite IT equipment and data architectures and, for the chosen applications, they must provide training and tools to frontline people. It is the usual strategic choice of "where to play." With limited resources, companies cannot do everything at once.

Second, someone has to lead the shift in corporate mind-set to one that data and information are an important company resource. Data is becoming a valuable resource like talent and money. Data and analytics groups are becoming like HR and finance – matrixed throughout the company. At a minimum, there is a corporate leadership group (the head of data and analytics) and several embedded groups (data analysts) in the businesses that report both to the business heads and to the corporate head of data and analytics or the CDO.

Third, the company must work to integrate and unite the many islands of data and analytics that exist throughout the organization. A lot of value comes from combining data from different sources both inside and outside the company. Resistance to sharing and combining data often arises depending on the strength of the organizational silos. Corporate leaders must create norms and values concerning information sharing, transparency, and trust. Each company tries to arrive at a situation where organizational units will have the data they need to execute their charter, but that data is also available to the rest of the company. In addition, those units should have reciprocal access to company data.

Finally, someone needs to resolve disputes caused by the new capability. As mentioned above, people are already charged with making decisions about advertising and new product features. Data and analytics will generate insights that lead to different decisions than does the current process. Those differences can lead to Moneyball-type clashes over data versus experience. The desired outcome is a blend of data and experience; a CDO is needed to see that disputes are settled with the right blend for the company. Other disputes such as issues around channel conflicts will arise. The digital technology enables disintermediation. The insurance agent was always a sacred cow at insurance companies. Now, young people are willing to buy insurance online and circumvent the agent. Inside the insurance companies, managers argue about whether and how to go about this new direct-to-consumer sale. The CDO needs to play a mediating role. In addition, jurisdictional disputes crop up between functions like the CIO, marketing, supply chain, and finance about who "owns" one digital activity or another. Again, the CDO needs to see that these disputes are discussed and resolved. For all of these reasons, someone must play the value-adding mediation role in all companies implementing big data.

Summary

Companies will need a power shift in their structures if they are to capitalize on big data analytics capability. The data and analytics newcomers need to be supported and integrated into the company's decision processes. If not, big data and the CDO will be like risk management and the CRO before the financial crisis. Grady and the scouts will not invite Peter to the decision-making meetings, and Art Howe will continue to be insubordinate and not play the right players. But, once a successful power shift is underway, the next step is to speed up decision making.

REAL-TIME DECISION MAKING

Another major big data challenge is to increase the speed of decision making. This is often referred to as increasing the "clock speed" of the organization. A computer has a clock which synchronizes the speed of the input unit, output unit, arithmetic unit, and memory unit. Historically, computer designers have been increasing the clock speed at which the computer operates. Similarly, organization designers need to increase their organizations' clock speed. Units such as advertising, customer management, new product development, and supply chain management have to synchronize around increasing clock speeds. The ultimate target is the making of decisions in real time.

Advertising

Today's advertising is transitioning from a "campaign" model to a "newsroom" model (Shetty & Wind, 2013). Traditionally, advertisers started their ad campaign planning in September for ads to be launched during the Super Bowl in February. They worked up a theme, shot many ads, narrowed them down to a few, bought time from the TV network, submitted their ads, and went to their Super Bowl parties to watch them. The media would report audience reactions to the ads on Monday morning following the game. The Nielsen ratings would arrive later that week on Friday. Coca-Cola and Audi went through this process for their 2013 Super Bowl ads, but unlike the other advertisers, they gathered in rapid response teams on the day of the Super Bowl. Then when their first ads were running they were looking at the Twitter feeds, the Facebook likes, and hits on their respective websites. Even before the first ad was finished playing, the teams were planning and making modifications to their second ads. Then the infamous power outage occurred. The brand teams that were already gathered in their "newsroom" jumped into action. The Audi team was quick to dig their rival, Mercedes Benz. The reason was that the blackout occurred in the Mercedes Benz USA Superdome stadium. Audi tweeted, "Sending some LEDs to the MBUSA Superdome right now", thereby plugging its own LED-laden vehicle. In order to respond in real time, these fast-response teams had to be supported by analytics to sift through all of the social media responses and make sense of them. Based on that data, the cross-functional team had to discuss insights, decide on a response, and post a tweet (an audio or video response). The blackout was indeed breaking news and the newsroom went into real-time action.

Another example of real-time advertising is the Old Spice ad for Red Zone Body Wash (Morrissey, 2010). The original ad was shown on traditional TV. It showed a very attractive man wearing only a towel around his waist applying the body wash. The response on Facebook and Twitter was way above normal, so the advertiser and the agency decided on a social media approach. They gathered a large team of writers, art directors, producers, editors, the actor, and social media specialists. They started with a couple of videos on YouTube and distributed them over Twitter. They targeted known influencers. The social media team then scoured the web for comments on the initial ads. They fed the funny comments, or those that came from interesting sources, to their "creatives" who turned them into humorous videos. The team was able to release several new videos per hour. The ads became a trending topic on Twitter. The promotion lasted for two days and the team created 200 videos, all in real time in response to topics coming from their viewers. This example shows the cross-functional newsroom advertising team in real-time action.

Community Management

Many companies have created Internet communities. People become members of the community when they register at the company's website or follow the company on Facebook, Twitter, or other social media sites. Communities are a valuable source of data for companies – but they become even more valuable when companies interact with their communities and do so in real time.

Nike is a good example (Piskorski & Johnson, 2012). It has always been active on the Internet. Nike launched its first website, nike.com, in 1998. It experimented with YouTube, MySpace, and other media sites. In 2005, it introduced NIKEiD, an online store. In addition to online direct sales, one of its features was that customers could design their own shoes, much like they could design their own computer on dell.com. Sales reached $100 million in a few years.

The big change came when Nike launched Nike+. Nike engineers became aware that virtually everyone was using iPods. They approached Apple about a partnership. The idea was to embed a sensor in Nike running shoes which could link wirelessly to an iPod and eventually to the iPhone and Android devices. The sensor could record distance, speed, and calories burned. The iPod could record these data and provide running music and other audio features. At the end of a run, the iPod could be uploaded to NikePlus.com and viewed and stored. The NikePlus.com website offered running tips, comparisons with other runners, shared workouts with friends, both on the Nike site and on Facebook and Twitter.

NikePlus.com built a following and in 2007 became the largest online running destination. Additional functionality was added to the website. For example, members could use the site to gather for group runs in many cities, or they could meet and gather after the runs. A

Nike+ GPS app was added to allow runners to map their runs. Then a database was built that now contains the largest collection of running maps around the world. One can ask for a suggestion for a steep course in Sonoma, CA where it is possible to take your dog. Next came the Nike+ GPS SportWatch. By 2013, there were more than seven million members of the Nike+ community.

Nike, like most consumer goods companies, has added a group of social network people to manage their communities (*The Economist*, 2013). Their responsibilities and management tools are expanding constantly. Some of the group's activities are to monitor the conversations on the web, create sub-communities, launch initiatives, and continuously manage the communities. This group runs like a newsroom; a 24/7 team monitors the conversations across all social media. They use analytics to do sentiment analysis to sense positive and negative sentiments about the brand. If there's a positive or negative spike, the team swings into action. When a factory collapses in Bangladesh, the team is quick to communicate that Nike manufactures no shoes in that country. If a positive spike occurs, the team investigates to see if it can be accentuated. In either case, the team explores what is behind the spike. They then say, "What can we learn from it? Should we act on it now? Should we pass on the learning to others?" Pop culture trends start unexpectedly from anywhere and spread quickly, like Gangnam Style. Which Nike shoes would the newsroom team suggest for Gangnam fashion?

Most companies believe that they have only scratched the surface in engaging with their communities. Nike discovered that dialogues with the community generate an enormous amount of data about running. This data can be analyzed and become the basis of new ideas for initiatives and products. The next step is to actively solicit ideas from the community through crowd-sourcing techniques. Indeed, Harley Davidson bypassed its ad agency and used crowd-sourcing to create the theme for a recent advertising campaign. Can crowd-funding be far behind (Winsor & Wind, 2013)?

New Product Development

Another function affected by big data is product development. As has Nike, General Electric and Bosch have embedded sensors in their products. Data from those sensors and other chips can be uploaded to the companies' websites. So when runners upload their data to NikePlus.com, it can be analyzed and compared to other runners. With the addition of GPS features, Nike can create many more applications that runners can access through NikePlus.com. As a result, Nike and these other companies find themselves in the software business. The software development process moves at Internet speeds, which approach real time.

The development of a new running shoe at Nike takes place over about an 18-month period. In contrast, the software development process is continuous and users co-develop the products. For example, LinkedIn and eBay launch new products and features several times a day. The new software development process, called the agile software development process, is a continuous and iterative cycle. The development team first creates a minimally featured product quickly and puts it on their company website. They invite users to try the products and report back on their experiences with it. The next day the team modifies the product based on the users' experiences. Through this continuous, iterative process, a new software product is created and available to users. The development teams expand and contract with the magnitude of the changes, but there is always a team evolving the product.

Another feature of software is that many companies are not just creating software products but rather software platforms. That is, Nike opens up the software code for NikePlus.com and makes it available to software developers. These developers then create running apps, which are accessible through NikePlus.com. Companies like Nike hire evangelists who go out and recruit software developers to write apps for Nike. For the good developers, Nike will share its vast data on running if they write apps exclusively for Nike. In January 2013, Nike partnered with a venture capitalist to create an incubator for startups that write software or create devices for running, which Nike can use for its customers and community (Banjo, 2013). So in addition to its user community, Nike is creating an ecosystem of device and software developers that will work with Nike and NikePlus.com to promote running and exercise. Nike engages continually with its communities of runners, developers, and suppliers to generate enormous amounts of data and ideas that can be analyzed for the shoes, software products, devices, or advertising ideas.

Organizing for Real-Time Decision Making

The question naturally arises about how to organize these real-time activities. There are a number of new units that must be integrated into the structure. Data and analytics talent must handle all of the incoming data and make sense of it. Software developers create new applications, and web designers regularly update the NikePlus.com site. Hardware engineers, who understand sensors and embedded chips, select and manage hardware vendors who make products like the SportWatch. Software evangelists recruit and manage partnerships with outside software vendors. Business managers run e-commerce websites, and social media experts manage the various communities. And, finally, digital marketing experts manage the process of taking real-time data to the analytics group, which produces real-time insights for decision makers, who then make real-time decisions. How should a company organize to implement its digital strategy that further differentiates its products and creates value?

One alternative is to integrate the software and hardware engineers into the product development function, and the digital marketing and social media experts into the marketing group. This alternative maintains the current functional structure and tends to be favored by the current managers. Another alternative is to combine all of the new talent into a digital unit and keep that digital unit intact. The company could add it as a new function in the business unit structure.

There are two arguments for a semi-independent unit. Operating independently, the unit can control its own activities and prove itself to the other units. As a new unit, it has to build its own capability and prove itself to others while earning credibility. In addition, a new unit always has a lot of trial and error until it discovers its success formula. Moreover, a new unit is fragile, and it needs independence as well as nurturing and developmental help from higher management. The second, related argument is that the unit is not just new; it is very different. It contains different specialists, each with their own language. But the real difference is the speed at which it has to operate. If it is a separate unit, it can operate at its own and faster pace. If it is embedded in other organizational units, it will have difficulty increasing its speed of decision making.

The new unit cannot be completely separate, however, because it is interdependent with the other functions. It must participate in the new product development process and pass ideas and information to Consumer Insights and Brand Advertising. As a result, the organization design must be more nuanced. The digital organization structure, along with Marketing, is shown in Figure 1.

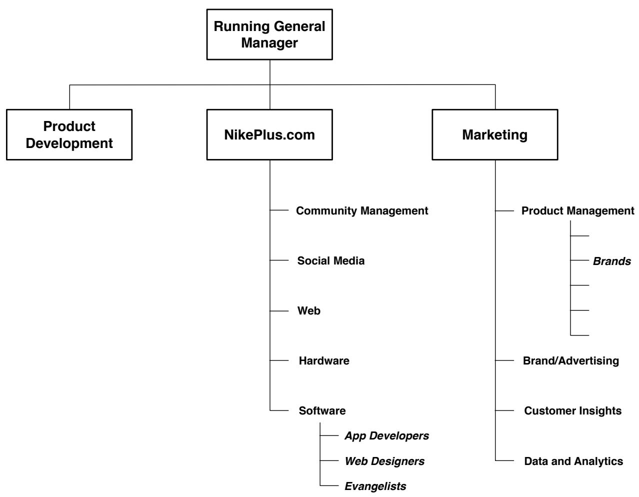

Fig. 1. Nike's Digital Functional Structure

The digital function consists of a community management unit, social media and mobile specialists, website management, hardware engineering, and a software function. These are all activities for which the digital unit is responsible. The units that operate in real time are the software, web, data analytics, and community management groups.

Another feature is needed to complete the design of the digital function. The governance feature is a steering committee chaired by the running business head. It consists of the heads of marketing, product development, sales, and digital. In the beginning, they meet weekly. The purpose of the committee is to see that all of the cross-functional linkages are working, and if not, to fix them. Once its initial growing pains are dealt with, the committee meets every other week or so to review plans, budgets, issues, and initiatives. This process links the new digital unit to the rest of the functions.

Supply Chain Management

The final function to be accelerated by big data is the supply chain. At P&G, the supply chain function meets in what is called the "control tower." Within the control tower, a cross-functional team meets in a special room called a "decision sphere" (see Figure 2). P&G has 42 decision spheres located throughout the company. These are specially designed rooms with video screens on the walls and computer access to various databases. The rooms are designed to foster real-time, cross-functional decision making. So when a paper machine's embedded sensors at the Pampers' plant in Wisconsin indicate that it requires maintenance, a plant shutdown is scheduled. If it looks like the machine will be down for a while, then the decision is made to supply Wal-Mart from the Albany, Georgia plant. The analytics capabilities are used to determine (and see) the best way to reroute trucks and still meet other delivery commitments to customers. This is an example of how big data facilitates real-time decisions in managing supply chains.

Fig. 2. A Procter & Gamble Decision Sphere

Summary

With the proper analytics capability, companies can make decisions in real time. They can involve their customers in dialogues about brands and gather ideas about new products and how to market them. Companies can use cross-functional teams that are in constant contact in a newsroom-like control tower and decision sphere to respond to real-time inputs. Such companies have increased the clock speed of their decision processes.

GENERATING REVENUE FROM BIG DATA

The third organization design feature triggered by big data is both a challenge and an opportunity. The changes in power and decision making described above will improve Nike's existing business – that is, Nike will sell more running shoes because of the added features of NikePlus.com. But the data, analytics, and insights can be revenue-producers themselves;

they have the potential to create entirely new businesses for Nike.

Companies such as Bosch and General Electric are putting sensors and microprocessors into all of their products. These companies anticipate that the services and software sales from the embedded devices will be a major source of growth in the coming years. Bosch has created a central unit, Bosch Software Innovation, which is to lead many of the new digital initiatives. One project is the design, installation, maintenance, and operation of an electric vehicle-charging infrastructure for Singapore. This huge project will provide software and data analysis for the government, retailers, fleet operators, utilities, and parking operators. The infrastructure will generate revenue from a range of services available through Bosch Software Innovation's Internet service platform.

In another example, large U.S. banks like JP Morgan and Wells Fargo provide reports on consumer trends to clients and other institutions that want to buy them. These banks have vast amounts of data from consumer use of credit and debit cards, checking accounts, ATM transactions, mortgages, and loans. The banks combine all of that data with publicly available data from governments. Then they apply their analytical capabilities to develop proprietary insights into consumer trends. In just a few hours, the banks can generate customized reports that slice the data into smaller and more narrowly defined market segments and geographies according to the specific demands of their clients. Such services generate increasing amounts of revenue and profits for the banks.

Citibank can provide this kind of data and insights on an international basis. It operates as a bank in 100 countries. That is, it takes deposits and makes loans in the local currency. The next largest bank is HSBC, which operates in 56 countries; no other banks are even close in size. So, Citibank has a competitive advantage that cannot be matched from a global perspective. It has the consumer data globally that Wells Fargo has nationally. Plus, it is one of the world's two largest foreign exchange providers. It is the largest cash management provider and number two in custody (securities safekeeping). That's in addition to its commercial lending operations in the 100 countries. Citibank can detect changes in trade patterns and economic conditions from an analysis of the basic (big) data. It says it can detect the new "silk roads" in emerging markets. Citibank sells its insights to companies like Zara and H&M to help them locate new stores and factories. Thus, banks see future growth coming from data and insights more than from their basic financial transactions businesses.

Nike is still the best example for the organizational design implications of big data. The Nike development history follows the Chandler (1962) model in which a new strategic emphasis is manifested in the organization's structure via a process he called "concatenation." Nike's recent strategy and structure transition show that big data is not just a new revenue source; it may be the start of a new organizational dimension. Let us briefly review the stages through which Nike has progressed.

Nike was founded in 1964. During the next several years, it became a fully functional, single-business company designing, manufacturing, marketing, and selling running shoes. The company went public in 1971. By the 1980s, two important developments occurred in its organizational structure. First, Nike diversified into other types of athletic footwear including basketball, tennis, soccer, and fitness shoes. It retained its functional structure but introduced lateral processes, such as cross-functional product teams, for the new types of footwear. Second, it expanded internationally by establishing sales, local marketing, and distribution subsidiaries. The organization became two-dimensional, with functions and regions reporting to the CEO. In the 1990s, it evolved into a three-dimensional organization by diversifying into sports apparel and sports equipment in addition to athletic footwear. Nike formed profit centers for footwear, apparel, and equipment, plus a separate business for golf. The supply chain, marketing, finance, and HR functions reported to the CEO in addition to the regions. In 2006, Nike made a big change to focus on customer groups. It created five categories: running, men's training, women's fitness, basketball, and soccer. Each customer category was a profit center and responsible for footwear, apparel, and equipment for its customer set. Nike thus created integrated solutions for customers who found them valuable. These categories were added to the previous structure, which still retained products, regions, and functions reporting to higher management. It is important to note that the categories were not just add-ons. The products were matrixed across categories. Women's fitness was still dependent on the apparel product line to determine fashion trends and help design this year's collection. All categories are still dependent on the latest footwear technologies. This change to categories, or customer segments, made Nike a four-dimensional organization. Then, in 2010, Nike created the Digital Sports Division.

The current Nike organization structure is shown in Figure 3. As indicated, Digital Sports is a profit center and has a number of responsibilities. First, it works with the categories in establishing Nike+ activities like the one in running described earlier. Nike has also added Nike+ Football, Nike+ Basketball, and Nike+ Kinect Training within the various category profit centers. Digital Sports supports the categories in creating and managing their communities and with new products like the GPS app. It is an example of another dimension being added to the Nike matrix, a manifestation of the strategic emphasis being placed on big data. Second, the division is developing its own products and its own sources of revenue. Nike+ FuelBand is a wristband that records distance traveled in a day, the number of steps taken, and calories burned. The stored results can be uploaded at the end of the day and compared with others in the user's age group. The division has its Nikefuel.com website, apps, apparel, and community. It has recently introduced a wristwatch with similar features. Now, Nike is competing in the wearable medical device market. More products are on the way. Third, the division takes the lead in building the Nike incubator described earlier. Thus, it is responsible for establishing an ecosystem of companies around the devices.

Fig. 3. Nike's Organization Structure (July 2013)

At the moment, the Digital Sports Division is small, but it may grow into a fifth dimension of strategy and organization. Digital Sports is not just an added product line; it is responsible for building digital capabilities in the other Nike units as well as growing its own revenue. It is similar to the Disney Interactive Division and the Bosch Software Innovation group. These companies are all following similar models of strategy and organization with respect to big data capability.

USING THE STAR MODEL TO ILLUSTRATE THE IMPACT OF BIG DATA

To summarize the impact big data has on an organization, I will use the Star Model (Kates & Galbraith, 2010) shown in Figure 4. I will describe how big data impacts each major element of an organization: strategy, people, structure, rewards, and processes.

Figure 4 shows that companies are adopting a dual strategy for implementing big data analytics capability. The first is to build a digital capability to make better and faster decisions, and to enhance existing products. Disney has developed a digital capability called Watch. Customers can sign up for Watch through their cable company. It will allow a viewer to watch any program from ESPN, the Disney Channel, and ABC on any device. The purpose is to sell more advertising on these Disney channels. The second strategy is to use data and analytics to create insights and custom reports that can be sold to customers and become a new profit center. Bosch, for example, supplies electrical components on many automobiles. It can take data from anti-rollover software, engine controls, anti-lock braking systems, and other sensors, analyze them and offer insights to drivers about safety, maintenance, and so on. It can offer these insights to the car owners directly or through the automobile dealer.

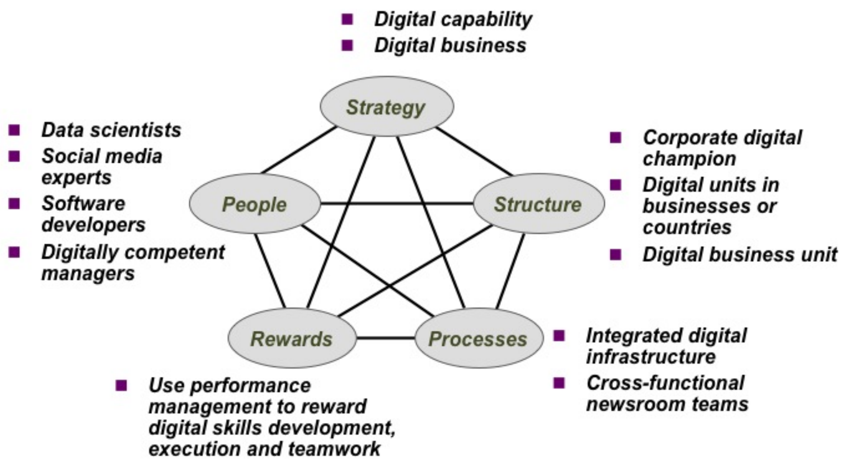

Fig. 4. Big Data's Impact on the Organization

In order to implement these strategies, companies must, of course, modify their organizations. Let's walk through the various changes that a company will have to make when it pursues a big data strategy. First, the company will need a champion for data and analytics on its leadership team. Whether it is a Chief Digital Officer (CDO), the CIO, or a digital division head, the digital leader needs to promote data as a strategic asset that can help grow the company and make it more successful. In addition, each business unit, customer segment, and country will need a digital unit to support its activities. Those digital units will report to both their respective profit center heads and the Chief Digital Officer. Such structural changes will constitute a shift in power to the digital units, and the company will need to rebalance its power structure and prepare for faster decision making. Finally, there will be a digital business unit (or group of business units) that will earn revenue and become a profit center. A new profit center and the subsequent matrix of digital capabilities will constitute a new dimension of organizational structure, much like functions, businesses, customer segments, and regions.

The aspiring big data company will also need to create information and decision processes to support the structure and strategy. First, the company will need to harness its information infrastructure to combine its various databases. For example, a bank will need to combine data from credit and debit cards, transaction accounts, mortgages, home equity loans, investments, and so on to paint a complete picture of a customer. Using this information, the bank can combine the data, perform analyses, and generate insights about financial trends among retirees, high net worth individuals, and Hispanic families with young children. Teams with representatives from the relevant product lines and functions (such as marketing and risk) can combine the bank's internal data with social network, customer relationship management, and other types of data and process them in real time. These teams are the equivalent of newsrooms or digital acceleration teams (DATs) as used by Nestlé. In all cases, the teams act on real-time insights to execute in real time.

In order to execute real-time decision processes, the organization needs the people and talent who are skilled in digital tools and who work effectively in teams. The human resource practices at Procter & Gamble are good examples. P&G has shifted its hiring practices to bring in digitally skilled experts, software developers, social media experts, and managers who are comfortable with quantitative decision practices. Each manager's progress is evaluated through the performance management system. In this way, the company can think holistically about the transition to the digital enterprise. The organizational change is not just deciding whether the company needs a CDO or not. It requires changes in all elements of the Star Model, which are aligned not only with each other but with the company's environment as well.

CONCLUSION

Big data is often portrayed as a potential opportunity for business firms and other types of organizations. Firms skilled in the use of big data, however, are already using their analytics capability to create strong competitive advantages. Those firms are now making many important decisions in real time and are able to keep pace with the rapidly changing environments of the digital age. Their experiences can be instructive for other interested organizations because the organizational, managerial, and cultural changes required by a big data analytics capability are considerable.

REFERENCES

Banjo S. 2013 (June 19). Nike takes a page from Apple's playbook. *The Wall Street Journal*, Technology Section.

Chandler AD. 1962. *Strategy and Structure*. MIT Press, Cambridge, MA.

Galbraith JR. 2014. *Designing Organizations*. 3rd edition. Jossey-Bass, San Francisco, CA.

Kates A, Galbraith JR. 2010. *Designing Your Organization: Using the Star Model to Solve 5 Critical Design Challenges*. Jossey-Bass, San Francisco, CA.

Morozov E. 2011. *The Net Delusion: The Dark Side of Internet Freedom*. Penguin, London, U.K.

Morrissey B. 2010 (July 15). How Old Spice ruled the real time web. *AdWeek*.

Piskorski M, Johnson R. 2012. Social strategy at Nike. Harvard Business School Case 9-712-484.

Schmidt E, Cohen J. 2013. *The New Digital Age: Reshaping the Future of People, Nations and Business*. Knopf, New York, NY.

Shetty B, Wind J. 2013 (February 15). Advertisers should act more like newsrooms. Harvard Business Review Blog Network.

The Economist. 2013. Less guff, more puff. May 18: 69-70.

Tushman M, Anderson P. 1986. Technological discontinuities and organizational environments. *Administrative Science Quarterly* 31: 439-465.

Winsor J, Wind J. 2013 (May 9). The end of traditional ad agencies. Harvard Business Review Blog Network.

JAY R. GALBRAITH

Professor Emeritus
IMD, Lausanne, Switzerland
E-mail: jay@jaygalbraith.com

THE INFORMATION PANOPTICON IN THE BIG DATA ERA

MARTIN BERNER • ENRICO GRAUPNER • ALEXANDER MAEDCHE

Abstract: Taking advantage of big data opportunities is challenging for traditional organizations. In this article, we take a panoptic view of big data – obtaining information from many sources and making it visible to all organizational levels. We suggest that big data requires the transformation from command and control hierarchies to post-bureaucratic organizational structures wherein employees at all levels can be empowered while simultaneously being controlled. We derive propositions that show how to best exploit big data technologies in organizations.

Keywords: Big data, organization design, information visibility, visibility paradox, control, empowerment

In her seminal book, *In the Age of the Smart Machine*, Zuboff (1988) observed that information systems not only automate various organizational processes; they also have the unique capacity of "producing" information by making activities, events, and objects more visible. Organizations with high information visibility are called *informated* organizations, and they "…operate very differently from the traditional assumption of imperative control" (Zuboff, 1988: 411). To be effective, informated organizations have to increase the information-processing capacities of their hierarchies or decentralize through lateral coordination mechanisms (Galbraith, 2012).

Recent technological advances offer huge opportunities to enhance the level of information visibility in organizations by leveraging big data. On the one hand, increased information visibility may help to empower employees. On the other hand, visibility can increase the surveillance of individuals which may have negative effects on employees' job satisfaction, performance, and motivation (Ball, 2010). In this article, we discuss information visibility and derive its implications for organizational behavior and design. First, we broadly describe information visibility in the era of big data and introduce the information visibility paradox. Next, we derive the implications of information visibility in the form of propositions about how big data can be exploited. Lastly, we close with some concluding remarks about our intended contribution and future research.

INFORMATION VISIBILITY IN THE ERA OF BIG DATA

"In the age of big data … the emphasis in industry has shifted to data analysis and rapid business decision making based on huge volumes of information" (Chen, Chiang, & Storey, 2012: 1182). Besides enabling more informed decisions, big data can provide value to organizations by offering new insights and automating business processes (Laney, LeHong, & Lapkin, 2013). Big data is associated with increased data volume, velocity, and variety but decreased data veracity (Schroeck et al., 2012). Accordingly, the size of datasets exceeds the abilities of many organizations in terms of capturing, storing, managing, and analyzing data (Manyika et al., 2011). For example, over 1.8 zettabytes (which translates to 1.8 trillion gigabytes) of data were created in 2011 (Gantz & Reinsel, 2011). The vast majority of this data still comes from processes inside organizations and not from external data sources (Schroeck

Journal of Organization Design
JOD, 3(1): 14-19 (2014)
DOI: 10.7146/jod.9736

et al., 2012). One reputable study suggests that the total data volume in organizations doubles every 18 months (Forrester Research, 2010). In terms of velocity, big data can be used for real-time decision making. Moreover, the availability of more and faster data means that its veracity becomes harder to determine.

Such changes in data characteristics call for corresponding changes in information processing and decision making. As data become cheaper, the complements to data become more valuable (McAfee & Brynjolfsson, 2012). Accordingly, new complementary analytics have evolved that ultimately establish new degrees of information visibility. Those analytics are based on new parallel technologies, such as MapReduce and Hadoop, that process large amounts of data at low cost. Further, big data analytics exploit new technologies, like in-memory and columnar stores, that analyze huge data sets in real-time. Such developments mark the beginning of big data use in day-to-day operations. Previously, access to information-centric systems like data warehouses was reserved for strategic decision support, while process-centric systems, such as Enterprise Resource Planning (ERP), support daily operations. System separation was introduced in the 1990s to allow the analysis of large datasets with good response performance. Recently, however, big data-related technologies like in-memory databases make system separation obsolete as "…transactional and decision-related data is managed in an integrative manner" (Loos et al., 2011: 394). Additionally, workers at the operational level have access to virtually unlimited external information via the Internet. Therefore, information visibility at the operational level increased significantly with the onset of the big data era. Organizations which provide analytical information to their operational decision makers perform better than those without analytical information provisioning at the operational level (Lock, 2010).

Digital companies such as Amazon have disrupted industries with new data-driven business models. Big data companies will also change traditional business as they can make more accurate predictions, better decisions, and precise interventions instead of relying on experience and intuition (McAfee & Brynjolfsson, 2012). Empirical research confirms the advantages of data-driven decision making and has identified a positive association with firm performance (Brynjolfsson, Hitt, & Kim, 2011). In summary, big data will trigger numerous changes in organizational decision making and design. It increases information visibility that enables data-driven decision making at both the strategic and operational levels (see Table 1).

Table 1. Comparison of Past and Future Information Visibility

Characteristic	Smart Machine Era	Big Data Era
Information Timeliness	Historical data	Real-time data
Information Sources	Self-created, high-quality datasets	Large amount of data including unreliable external datasets
Information Reach	Strategic level	Strategic and operational levels
Information Relevance for Decision Making	Low (experience-driven decision making)	High (data-driven decision making)

THE INFORMATION VISIBILITY PARADOX

Information visibility has paradoxical characteristics. On the one hand, greater visibility "… serves as a means of empowering … to make decisions which used to be formally referred upwards or to other departments" (Sia et al., 2002: 24). On the other hand, visibility offers new possibilities to keep subordinates as well as peers under surveillance and even to infringe on their personal privacy. Studies about this phenomenon are often based on the metaphor and theoretical lens of the *information panopticon* (Elmes, Strong, & Volkoff, 2005; Sia & Neo, 2008; Sia et al., 2002; Zuboff, 1988). The panopticon is a special design of an early nineteenth-century prison. An observation tower in the middle of a circular prison enables guards to view every cell. This creates "… a state of conscious and permanent visibility that assures the automatic function of power", where people behave as if they are under constant control (Foucault, 1979: 201). The psychological effects of such visibility are also evident in the context of information systems, where the knowledge that information is potentially visible for others likewise induces self-control (Zuboff, 1988).

The panopticon building creates hierarchical visibility for the guards, but the information panopticon in an organization offers horizontal visibility for peers and even subordinates. In the past, separate information-centric systems provided visibility mainly for management. In this regard, old information technologies increased bureaucracy by establishing more rationality and control along the organizational hierarchy (Weber, 1922). In contrast, horizontal visibility can empower workers and brings new opportunities for the creation of post-bureaucratic organizations (see Figure 1). Horizontal visibility is enabled by the blurring of information-centric and process-centric system boundaries as well as the accessibility of big data at the operational level. As Zammuto et al. (2007: 752) state, "[Integrated] enterprise systems decreased the need to move information through a hierarchy, allowing people to organize around the work itself… Everyone working on a particular process… could now use the process-based IT system to see and understand the whole work flow." In other words, empowerment requires that workers – and not only management – have an appropriate level of visibility.

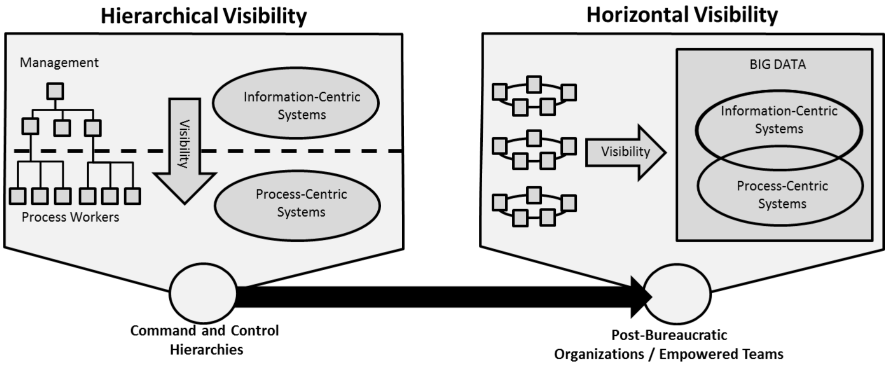

Fig. 1. Comparison of Hierarchical and Horizontal Visibility

Elmes, Strong, & Volkoff (2005: 29) found that employees perceived "… visibility of information as empowering and took action based on this additional information." Psychological empowerment is defined as any increase in worker power that enables them to achieve organizational objectives (Seibert, Silver, & Randolph, 2004). It is specified as an individual-level motivational state shaped by the work environment and manifested in a set of four cognitions: meaning, competence, self-determination, and impact (Spreitzer, 1995; Thomas & Velthouse, 1990). Work context antecedents of psychological empowerment include access to information about unit performance and strategy (Spreitzer, 1995; Thomas & Velthouse, 1990). Empirical research has found a strong positive relationship between employees' access to information and their feeling of psychological empowerment (Laschinger et al., 2004; Spreitzer, 1995). The existing literature also states that employees become more effective, innovative, and satisfied with increased levels of empowerment (Spreitzer, 1995; Spreitzer, Kizilos, & Nason, 1997). This confirms earlier observations that information access rules are an IT-enabled constitution of power (Orlikowski & Robey, 1991; Zuboff, 1988).

But information visibility not only empowers workers, it also controls them. Theoretically grounded in the information panopticon, Sia et al. (2002) call information systems that enable the surveillance of employees "panoptic control." Compared to enterprise resource planning (ERP) systems, which information panopticon research has studied so far (Elmes et al., 2005; Sia et al., 2002), big data technologies have higher tracking and information visibility capabilities.

Some empirical research has shown that information visibility can increase panoptic control and empowerment simultaneously (Elmes et al., 2005; Sia & Neo, 2008; Sia et al., 2002). Thus, information visibility is a two-edged sword that enables employees to see and do more, while it makes them more visible to others at the same time (Elmes et al., 2005).

LEVERAGING INFORMATION VISIBILITY

Panoptic control does not necessarily have a negative impact on employees. Employees expect performance assessments based on information about their activities, but dysfunctions such as resistance or sabotage arise if surveillance goes beyond what is reasonable or necessary (Ball, 2010). Control dysfunctions are behaviors that are "inconsistent with the best interests of the organization", potentially resulting in lower job satisfaction and individual performance (Jaworski, 1988: 23). Therefore, it depends on the perceived adequacy of panoptic control whether it triggers dysfunctional behavior or intended self-control. Correspondingly, the growing big data literature recommends that organizations have to implement privacy protection rules and define data ownership in order to reap the benefits of big data (Lund et al., 2013). A case study underlines this: A company that implemented sophisticated real-time performance dashboards was not able to gain process improvements because its workforce showed a strong resistance to the measurement initiatives in fear of complete transparency of individual performance (Cleven, Winter, & Wortmann, 2011). Increased transparency from new technologies requires adequate protection of private data from superiors and, since horizontal visibility is increasing, from co-workers. To realize benefits from the usage of big data, we propose that organizations consider the following:

Proposition 1: Control dysfunctions of big data usage are lower if privacy protection and data ownership rules are implemented.

Empirical studies confirm that psychological empowerment is significantly and positively related to individual performance and job satisfaction (Seibert et al., 2004; Spreitzer et al., 1997). However, the benefits of psychological empowerment are moderated by different factors. For example, the less structured the work context is, the higher the positive influence of psychological empowerment (Thomas & Tymon, 1994). Also, the empowering effects of business process re-engineering are greater for knowledge-intensive processes than for standardized processes where the disciplinary effects of visibility are more dominant (Sia & Neo, 2008). Accordingly, we conclude the following for big data usage in organizations:

Proposition 2: The psychological empowerment benefits of big data usage are higher for knowledge-intensive processes than for standardized processes.

Information panopticon research identified that the empowering outcome of information systems implementation depends on "… clear management intentions to break away from pre-existing structures" (Sia et al., 2002: 35). An empowerment climate is positively related to individuals' perceived psychological empowerment and includes open information sharing, employee autonomy, and team accountability as key organizational practices (Seibert et al., 2004). Furthermore, Davenport & Beers (1995: 74) conclude that "… giving line workers… information in real time without empowering them to act on it is, at best, wasteful, and, at worst, harmful." Thus, we propose:

Proposition 3: The psychological empowerment benefits of big data usage are higher for organizations with a strong empowerment climate.

An important dimension of psychological empowerment is the competence of employees (Spreitzer, 1995). Research indicates the need for training a large number of so-called data scientists in order to realize the potential of big data (Lund et al., 2013; Schroeck et al., 2012). Although we share the call for more data scientists, the switch from strategic to operational real-time use of big data calls for analytical skills development within existing operational roles as well as more decision competency. This leads to our final proposition:

Proposition 4: The psychological empowerment effects of big data usage are higher if analytical skills and decision competencies of operational employees are increased.

CONCLUSION

Our conceptual article links big data to information visibility, describes how visibility affects employee control and empowerment, and provides propositions that help organizations to exploit big data opportunities. Organizations can enhance information visibility in decision making by appropriate big data analytics. We use a panoptic lens to explore the control and empowerment paradox associated with big data in an organizational context. Horizontal information visibility – particularly for employees at the operational level – requires changes from hierarchical structures to post-bureaucratic structures, as less information moves through the hierarchy and workers can organize instead around the work itself. The propositions that we offer identify key variables in information processing and decision making, and they suggest research which could unlock the potential of big data in organizations.

REFERENCES

Ball K. 2010. Workplace surveillance: An overview. *Labor History* 51(1): 87–106.

Brynjolfsson E, Hitt LM, Kim HH. 2011. Strength in numbers: How does data-driven decision making affect firm performance? *SSRN Electronic Journal*. Accessed December 10, 2013: http://www.ssrn.com/abstract=1819486.

Chen H, Chiang RHL, Storey VC. 2012. Business intelligence and analytics: From big data to big impact. *MIS Quarterly* 36(4): 1165–1188.

Cleven A, Winter R, Wortmann F. 2011. Process performance management: Illuminating design issues through a systematic problem analysis. *Proceedings of the 2011 ACM Symposium on Applied Computing*: 280–286.

Davenport TH, Beers MC. 1995. Managing information about processes. *Journal of Management Information Systems* 12(1): 57–80.

Elmes MB., Strong DM, Volkoff O. 2005. Panoptic empowerment and reflective conformity in enterprise systems-enabled organizations. *Information and Organization* 15(1): 1–37.

Forrester Research. 2010. Forrester TechRadar: Enterprise data integration. Accessed December 10, 2013: http://www.forrester.com/Forrester+TechRadar+Enterprise+Data+Integration+Q1+2010/fulltext/-/E-RES56028.

Foucault M. 1979. *Discipline and Punish: The Birth of the Prison*. Peregrine Books, London, U.K.

Galbraith JR. 2012. The future of organization design. *Journal of Organization Design* 1(1): 3–6.

Gantz BJ, Reinsel D. 2011. Extracting value from chaos state of the universe. Accessed December 10, 2013: http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf.

Jaworski BJ. 1988. Toward a theory of marketing control: Environmental context, control types, and consequences. *Journal of Marketing* 52(3): 23–39.

Laney D, LeHong H, Lapkin A. 2013. What big data means for business. *Financial Times*. Accessed December 10, 2013: http://www.ft.com/intl/cms/s/0/b1dec7f4-b686-11e2-93ba-00144feabdc0.html#axzz2gBZKTk4m.

Laschinger HKS, Finegan JE, Shamian J, Wilk P. 2004. A longitudinal analysis of the impact of workplace empowerment on work satisfaction. *Journal of Organizational Behavior* 25: 527–545.

Lock M. 2010. Operational intelligence: Boosting performance with "right-time" business insight. Accessed December 10, 2013: http://www.aberdeen.com/aberdeen-library/6528/RA-operational-business-intelligence.aspx.

Loos P, Lechtenbörger J, Vossen G, Zeier A, Krüger J, Müller J, Lehner W, Kossmann D, Fabian B, Günther O, Winter R. 2011. In-memory databases in business information systems. *Business & Information Systems Engineering* 3(6): 389–395.

Lund S, Manyika J, Nyquist S, Mendonca L, Ramaswamy S. 2013. *Game Changers: Five Opportunities for US Growth and Renewal*. McKinsey Global Institute, San Francisco, CA.

Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers HA. 2011. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, San Francisco, CA.

McAfee A, Brynjolfsson E. 2012. Big data: The management revolution. *Harvard Business Review* October 2012: 1–9.

Orlikowski WJ, Robey D. 1991. Information technology and the structuring of organizations. *Information Systems Research* 2(2): 143–169.

Schroeck M, Shockley R, Smart J, Romero-Morales D, Tufano P. 2012. Analytics: The real-world use of big data. Accessed December 10, 2013: http://www-935.ibm.com/services/

us/gbs/thoughtleadership/ibv-big-data-at-work.html.
Seibert S, Silver S, Randolph W. 2004. Taking empowerment to the next level: A multiple-level model of empowerment, performance, and satisfaction. *Academy of Management Journal* 47(3): 332–349.
Sia SK, Neo BS. 2008. Business process reengineering, empowerment and work monitoring: An empirical analysis through the panopticon. *Business Process Management Journal* 14(5): 609–628.
Sia SK, Tang M, Soh C, Boh WF. 2002. Enterprise resource planning (ERP) systems as a technology of power: Empowerment or panoptic control? *The DATA BASE for Advances in Information Systems* 33(1): 23–37.
Spreitzer GM. 1995. Psychological empowerment in the workplace: Dimensions, measurement and validation. *Academy of Management Journal* 38(5): 1442-1465.
Spreitzer GM, Kizilos MA, Nason SW. 1997. A dimensional analysis of the relationship between psychological empowerment and effectiveness, satisfaction, and strain. *Journal of Management* 23(5): 679–704.
Thomas K, Tymon W. 1994. Does empowerment always work: Understanding the role of intrinsic motivation and personal interpretation. *Journal of Management Systems* 6(3): 1-13.
Thomas K, Velthouse B. 1990. Cognitive elements of empowerment: An "interpretive" model of intrinsic task motivation. *Academy of Management Review* 15(4): 666–681.
Weber M. 1922. *Wirtschaft und Gesellschaft - Grundriß der Verstehenden Soziologie*. Mohr, Tübingen, Germany.
Zammuto RF, Griffith TL, Majchrzak A, Dougherty DJ, Faraj S. 2007. Information technology and the changing fabric of organization. *Organization Science* 18(5): 749–762.
Zuboff S. 1988. *In the Age of the Smart Machine: The Future of Work and Power*. Basic Books, New York, NY.

MARTIN BERNER

Institute for Enterprise Systems
University of Mannheim
E-mail: berner@es.uni-mannheim.de

ENRICO GRAUPNER

Institute for Enterprise Systems
University of Mannheim
E-mail: graupner@es.uni-mannheim.de

ALEXANDER MAEDCHE

Institute for Enterprise Systems
University of Mannheim
E-mail: maedche@es.uni-mannheim.de

ORGANIZATIONAL MODELS FOR BIG DATA AND ANALYTICS

ROBERT L. GROSSMAN • KEVIN P. SIEGEL

Abstract: In this article, we introduce a framework for determining how analytics capability should be distributed within an organization. Our framework stresses the importance of building a critical mass of analytics staff, centralizing or decentralizing the analytics staff to support business processes, and establishing an analytics governance structure to ensure that analytics processes are supported by the organization as a whole.

Keywords: Organizational structures for analytics, big data, analytic governance, organizing data scientists

There is little debate these days about the importance of big data and analytics in supporting the strategic goals of an organization (Davenport, 2006; Manyika, et al., 2011), but there is as yet no consensus about how best to organize analytics efforts within the organization and what core analytics processes the organization must support. In this article, we introduce a framework that breaks big data and analytics into several processes and shows how those processes fit within the organization, and we discuss how an appropriate analytics governance structure can enable an organization to extract business value and competitive advantage from big data.

Following Laney (2001), we consider *big data* as data whose volume, velocity, and variety make it difficult for an organization to manage, analyze, and extract value using current or conventional methods and systems. We use the term *analytics* as the process that extracts value from data through creating and distributing reports, building and deploying statistical and data-mining models, exploring and visualizing data, sense-making, and other related techniques. Data may be internal or external to the organization; processing may be real-time, near real-time, or batch; and any combination of these is possible.

A FRAMEWORK FOR ORGANIZING ANALYTICS

Our organizational framework seeks to integrate analytics, business knowledge, and information technology (see Figure 1), and it is based on four main questions:

1. Does the organization view data and analytics as a key function of the organization, similar to the way that finance, information technology, sales and marketing, product development, etc. are viewed as functions of the organization? Analytics must be perceived as valuable to the business units in order for it to be integrated into operations.
2. Is there a critical mass of data scientists? Without a critical mass of data scientists, there is insufficient domain knowledge to address all the problems of interest. Also, there is not deep enough knowledge of the analytics infrastructure to obtain or create the needed data and to manage the data that is obtained. Finally, there may not be deep enough knowledge to deploy statistical and data-mining models in operations.
3. Are there data scientists with sufficiently deep knowledge of the business unit domains? Without such knowledge, it is difficult to build models that bring value to the business unit. Deep knowledge and complex business problems tend to spawn specialization. It is important for an analytics group to include a mixture of data scientists, some of whom are generalists and others who are specialists.

Journal of Organization Design
JOD, 3(1): 20-25 (2014)
DOI: 10.7146/jod.9799

4. Is there an adequate analytics governance structure? A governance structure helps stakeholders make decisions that prioritize big data opportunities, obtain the required data, deploy analytical models, and support measurement of the business impact of the models.

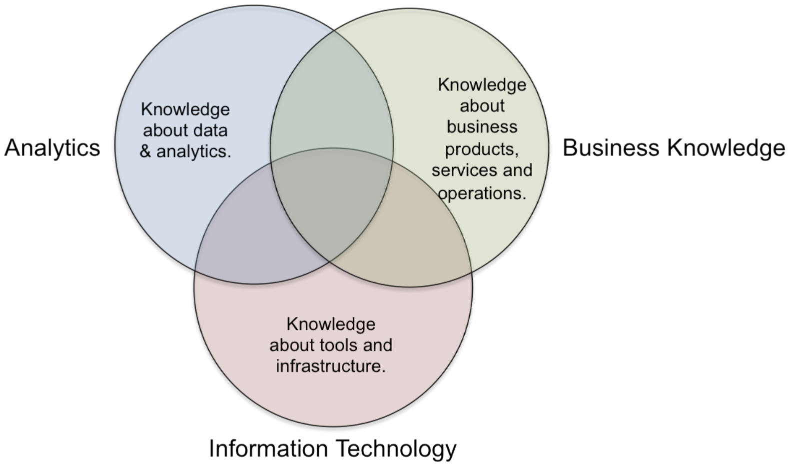

Fig. 1. The knowledge required by data scientists

We call our framework the CSPG framework – for analytics Culture, Staffing, Processes, and Governance (see Table 1). The CSPG framework orients the organization designer to establishing a culture for big data and analytics; hiring, training, and organizing a group of analytics staff; developing the required analytics processes; and setting up a robust analytics governance structure. Starting with culture, corporate-level executives must recognize the need to organize big data and analytics as an organizational function that is given broad responsibility and authority for data assets and which is analogous to other major functions in the organization. The analytics leader has the responsibility for hiring and managing the best data scientists, ensuring that the appropriate analytics opportunities are identified and explored, acquiring the appropriate internal and external data, and setting up and operating the analytics governance structure.

Table 1. A summary of the CSPG Framework

	Analytics at the Department/ Unit Level	Analytics at the Organizational Level
Analytics Culture	Are big data and analytics viewed as an organizational function and is there a big data/analytics department or unit to support this function?	Are big data and analytics integrated into corporate strategy? Is there a senior leader advocating for big data and analytics? If not, put a senior leader in charge of big data and analytics with this charge. Is data (both internal and external) that can provide value being used?
Analytics Staff	Does the analytics department have the right people with the right degree of analytic specialization, IT knowledge, and business knowledge?	Are there analytic team members in the right departments within the organization and is there a critical mass of analytic talent? If not, rebalance the analytic staff or change the centralization/decentralization of the analytics staff as required.
Analytics Processes	Does the analytic department have analytic processes in place to build analytic models, deploy analytic models, and measure their business impact?	Does the organization have the analytic processes in place to select analytic opportunities, provide data to the data scientists, build analytic models, deploy analytic models, and measure the business value generated? Is there an analytic governance structure in place to support and to coordinate the correct analytic processes?

The CSPG framework requires that there be a critical mass of analytics staff (data scientists). Analytics staff must be able to obtain and manage data; build statistical, predictive, and data-mining models; and deploy those models. The analytics leader, along with corporate management, must decide on where to locate the analytics function within the organization (discussed in the next section). Essentially, the analytics staff can be centralized or decentralized, with hybrid approaches available as well.

The third component of the CSPG framework concerns the analytics processes themselves. Big data presents many opportunities if those processes can be properly created and managed. Data can be traded among organizations, products can be augmented to produce data, assets can be digitized, data can be combined within and across industries, and so on (Parmar, Cohn, & Marshall, 2014). The more sophisticated the analytics processes become, the more opportunities that can be pursued. The organizational aspects of analytics processes are discussed below.

The final component of the CSPG framework is analytics governance. Because big data and analytics are new to many organizations, analytics governance structures are not well defined. Senior corporate leaders are responsible for setting up the governance structure, and they are responsible for monitoring and improving it as experience accumulates. Analytics governance structure is discussed below.

Broadly speaking, the CSPG framework presented here can be thought of as an application of the Star Model (Galbraith, 2008) to the analytics function. The design of the analytics function must be complete in the sense that it covers people, structure, rewards, and so on, and each component of the analytics function must be aligned with the others and with the larger corporate organization.

LOCATION OF THE ANALYTICS FUNCTION WITHIN THE ORGANIZATION

There are three basic models for locating the analytics function within the organization, all of which involve well-known tradeoffs between centralization and decentralization. One model centralizes analytics by placing the data scientists in a single unit. This model is the easiest way to achieve critical mass, obtain necessary data, drive an integrated infrastructure, and gain the required expertise to efficiently test and deploy various statistical, predictive, and data-mining models. When analytics is centralized, however, the data scientists may be far away from the business units they are supposed to support. The challenge in such a structure is for the data scientists to understand the various business units and their needs. In addition, there is the issue of where the analytics department should report within the organization. Should it report to a functional area such as finance, IT, R&D, or marketing, or should it report to the very top of the organization?

A second organizational approach is to decentralize analytics and place a group of data scientists in each business unit. This approach makes it easier for data scientists to collaborate with their respective business units and to tailor their models to each unit's needs. The main tradeoff is difficulty in achieving critical mass on enterprise-wide problems and opportunities. A closely related question is whether each group has the expertise required to create datasets and to deploy analytical models.

The third model is a hybrid approach in which a critical mass of data scientists is housed in a central unit, and the remaining data scientists are distributed throughout the organization. One common hybrid model is to set up an analytics or big data "center of excellence" that the distributed data scientists can draw on as appropriate. Another is to centralize the data scientists that interact with the IT organization, or those that manage the data, or those that deploy the models.

None of these three models provides a perfect organizational solution; each involves tradeoffs. From a design perspective, managers must recognize the tradeoffs associated with each model and make their location choice accordingly.

ANALYTICS PROCESSES

Generally speaking, the analytics function is composed of analytics models, analytics infrastructure, and analytics operations. Analytics models are statistical, predictive, or data-mining models that are empirically derived from data using generally accepted statistical methods. A key analytics process is building models. This is usually done by statisticians, modelers, or, to use the new name, data scientists (Press, 2013). As discussed above, data scientists may be located within a single department or group, attached to business units, or a combination of both. If the data scientists are centralized in a single unit, it is often called the analytics department. In addition to building models over data, analytics also includes summarizing data in reports (now called descriptive analytics), ad hoc querying of data, exploring data with visualizations, sense-making, and other techniques.

Analytics infrastructure refers to the software components, software services, applications, and platforms for managing data, processing data, producing models, and using models to generate alerts, take actions, and make decisions (Grossman, 2009). The key processes associated with analytics infrastructure are managing the data required by the organization and deploying the models and other analytics that are incorporated into the organization's products, services, and operations. It is becoming more common to use computer languages for describing analytics (Data Mining Group, 2012) so that analytics can be more easily deployed. These processes involving analytics infrastructure are usually performed by the information technology (IT) organization.

Analytics operations refers to the various processes that result in the outputs of models being used to make decisions and to take actions that bring business value. Analytics operations ensures that the results of models are integrated into an organization's products, services, and operations. In an analytics department, data scientists identify the data needed, acquire the data, work collaboratively with business units to build models, and then work with the IT group to deploy the models into the organization's operations. Data can be a combination of data internal to the organization, collected by the organization, or purchased by the organization. The IT department is normally involved if data is generated by the organization or collected by the organization.

An organization requires a critical mass of data scientists so that their expertise as a whole extends across these three analytics processes. The team as a whole must be able to: identify relevant data (both internal and external), manage the data required for analytics, build the needed analytics models, and deploy the models that are built into products, services, and internal systems. Multiple parts of an organization can be involved in analytics processes. Typically, a business unit sponsors the model, an analytics department builds the model, an IT unit supplies and manages the data, and an operations unit deploys the model. With so many diverse pieces of an organization involved, an analytics governance structure is critical.

ANALYTICS GOVERNANCE STRUCTURE

There are three main challenges that organizations face when trying to extract value from big data using analytics.

1. *Identifying and resourcing analytics opportunities.* The first challenge is identifying which analytics opportunities to pursue, building the business case for those opportunities, and obtaining the required resources. Analytics opportunities belong to stakeholders within the various business units and functional areas of the organization. Opportunities can also originate outside the organization and must somehow be identified and subjected to the analytics process.
2. *Obtaining the data.* The second challenge is to obtain all of the necessary data in a consistent and timely fashion. It is usually difficult for most modeling groups to obtain the data they require in a timely fashion unless they have their own datamart, data warehouse, or distributed data processing system (White, 2009). In most organizations, the IT group controls access to the data.
3. *Deploying the models.* The third challenge is to deploy the models into operations or production systems in a consistent and timely fashion. Deployment challenges can directly impact analytics' efficacy. In most organizations, the IT group controls how models are deployed into products, services, and operations.

These are challenges for most organizations since the modeling group must work with other components of the organization to identify analytics opportunities, obtain the necessary data, and deploy the resulting models. The role of an analytics governance structure is to put in place an individual (the analytics leader) with sufficient authority to overcome these three challenges. An analytics governance structure must also include mechanisms for identifying, communicating, and resolving issues that are holding up analytics projects. Lastly, the governance structure requires a mechanism for providing sufficient resources for analytics projects and for balancing priorities between analytics projects and other corporate projects. At this stage of evolution of analytics governance structures, a complete set of parameters for designing a governance structure does not exist. We suggest the following preliminary parameters:

1. Ensure that sound long-term decisions about analytics are reached and that investments in analytics generate business value.
2. Operate in such a way that data, derived data, and analytics products are protected and managed in a secure and compliant fashion.
3. Operate in such a way as to make sure that there is accountability, transparency, and traceability to those who are funding analytics projects, to those who are developing and supporting analytics resources, and to those who are making use of analytics resources.
4. Provide an organization structure to ensure that the necessary analytics resources are available; that data is available to those developing analytics; that analytics can be deployed; that the impact of analytics is quantified and tracked; and that data, derived data, and data products are managed in a secure and compliant fashion.

These design parameters can be achieved by using governance committees:

- An analytics governance committee that includes senior management and representatives from the IT organization and various business stakeholders. This committee helps prioritize analytics opportunities; obtain resources for analytics projects; and ensure that those building the models get the data required, that the models that are built get deployed, and that deployed models measure the business value that they generate.
- An analytics technical policy committee that determines what data, analytics applications, processes, best practices, and standards are used across the organization.
- An analytics security and compliance committee that oversees the security and compliance of data and analytics processes and applications.
- An analytics data management and data quality committee that ensures the organization's data and metadata are accurate, complete, and consistent.

CONCLUSION

Organizations that desire to derive value from big data through analytics are more likely to succeed if they pay attention to the following four aspects of how analytics is viewed and organized: 1) Do senior leaders in the organization recognize the importance of analytics? 2) Is there a critical mass of data scientists who understand the organization and does the breadth of their expertise span not just building analytic models, but also deploying them? 3) Do the data scientists in the organization understand the various processes required for selecting the right models to build; building them correctly; and deploying them into operational systems and processes so that value is generated? 4) Is there an analytic governance structure in place to support analytics and to integrate analytics and big data into the organization's overall strategy?

ADDITIONAL INFORMATION

The views expressed in this paper are the views of the individual authors and do not necessarily reflect the views, opinions, intentions, plans, or strategies of their employers.

REFERENCES

Data Mining Group. 2012. Predictive Model Markup Language (PMML). Accessed April 7, 2014: www.dmg.org.

Davenport, TH. 2006. Competing on Analytics. *Harvard Business Review* 84(1): 99-107.

Galbraith JR. 2008. Organization design. In T.G. Cummings (Ed.), *Handbook of Organization Development*: Sage Publications, Inc., Thousand Oaks, CA.

Grossman RL. 2009. What is analytic infrastructure and why should you care? *SIGKDD Explorations Newsletter* 11(1): 5-9.

Laney D. 2001. 3D Data Management: Controlling Data Volume, Velocity, and Variety. META Group. Accessed April 7, 2014: http://blogs.gartner.com/ doug-laney/files/2012/01/ ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Manyika J, Chui M, Brown B, et al. 2011. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Accessed April 7, 2014: http:// www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_ innovation

Parmar R, Cohn DL, Marshall A. 2014. Driving innovation through data. IBM Institute for Business Value. Accessed April 7, 2014: www-935.ibm.com/services/us/gbs/ thoughtleadership/innovation-through-data/

Press G. 2013. A very short history of data science. Forbes. Accessed April 7, 2014: www. forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/

White T. 2009. *Hadoop: The Definitive Guide*. O'Reilly Media, Sebastopol, CA.

ROBERT L. GROSSMAN

University of Chicago
Computation Institute
E-mail: robert.grossman@uchicago.edu

KEVIN P. SIEGEL

Visa U.S.A.
E-mail: kpsiegel@yahoo.com

COLLABORATIVE APPROACHES NEEDED TO CLOSE THE BIG DATA SKILLS GAP

STEVEN MILLER

Abstract: The big data and analytics talent discussion has largely focused on a single role – the data scientist. However, the need is much broader than data scientists. Data has become a strategic business asset. Every professional occupation must adapt to this new mindset. Universities in partnership with industry must move quickly to ensure that the graduates they produce have the required skills for the age of big data. Existing curricula should be reviewed and adapted to ensure relevance. New curricula and degree programs are needed to meet the needs of industry.

Keywords: Big data, data science, organiztion design, big data jobs, big data talent/skills gap

A recent McKinsey Global Institute study forecasts a significant shortfall in big data skills in the U.S.: “By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions” (Manyika et al., 2011: 3). The big data skills shortage is global, and every region will face similar challenges. These numbers do not tell the whole story, however, because the lack of skilled big data practitioners of all types is limiting the ability of business to derive value from big data (Kelly, 2013). There is a talent shortfall as well in data strategy and in a wide variety of technical data management positions (Olavsrud, 2013; Rowe, 2013; Yerak, 2013), largely due to the shortfall in university, professional, and executive education programs designed to produce the talent needed to fill the growing demand for every type of big data professional.

Although the data scientist role gets the most attention, rebuilding businesses to strategically use data requires redefining existing jobs, creating new specialty roles, and perhaps creating new C-suite or redefining existing C-suite roles. Businesses will need highly skilled and experienced information strategists, data architects, data governance professionals, and visualization experts, as well as chief analytics officers (Advani, 2012) or chief data officers (Aiken & Gorman, 2013), all with broad data skills and most with highly specialized skill sets. In this article, I describe the needed new big data specialists and offer a set of recommendations about how to develop them.

WIDENING THE FOCUS BEYOND THE DATA SCIENTIST

Exactly what a data scientist is and what skill set is required is a subject of intense debate (Harris, 2013). No official standards exist yet, and consequently anyone can call himself or herself a data scientist. Universities and industry need to collaborate to define the profession and set basic standards. Questions that need to be answered include: Do data scientists need to be PhDs, or is data science an applied role? Is it reasonable to expect that individuals possess all required skills or are team approaches preferred? Must every data scientist be an expert in machine learning? What are the core skills required by a data scientist (individual or team), and how are they different from what is being taught in existing statistics or business analytics programs? Until these and other questions are asked and answered, both recruiting

Journal of Organization Design
JOD, 3(1): 26-30 (2014)
DOI: 10.7146/jod.9823

and educating data scientists will be major challenges for industry and academia. Additional considerations include what leadership and management skills (IT and non-IT) are required to ensure that firms and other organizations are leveraging their data assets. Even more important than adding specificity to the data scientist's job and role is a wider focus on other big data jobs. These include information strategists, information systems professionals, and data governance and ethics professionals.

The Information Strategist

What makes some companies more adept than others at leveraging data as a strategic asset (Deutscher, 2013)? One example is Amazon's decision to retain every customer's entire purchase history for operational use while many, if not most, other companies simply store the data for a brief period and then delete it forever. This affords Amazon deep customer insight which enables superior recommendations for future purchases. Information strategy is an emerging discipline that concerns itself with all aspects of data as a business asset (Adler, 2013). Today these business-focused data skills are largely developed on the job as university data management courses concern themselves with technical issues such as data models. Strategic business-level data management is not prominent in most curricula, but there is a pressing need for graduates with both business and industry acumen.

Big Data Information Systems Professionals

Most information systems curricula provide students with a solid introduction to traditional database concepts such as structured query language, relational database, and data warehousing. Businesses, however, need data professionals with broader and deeper analytics skills that enable them to tackle the full breadth of today's data management technologies and challenges, including security, privacy, master data, Hadoop, real-time streaming data, real-time predictive analytics, cloud, and mobile. Businesses need, for example, data professionals who understand when ACID (ACID, 2014) databases are mandatory and when compromises can be made for scalability because eventual consistency is good enough (Bailis & Ghodsi, 2013).

One of the fastest-growing big data jobs is that of data engineer (Data Engineer Job Trends, 2014; Data Engineer Jobs, 2014). The data engineer has deep knowledge of relational databases and NoSQL databases such as Hadoop, can integrate data from diverse data sources, and can design data-driven services. Data engineers work in tandem with data scientists (Walker, 2013). Big data architects (van Rijmenam, 2013) are senior technical staff capable of designing large-scale big data solutions that often span legacy systems and newer systems that handle real-time machine and sensor data (Machine-Generated Data, 2014). Information systems curricula need to evolve quickly to better prepare students for these emerging jobs and roles.

Data Governance and Ethics Professionals

These days failures in data security and governance regularly create public embarrassments for companies (Duhigg, 2012). Even those thought of as among the most tech-savvy have made major blunders. Big data introduces ethical challenges as well. Just because you can obtain the home address of everyone who owns a gun, should you publish that information for all to see (Maas & Levs, 2012)? De-identification is often used to protect the privacy of individuals by deleting or obscuring actual names, addresses, social security numbers, and other data elements which uniquely identify individuals. However, simple de-identification techniques are often insufficient to protect privacy in the age of big data since it is often possible to re-identify individuals by combining data from multiple public sources with your de-identified data (Re-identification, 2014). Data governance is a major challenge for every organization, public or private (Aiken, Allen, Parker, & Mattia, 2007). Unfortunately, data governance is largely absent from academic curricula. Data ethics courses, in particular, are rare and when offered are simply an elective course even though the argument can be made that the ethical handling of data should be a skill for every professional who works with big data. The skills for common data governance jobs, such as data steward and data quality analyst, are usually acquired on the job, as the data governance profession has not been fully developed by universities.

A COMPREHENSIVE TALENT DEVELOPMENT APPROACH

Business and academia must collaborate to clearly define the big data knowledge and skill sets required across the organization. Every profession, whether business or technical, will be impacted by big data and analytics (Laster, 2010; Miller, 2010; Parry, 2010; Vaidhyanathan, 2010). Top academic business programs, such as the Kellogg School of Management at Northwestern University, are embracing the big data phenomenon (Zettlemeyer, 2013). Law schools and law firms are responding to the need by changing curricula and practices (IPIC Academic Program, 2014). Some universities, such as Lehigh University, have created laboratories where students use technology to analyze data specific to the field of study (Financial Services Laboratory, 2014). Laboratories like these are vital to ensure that students develop a practical understanding of how to apply data and analytics skills in the real world. But focusing solely on data-specific jobs is not broad enough. Data and analytics literacy must become an expectation across all curricula, regardless of the ultimate field or degree pursued. Graduates without big data skills will not be prepared for the business challenges they will face upon entering today's workforce.

Given the growing emphasis on data science and business analytics, businesses will move to aggressively recruit talent and re-train existing staff with a focus on analytics. However, analytics skill alone is not sufficient to succeed. Differentiating your business will require a comprehensive strategy that considers data as a core business asset. The pressures to innovate and differentiate will only become more intense, and consequently the big data talent shortage will become even more acute. A major challenge for many universities is that the skills required by all big data and analytics professionals will require cross-program collaboration to produce the needed "T-shaped professionals" (Brooks, 2012). The vertical stem of the T is a foundation of deep disciplinary skill. The horizontal bar of the T adds the breadth of skill necessary to work across an organization with the ability to influence others, collaborate across disciplines, and develop creative solutions to complex business problems. Universities must respond by adapting curricula to the needs of business. Big data professionals will need deep and broad skills spanning both technical and business domains. Big data information systems graduates will need math and statistics, machine learning, predictive analytics, decision management, computer science and programming, data ethics, information law, information privacy, data security, data and information theory, and visualization and communication (the arts), in addition to the core information systems, database, data warehousing, and data mining education they receive today. For example, not only will business firms and other organizations require big data professionals who understand basic machine learning concepts and can apply existing algorithms to a solution, but those organizations striving to differentiate themselves in an increasingly competitive market will need hard-to-find, highly skilled, machine-learning data scientists who can invent innovative machine-learning algorithms to underpin groundbreaking solutions. Lastly, the challenge is not simply about IT and technology. Jobs spanning the entire business spectrum, including legal, sales, marketing, finance, product development, manufacturing, and operations, will be impacted by the big data phenomenon.

RECOMMENDED ACTIONS

Academic leaders in partnership with industry and government need to assess the rapidly changing technological landscape and create new curricula and programs to develop talent for the increasing number of big data jobs. By working collaboratively with industry partners, academic leaders must evaluate all existing curricula and programs to determine where and how data and analytics knowledge and skills can be infused into the curriculum to ensure graduates have the skills industry requires to compete in a big data world. Academia, industry, and government should join together to create a national consortium to address the big data and analytics skills challenge. That consortium would:

- Create formal definitions of prioritized jobs such as data scientist, information strategist, big data architect, and data governance professional.

- Establish curriculum requirements and accreditation standards for programs designed to produce the required knowledge and skills for specialty jobs. Use workforce analytics (Ringo, 2012) to provide actionable feedback to ensure that curricula rapidly evolve to meet the evolving needs of industry.
- Set minimum standards for data and analytics literacy required by all students in the age of big data. Create and deliver literacy training via massive online open courses (MOOCs).
- Create open online communities around shared interests to engage industry, government, and academia.
- Partner with industry organizations such as IT-oLogy and the National Consortium for Data Science (National Consortium for Data Science, 2014) to establish strong internship programs and increase collaboration between business and academia.
- Foster the creation of textbooks and courseware to address both literacy and specialized skills at all levels from undergraduate to executive education.
- Establish working groups to address key data policy issues such as information security, individual privacy, and the ethical use of big data.

By working collaboratively, industry, academia, and government can begin to close the knowledge and skills gaps outlined in this article and better prepare students, managers, and professionals for big data jobs of the future.

REFERENCES

ACID. 2014. *Wikipedia*. http://en.wikipedia.org/wiki/ACID (Accessed March 29, 2014)

Adler S. 2013 (February 1). I am an information strategist. *IBM Data Magazine*. http://ibmdatamag.com/2013/02/i-am-an-information-strategist/

Advani D. 2012 (July). Chief Analytics Officer: The newest member of the C-suite. The Atlantic: Smarter CMO Solutions Blog. http://www.theatlantic.com/sponsored/ibm-power-data/archive/2012/07/chief-analytics-officer-the-newest-member-of-the-c-suite/260570.

Aiken P, Allen MD, Parker B, Mattia A. 2007. Measuring data management practice maturity: A community's self-assessment. *Computer* 40(4): 42–50. http://www.computer.org/csdl/mags/co/2007/04/r4042.html

Aiken P, Gorman MM. 2013. *The Case for the Chief Data Officer: Recasting the C-Suite to Leverage Your Most Valuable Asset* (1st ed.). Morgan Kaufmann, San Francisco, CA.

Bailis P, Ghodsi A. 2013 (April 9). Eventual consistency today: Limitations, extensions, and beyond. *ACM queue*. http://queue.acm.org/detail.cfm?id=2462076

Brooks K. 2012 (April 19). Career success starts with a "T." *Psychology Today*. http://www.psychologytoday.com/blog/career-transitions/201204/career-success-starts-t

Data Engineer Job Trends. 2014. Indeed.com. http://www.indeed.com/jobtrends?q=%22data+engineer%22&relative=1 (Accessed March 30, 2014)

Data Engineer Jobs. 2014. *LinkedIn*. https://www.linkedin.com/job/q-data-engineer-jobs (Accessed March 29, 2014)

Deutscher M. 2013 (March 11). IBM's CEO says big data is like oil, enterprises need help extracting the value. *SiliconANGLE*. http://siliconangle.com/blog/2013/03/11/ibms-ceo-says-big-data-is-like-oil-enterprises-need-help-extracting-the-value/

Duhigg C. 2012. How companies learn your secrets. *NYTimes.com*. http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=0

Financial Services Laboratory. 2014. *Lehigh University*. http://fsl.lehigh.edu/ (Accessed March 29, 2014)

Harris D. 2013 (July 11). Kaggle now has 100K data scientists, but what's a data scientist? *GigaOM.com*. http://gigaom.com/2013/07/11/kaggle-now-has-100k-data-scientists-but-whats-a-data-scientist/

IPIC Academic Program. 2014. Michigan State University College of Law. http://www.law.msu.edu/ipic/program.html (Accessed March 29, 2014)

Kelly J. 2013 (September 19). Enterprises struggling to derive maximum value from big data. *Wikibon*. http://wikibon.org/wiki/v/Enterprises_Struggling_to_Derive_Maximum_Value_from_Big_Data

Laster J. 2010 (May 28). Scholars scale up music studies. *The Chronicle of Higher Education*. http://chronicle.com/article/Scholars-Scale-Up-Music/65709/

Maas K, Levs J. 2012 (December 25). Newspaper sparks outrage for publishing names, addresses of gun permit holders. *CNN.com*. http://www.cnn.com/2012/12/25/us/new-york-gun-permit-map/index.html

Machine-Generated Data. 2014. *Wikipedia*. http://en.wikipedia.org/wiki/Machine-generated_data (Accessed March 29, 2014)

Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C. 2011. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Miller MH. 2010 (May 28). The spoken word, earchable for scholarship. *The Chronicle of Higher Education*. http://chronicle.com/article/The-Spoken-Word-Searchable/65706/

National Consortium for Data Science. 2014. http://data2discovery.org/ (Accessed March 29, 2014)

Olavsrud T. 2013 (September 25). Enterprises confident about tackling big data initiatives. *CIO.com*. http://www.cio.com/article/740321/Enterprises_Confident_About_Tackling_Big_Data_Initiatives

Parry M. 2010 (May 28). The humanities go Google. *The Chronicle of Higher Education*. http://chronicle.com/article/The-Humanities-Go-Google/65713/

Re-identification. 2014. Electronic Privacy Information Center. http://epic.org/privacy/reidentification/ (Accessed March 29, 2014)

Ringo T. 2012 (March 23). Workforce analytics isn't as scary as it sounds. Harvard Business Review Blog. http://blogs.hbr.org/2012/03/workforce-analytics-is-not-as/

Rowe A. 2013 (July 6). The big search for big data candidates. Wanted Analytics Blog. http://www.wantedanalytics.com/insight/2013/07/06/the-big-search-for-big-data-candidates/ (Accessed March 29, 2014)

Vaidhyanathan S. 2010 (June 3). Crunching words in great numbers. *The Chronicle of Higher Education*. http://chronicle.com/article/Crunching-Words-in-Great/65777/

van Rijmenam M. 2013 (May 6). Big data solutions architect job description. Big Data Startups Blog. http://www.bigdata-startups.com/job-description-big-data-solutions-architect/.

Walker M. 2013 (July 2). Data scientists vs. data engineers. Data Science Central. http://www.datasciencecentral.com/profiles/blogs/data-scientists-vs-data-engineers

Yerak B. 2013 (August 25). In growing field of big data, jobs go unfilled. *Chicago Tribune*. http://articles.chicagotribune.com/2013-08-25/business/ct-biz-0825-big-data-20130825_1_big-data-data-sources-data-analysts

Zettlemeyer F. 2013. The four elements of a big data mindset. Kellogg School of Management, Northwestern University. http://www.kellogg.northwestern.edu/news_articles/2013/06272013-big-data-mindset.aspx

STEVEN MILLER

Talent & Skills Ecosystem Strategist
IBM
E-mail: steven.miller@us.ibm.com

BIG DATA – BIG DEAL FOR ORGANIZATION DESIGN?

JANNE J. KORHONEN

Abstract: Analytics is an increasingly important source of competitive advantage. It has even been posited that big data will be the next strategic emphasis of organizations and that analytics capability will be manifested in organizational structure. In this article, I explore how analytics capability might be reflected in organizational structure using the notion of "requisite organization" developed by Jaques (1998). Requisite organization argues that a new strategic emphasis requires the addition of a new stratum in the organization, resulting in greater organizational complexity. Requisite organization could serve as an objective, verifiable criterion for what qualifies as a genuine new strategic emphasis. Such a criterion is necessary for research on the co-evolution of strategy and structure.

Keywords: Big data, organization design, organization structure, structural evolution, strategy and structure, requisite organization

Big data – large sets of data that can be captured, communicated, aggregated, stored, and analyzed – is widely regarded as the next frontier for innovation, competition, and productivity (McKinsey Global Institute, 2011). Davenport and Harris (2006) studied "analytics competitors" – a handful of organizations that have made a commitment to and achieved proficiency in quantitative, fact-based analysis as a competitive differentiator – and noted that those organizations create centralized groups to ensure consistency and data sharing throughout the enterprise and that their shift to analytics is driven from the very top, ideally by the CEO. Galbraith (2012a,b, 2014) posits that big data and analytics may provide a basis for a new structural dimension (in addition to functions, businesses, countries, and customers) that will be "concatenated" (Chandler, 1962) to the organization structures of the future.

My point-of-view article proposes an objective, verifiable criterion for what qualifies as a genuine new strategic dimension such as big data. "Requisite organization," a long-established framework originated by Jaques (1998), is put forward as a yardstick of organizational complexity that pertains to both strategy and structure. Following the requisite organization logic, each concatenation of a strategic dimension would structurally require a new requisite stratum, reflecting a stepwise increase in organizational complexity.

STRUCTURE FOLLOWS STRATEGY

The evolution of organizational structure transpires through a dialectic process of differentiation and integration, with greater differentiation resulting in more organizational complexity (Lawrence & Lorsch, 1967). The pattern of differentiation and integration is manifested in structural adjustments to meet changes in strategy resulting from an expansion of the organization's activities. Chandler (1962) theorized that structural evolution is the cumulative result of several basic strategies, a process he called "concatenation." Each new dimension of organizational structure pertains to a new strategy concatenated to the previous ones. The four basic strategies and respective structures identified by Chandler (1962) are shown in Table 1.

Journal of Organization Design
JOD, 3(1): 31-36 (2014)
DOI: 10.7146/jod.13261

Table 1. Structure Follows Strategy

Strategy	Structure
Expansion of Volume	An administrative office of a field unit (e.g., plant, sales office) to handle one function in one local area
Geographical Dispersion	A departmental structure and headquarters to administer several local field units
Vertical Integration	A central office to administer multiple departments each of which is responsible for a major function (e.g., manufacturing, sales, research)
Diversification	A general office to administer multiple self-contained divisions

Source: Chandler (1962)

Using the concept of concatenation, Galbraith (2012a,b) describes two more recent strategies, international growth (Stopford & Wells, 1972) and customer focus (Galbraith, 2005, 2010) as well as their structural manifestations. In the case of international growth, the line organization is geographic (typically in the case of B2C enterprises) or based on worldwide business units (typically in the case of R&D-intensive B2B enterprises), with the other dimension being matrixed to the dominant profit center structure. In response to customers' preference to be served through global account units, large global firms such as Walmart, Accenture, and Procter & Gamble have added a fourth dimension to their structure, organizing the "front end" of the value chain by customers. Galbraith (2012a,b) further posits that big data may provide the basis for a new dimension to be concatenated to organization structures of the future. However, he does not specify how a big data strategy will be manifested in an organization's structure.

REQUISITE ORGANIZATION: YARDSTICK OF STRUCTURAL EVOLUTION

In his rigorous conceptual and empirical research that spanned several decades, Jaques (1976, 1998, 2002; Jaques & Cason, 1994; Jaques & Clement, 1991) recognized that organizations exhibit a hierarchical ordering of work complexity that reflects differences in human capability. Role complexity increases discontinuously in specific steps, stratifying different kinds of work into natural layers or "strata." As shown in Table 2, Strata I–IV pertain to the symbolic-verbal order of complexity, embracing activities from day-to-day first-line work to the middle management levels. The conceptual-abstract order of complexity covers Strata V and beyond at the higher management levels, typically in the corporate realm (Jaques, 1998). The major roles at each stratum, along with their time span of discretion, are also shown in Table 2.

Table 2. Requisite Organization

Order of Complexity	Stratum	Role	Time Span of Discretion
Conceptual-Abstract	VIII	Super-corporation CEOs	50+ years
	VII	COOs of large corporations	20–50 years
	VI	Corporate Executive VPs	10–20 years
	V	Business unit Presidents	5–10 years
Symbolic-Verbal	IV	General Managers	2–5 years
	III	Managers of mutual recognition units; senior professionals	1–2 years
	II	First-line managers; professional specialists	3 months to 1 year
	I	First-line manual and clerical workers	1 day to 3 months

Source: Jaques (1998)

Following Jaques (1998), I hypothesize that each structural evolution pertaining to the concatenation of a new strategy requires an additional stratum. Viewed against the backdrop of requisite organization, the strategy of volume expansion can be seen as Stratum III strategy. The corresponding structural response is a field unit in Chandler's (1962) nomenclature, such as a sales office or retail outlet. The strategy of geographical expansion to larger markets calls for decentralization to a departmental structure. The autonomy of field units begets the crisis of control (Greiner, 1972), where direct control over the organization as a whole is lost. A departmental headquarters (Chandler, 1962) is required at Str-IV to coordinate a number of Str-III units. Management is less direct, and operating decisions are delegated through policy edicts (Scott, 1971). As coordination systems and processes outgrow their original intention and become overly bureaucratic, vertical integration (Chandler, 1962) toward a Str-V structure is needed. The “red-tape crisis” (Greiner, 1972) is resolved through less formal mechanisms such as normative control and interpersonal cooperation. The Str-V structure is a “unified whole system” (Jaques, 1998) that gets its closure in a central office (Chandler, 1962) that integrates different functions into a coherent organizational whole. Diversification strategy (Chandler, 1962) denotes a shift beyond a single Str-V organization. There is no internal solution for further growth, so a diversification strategy moves the firm into new markets and businesses. Chandler (1962) suggests an M-form organization to accommodate diversification, with a general office to administer multiple quasi-autonomous divisions. In terms of requisite organization, this form pertains to a Str-VI corporate structure with Str-V business units.

Extrapolating from the basic strategies identified by Chandler (1962) along the lines suggested by Galbraith (2012a,b), the international growth and customer-focus strategies would denote Str-VII and Str-VIII, respectively, in the requisite organization scheme. Figure 1 exhibits a three-dimensional matrix structure. As described by Galbraith (2010), in Nestlé’s matrix structure the businesses and functions report to the country manager and to their respective corporate units forming a three-dimensional matrix. In terms of requisite organization, this structure would be a Str-VII organization with global and local business units at Str-V.

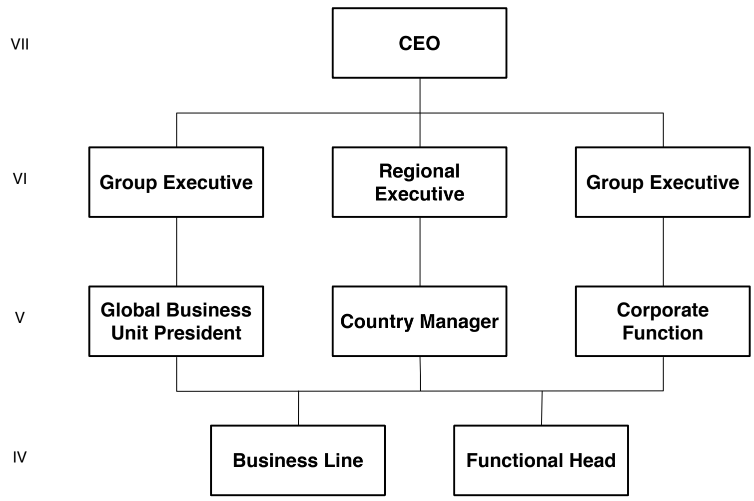

Fig. 1. Str-VII three-dimensional matrix structure

Figure 2 illustrates a four-dimensional matrix structure in which regional units report both to regional teams and to regional heads of the global business units. On the other hand, the regional teams report to both their regional manager and to the customer team. In terms of requisite organization, this structure would be a Str-VIII organization with regional units at Str-III; regional, global, and customer-specific business units at Str-V; a division to back-end and front-end at Str-VII; and the cross-over CEO at Str-VIII. Such a structure is very complex and is only manifested in the largest multinational corporations such as Procter & Gamble (Galbraith, 2010, 2012b).

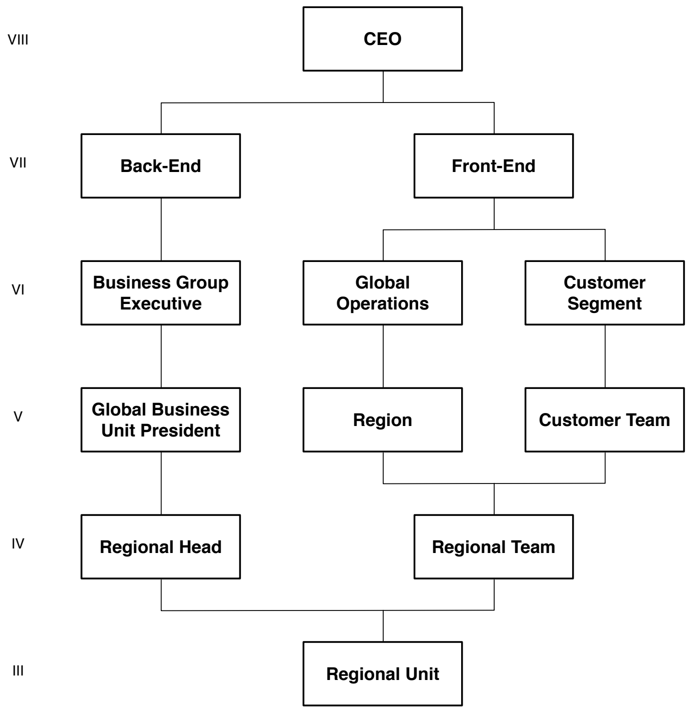

Fig. 2. Str-VIII four-dimensional matrix structure

SO WHAT'S THE BIG DEAL?

Returning to Galbraith's (2012a,b, 2014) contention that big data and analytics would provide a basis for a new structural dimension in organizations of the future, the question naturally arises as to what will be the basic strategy pertaining to big data and how it will be reflected in organizational structure through the concatenation process. While analytics seems to be an important source of competitive advantage, I contend that future concatenation should meet some verifiable, objective criterion. As requisite organization (Jaques, 1998) describes and prescribes the ordering of work complexity in organizations, I suggest that the criterion for concatenation would be that a new stratum is required to incorporate a big data and analytics strategy into the organizational structure. As per my hypothesis, there would then be a litmus test for determining whether big data, indeed, represents a basis for a new strategic dimension. Extrapolating from the observations above, competition on analytics would thereby require a Str-IX structure.

Galbraith (2012b: 10) rightfully asks if new strategic dimensions can be added and embedded in already complex global enterprises: "First, will the growth drivers continue to create additional organizational dimensions as markets, channels, and media become more fragmented and specialized? And, second, can organizations continue to create the integrating mechanisms needed to handle more complex interdependence?" The largest "super-organizations" of the world are at Str-VIII complexity, while Str-IX is concerned with the shaping of societies (Jaques, 1986). Few sources in the work levels literature even refer to strata beyond Str-VII. Ivanov (2013) recognizes the theoretical existence of Str-IX but notes that organizations of this level of complexity have not been found.

Galbraith (2012b) sees inklings of a new structural dimension in the emergence of increasingly substantial analytics units in a few actual organizations. He refers to an anonymous (but real) company whose organizational unit called Decision Analytics provides services for all of the firm's businesses and country organizations. He also considers Nike's Digital Division, responsible for building digital capabilities in other units, as something that may grow to a fifth dimension of strategy and organization (Galbraith, 2014). Just like the third (geographical) dimension grows out of an initial international division (Galbraith, 2010), the fifth dimension may be preceded by such onsets.

Assuming that a new strategic dimension would entail a new requisite stratum, an extrapolation from three- and four-dimensional structures (as depicted in Figures 1 and 2) might result in something like the organizational structure depicted in Figure 3. As indicated, the enterprise would be divided into the business organization and the analytics organization, the latter "informating" (Zuboff, 1985) the former. This would bring to full life Zuboff's (1985: 9-10) decades-old notion wherein organizations "…recreate their own images in the form of detailed, real-time, integrated databases which give access to internal operations and external business data and can be reflexive enough to organize, summarize, and analyze aspects of their own content."

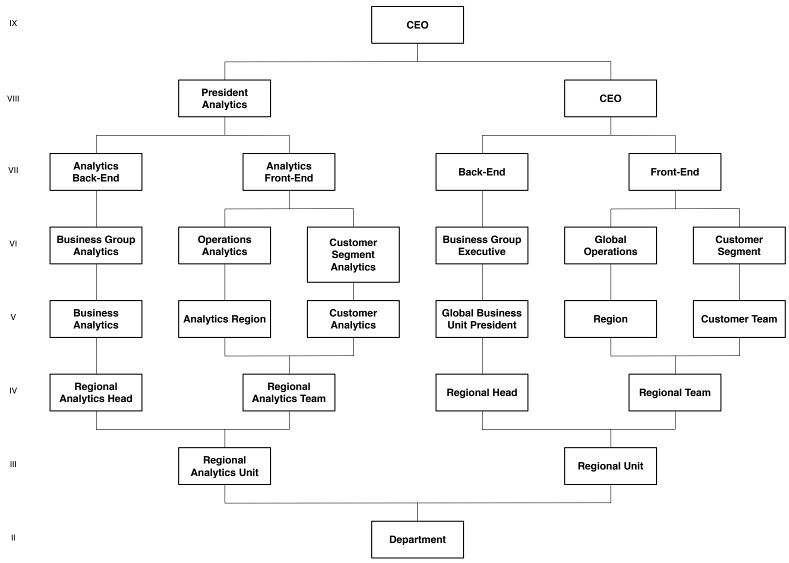

Fig. 3. A hypothetical Str-IX five-dimensional matrix structure

For the sake of argument, the analytics organization in Figure 3 mirrors the business organization (here simply a replica from Figure 2) at all lines and levels. In an actual organization, it would probably not mirror the business organization as closely, but given the predicted exalted role of analytics, one might expect in a five-dimensional structure that business managers at all levels will have peers in the analytics organization. Should an organization of this size and complexity (nine strata and five strategic dimensions) actually emerge, it would indeed be a big deal as no organization of this complexity presently exists.

CONCLUSION

In this point-of-view article, I have addressed the important relationship between organizational strategy and structure. Both Chandler (1962) and Jaques (1998) have shown how a new strategic emphasis is manifested in organizational structure, introducing the concepts of concatenation and requisite organization, respectively. In his various writings, Galbraith has chronicled the new strategic and structural dimensions as they have emerged. Moving forward, organization design scholars can and should use these ideas and concepts to objectively measure and document the evolution of organizational complexity.

REFERENCES

Chandler AD. 1962. *Strategy and Structure: Chapters in the History of the American Enterprise*. MIT Press, Cambridge, MA.

Davenport TH, Harris JG. 2006. *Competing on Analytics: The New Science of Winning*. Harvard Business School Press, Boston, MA.

Galbraith JR. 2005. *Designing the Customer-Centric Organization: A Guide to Strategy, Structure, and Process*. Jossey-Bass, San Francisco, CA.

Galbraith JR. 2010. The multi-dimensional and reconfigurable organization. *Organizational Dynamics* 39(2): 115–125.

Galbraith JR. 2012a. The future of organization design. *Journal of Organization Design* 1(1): 3–6.
Galbraith JR. 2012b. The evolution of enterprise organization designs. *Journal of Organization Design* 1(2): 1–13.
Galbraith JR. 2014. *Designing Organizations: Strategy, Structure, and Process at the Business Unit and Enterprise Levels*. 3rd edition. Jossey-Bass, San Francisco, CA.
Greiner LE. 1972. Evolution and revolution as organizations grow. *Harvard Business Review* 50(4): 37–46.
Ivanov S. 2013. Defects in modern organizations: Field findings and discovery. *International Journal of Innovation, Management and Technology* 4(2): 204–208.
Jaques E. 1976. *A General Theory of Bureaucracy*. Heinemann Educational, London, U.K.
Jaques E. 1986. The development of intellectual capability: A discussion of stratified systems theory. *Journal of Applied Behavioral Science* 22(4): 361–383.
Jaques E. 1998. *Requisite Organization: A Total System for Effective Managerial Organization and Managerial Leadership for the 21st Century*. Revised edition. Cason Hall & Co., Baltimore, MD.
Jaques E. 2002. *The Life and Behavior of Living Organisms: A General Theory*. Praeger Publishers, Westport, CT.
Jaques E, Cason K. 1994. *Human Capability: A Study of Individual Potential and Its Application*. Cason Hall & Co., Arlington, VA.
Jaques E, Clement SD. 1991. *Executive Leadership: A Practical Guide to Managing Complexity*. Cason Hall & Co., Arlington, VA.
Lawrence PR, Lorsch JW. 1967. *Organization and Environment: Managing Differentiation and Integration*. Harvard University. Graduate School of Business Administration, Boston, MA.
McKinsey Global Institute. 2011. Big data: The next frontier for innovation, competition, and productivity.
Scott BR. 1971. *Stages of Corporate Development, Part 1*. Harvard University. Graduate School of Business Administration, Boston, MA.
Stopford JM, Wells LT. 1972. *Managing the Multinational Enterprise: Organization of the Firm and Ownership of the Subsidiaries*. Longmans, London, U.K.
Zuboff S. 1985. Automate/informate: The two faces of intelligent technology. *Organizational Dynamics* 14(2): 5–18.

JANNE J. KORHONEN

Doctoral Researcher
Aalto University School of Science
E-mail: janne.korhonen@aalto.fi

BIG DATA AND ORGANIZATION DESIGN

KEY CHALLENGES AWAIT THE SURVEY RESEARCH FIRM

TIM J. GABEL • CATHY TOKARSKI

Abstract: Digital data is everywhere, and its ubiquity is causing profound changes in our personal lives and in the functions of government, business, and academia. Organizations of all sizes and purposes are seeking to take advantage of the big data tsunami and the opportunities it presents. RTI International, a non-profit U.S. research organization, draws more than 80 percent of its $760 million in annual revenues from competitive grants and contracts funded by the U.S. government. The organization is rich in talent and expertise but not currently aligned in a way that meets big data's challenges. To thrive in this rapidly changing environment, RTI must determine how to seize opportunities big data presents, survive the threats posed by big data, and offer its clients expanded services. How well RTI responds to these challenges will determine its role in the search for solutions to the major social and scientific problems of our day.

Keywords: Big data, RTI International, survey research, data science, big data workforce

Forward-thinking organizations and their leaders around the world are wrestling with the ”big data revolution” and its impact on their businesses. Some firms are mining mountains of process data to fine-tune manufacturing operations. Others are sifting through Facebook postings and Twitter feeds to understand customer sentiments about their products and services. Software vendors are marketing sector-specific applications – ”insight solutions” – to assist in this process. Meanwhile, high-tech companies are collaborating with consumer-ratings groups on mobile applications that use government-financed research to compare the safety and effectiveness of various medications and offer the best recommendation to a patient (Comstock, 2013). To help organizations store, aggregate, and retrieve these ever-expanding forms of information, cloud computing vendors are reshaping the data storage landscape.

In 2010, more than four billion people, or 60 percent of the world's population, were using mobile phones (Manyika et al., 2011). About 12 percent of them were using smartphones, a percentage that's growing more than 20 percent per year. Big data is estimated to represent potentially $300 billion to the U.S. healthcare system through greater efficiency and delivered value. Global personal location data, which can quickly identify the most efficient route from Point A to Point B, represents as much as $100 billion in revenue for service providers.

Since 2000, the amount of information collected by the federal government has increased at a mind-boggling rate (TechAmerica Foundation, 2012). In 2009, the federal government produced 848 petabytes of data, and U.S. healthcare data on its own reached 150 exabytes. Five exabytes of data would contain all of the words ever spoken by human beings on earth. As big data permeates every aspect of how we live and conduct business, most organizations fear standing on the sidelines while others figure out how to use big data to their advantage.

JOD, 3(1): 37-45 (2014)
DOI: 10.7146/jod.9753

In response to these and numerous other developments, many organizations are attempting to prepare themselves to take advantage of the big data tsunami and the opportunities it presents. Some are opening new divisions or setting up ”predictive analytics” initiatives. In many cases, this involves hiring new talent – data scientists and other professionals – to help lead their efforts. While data-savvy companies may be on the hunt for the next Nate Silver, the statistical whiz who correctly predicted the outcome of the 2008 and 2012 presidential elections, their efforts may be hindered by dire predictions of a looming shortage of qualified data analysts and managers (Davenport & Patil, 2012; Rooney, 2012).

RTI International, a survey research firm, is an organization that is rich in talent and expertise but not currently aligned in a way that meets big data's challenges. To thrive in this new environment, we must address questions such as:

- How do we re-calibrate technological savvy and subject matter expertise in order to meet emerging business opportunities associated with big data?
- How do we organize staff so that subject matter experts and more data-savvy junior staff can easily share their expertise across disciplines?
- How do we avoid investing resources in massive projects (the ”build it and they will come” mentality) before truly understanding clients' needs and requirements?

Our article aims to share a practitioner's perspective on the challenges of restructuring a knowledge-worker company in the midst of the big data revolution. In some cases, these challenges include retooling fundamental human resource processes such as recruiting and hiring, performance management, and talent development. As organizations begin to orient themselves and their workforce to meet big data's demands, new areas of opportunity are likely to emerge, but the path forward will not be clear and straightforward. After briefly describing RTI International, we discuss the main challenges to the organization posed by big data. Based on RTI's experiences to date, we derive several implications for organizations in general.

RTI INTERNATIONAL: BACKGROUND, MISSION, AND CAPABILITIES

RTI International is an independent, nonprofit research institute that provides research, development, and technical services to government (local, state, and federal) and commercial clients worldwide. Founded in 1958 with the creation of North Carolina's Research Triangle Park, RTI was conceptualized as a partnership between North Carolina's business leaders, the state government, and the region's three major universities (University of North Carolina-Chapel Hill, North Carolina State University, and Duke University). With a mission of improving the human condition by turning knowledge into practice, RTI leverages its research and technical capabilities to solve critical social and scientific problems.

The company's annual revenues are approximately $760 million, with over 80 percent of this sum derived from competitive grants and contracts funded by the U.S. government. More than 3,700 staff from 250 scientific and technical disciplines work in eight U.S.-based and ten international offices. Staff members carry out approximately 1,800 funded research projects annually, many of which result in peer-reviewed publications and/or adjudicated government statistical reports.

RTI is organized into four business units:

- Social, Statistical, and Environmental Sciences – Program areas include criminal justice and behavioral health, environmental sciences, statistics and epidemiology, and survey and computing sciences.
- Discovery Science and Technology – Program areas include energy technology, materials and electronic technology, organic and medicinal chemistry, pharmacology and toxicology, biomarkers and systems biology, and analytical chemistry and pharmaceutics.
- International Development – Program areas include governance and economic development, international education, and global health.
- Health Solutions – A specialized group focused entirely on services for the pharmaceutical and biotechnology sectors.

The Social, Statistical, and Environmental Sciences (SSES) business unit, led by the first author, is RTI's largest, with approximately $385 million in annual revenue, 1,500 professional staff, and 1,200 temporary data collectors. Gabel's role is to set the strategic direction for the group and ensure that it is appropriately organized to face the future.

BIG DATA'S CHALLENGES TO RTI INTERNATIONAL

Big data's transformative nature presents unique challenges to RTI, whose organizational structure promotes deep subject matter expertise and research capacity. While this structure has yielded significant benefits, it can inhibit the type of cross-disciplinary innovation that big data demands. For RTI to become a stronger organization in the future, it must successfully address three large challenges posed by the big data phenomenon.

Seize Opportunities

RTI International was founded more than 50 years ago and has long been organized along traditional scientific/academic disciplines (e.g., economics, statistics, engineering, chemistry). Even when reorganizations have taken place, the cultural affinities of staff tend to run along disciplinary lines. Likewise, RTI's human resource system (e.g., job descriptions, job families and functions, job expectations and promotion criteria) mirror that of academic disciplines. For example, the most common job titles in SSES, and the number of people holding each job, are: Public Health (335), Economics/Health Economics (197), Systems Analysis and Programming (196), Statistics (156), Survey Methodology/Operations (149), Education (126), Environmental Science (95), and Epidemiology (47). Nearly 40 percent of our technical staff have terminal degrees, and 65 percent have at least a master's degree. About 53 percent of RTI's workforce is female, and 29 percent are 35 years of age or younger. Average employee tenure is just under eight years.

Our well-tested and successful business model uses a matrix management approach, assembling cross-department teams for specific proposal and project efforts. Over time, this approach has produced a culture of specialization and its attendant demands and rewards. Despite the successes that this model has brought to RTI in the past, we recognize that it is unlikely to succeed in the era of big data, which demands the blending of disciplines, especially across the boundaries of subject matter, statistics, and computing. We are trying to come to grips with the best way to organize ourselves to address the challenges and opportunities that big data will present.

Yet it is not obvious how to go about that. RTI's leadership and business units are seeking how to address the challenges of big data in the clearest and most realistic manner. Our statistics group sees it primarily as an analytics issue, with a clear emphasis on their statistical sampling expertise. Our technology group sees it as a high-performance computing problem. And our subject matter experts see enormous potential in the power of virtually unlimited sample sizes. RTI's legal office and Institutional Review Boards see exponential increases in data privacy risks. Not surprisingly, each views the big data phenomenon from its unique functional perspective. Fortunately, we are in a position to leverage our talent to capitalize on the opportunities big data presents, but it will require us to break down silos, revise our job descriptions and hiring practices to attract staff with blended skills, and re-think the way we create our project teams. To begin this process, SSES initiated a re-organization this year that more tightly coordinates our data, computing, statistical, and analytics resources. In one set of moves, staff from our computing, biostatistics, and epidemiology departments were merged into a single new research center with a defined focus on managing health research data. In another move, a set of sampling statisticians, programmers, and data managers were combined to form a new data science and statistical methods center. While seemingly modest, both announced changes were perceived as controversial, as they moved researchers out of traditional discipline-focused departments.

In organizations like RTI, such changes can elicit a strong, and even emotional, response from staff. In Gabel's judgment, this is because it strikes at the core of employees' professional identity, especially if the corporate culture has long valued and celebrated a disciplinary focus. Employees attach great meaning to the name of the organizational unit in which they reside. Moving a statistician out of the "statistics department" and into the "health data management" or "energy analytics" department raises questions around career path and professional growth, job performance expectations, access to appropriate mentoring, and the professional risk of overly narrow specialization. These are all reasonable concerns for employees when they are wondering, "What do these changes mean for me and who I am?"

To help overcome such concerns, our reorganization included a robust communication and change-management plan. Town hall and small-group meetings addressed concerns head-on, and staff members helped develop and review responses to frequently asked questions. At every step, we tried to ensure that we could answer the "Why are we doing this" question with market data and explanations of how this would better prepare RTI to capture current and future market opportunities. The goal in adopting this approach is twofold: to more closely align our subject matter and data science expertise and to sustain those collaborations within RTI's specific disciplines. In effect, our reorganization has built dotted-line bridges that connect RTI's subject matter and data science experts as they create new communities while allowing experts to maintain their professional identities.

We anticipate that our new organizational matrix will evolve as the market matures and our clients gain a deeper understanding of their business needs. We have adjusted our organizational matrix to support the needs of the market, such as with the creation of RTI's multidisciplinary Center for the Advancement of Health IT in 2010. We will continue to make adjustments to become more market and customer-oriented in the future.

To enhance our new organizational approach, RTI recently launched a customer relationship management platform called Salesforce.com. This robust tool, which is used by many organizations to support their sales and business development functions, includes a social media module called Chatter. Similar in look and feel to Facebook, Chatter allows RTI staff to form and communicate through cross-disciplinary groups around cutting-edge topics such as Big Data/Big Science, Global Health Informatics, Implementation Science, and Education and Workforce Development for 2025. Still in its early months of implementation, Chatter is proving to be a popular and time-efficient mechanism for building virtual expert communities within RTI. We are already seeing the organic formation of Chatter communities, including discipline-focused groups comprised of staff formerly in the same administrative unit. Of course, neither our re-organization nor our collaborative technology roll-out replaces the need for our domain experts to keep themselves up-to-date in their respective areas of expertise. We are actively encouraging our subject matter experts to collaborate on research projects and peer-reviewed publications across business silos, and to retain their membership and participation in professional societies and associations aligned to their expertise. We are also encouraging them to think about, study, and publish on the impact of big data on their discipline.

Survive Threats

RTI's second big data challenge can be likened to the daunting task of "rebuilding the airplane engine in mid-flight." As the largest business unit at RTI, SSES has a particular responsibility for maintaining our core capabilities while also positioning us for the future. In other words, we must find a way to maintain an engine that risks sputtering while operating at full capacity. The airplane engine is RTI's survey research business. Much of SSES's revenue comes from conducting scientifically rigorous, statistically representative survey projects (face-to-face household surveys, establishment surveys, and telephone surveys) for RTI's federal government clients. As information from big data's many sources is now available on a 24/7 basis, the intrinsic value of the statistically representative survey must be redefined.
The advent of big data is forcing a paradigm shift in the federal government's statistical system, as reflected in this blog post (http://directorsblog.blogs.census.gov/2011/05/31/designed-data-and-organic-data/) by Robert Groves, the former Director of the U.S. Census Bureau:

> *We're entering a world where data will be the cheapest commodity around, simply because the society has created systems that automatically track transactions of all sorts. For example, internet search engines build data sets with every entry, Twitter*

generates tweet data continuously, traffic cameras digitally count cars, scanners record purchases, RFIDs signal the presence of packages and equipment, and internet sites capture and store mouse clicks. Collectively, the society is assembling data on massive amounts of its behaviors. Indeed, if you think of these processes as an ecosystem, it is self-measuring in increasingly broad scope. Indeed, we might label these data as "organic," a now-natural feature of this ecosystem.

In his article on the three eras of survey research, Groves (2011) distinguishes between "organic" data and "designed" data, or data collected via well-designed questionnaires in order to answer hypothesis-driven policy questions. Like the Census Bureau, RTI staff are experts at carrying out complex efforts to collect, manage, and analyze what Groves refers to as designed data. We are not, however, adequately prepared to change our core business in response to the onslaught of organic data generated by big data. Groves continues to be a thought leader on this subject. Now provost of Georgetown University, he welcomed the creation of the Massive Data Institute at the university's newly formed McCourt School of Public Policy (Kerr, 2013). Groves noted the institute's potential to capitalize on the explosive growth of quantitative public information through sites such as www.data.gov to help frame policy issues and train new generations of government leaders. Such training can also help address critical staffing needs at government agencies, where "just having the talent that can navigate these files" is sorely needed (Anderson, 2013).

The formation of the Massive Data Institute and the potential it holds for addressing the pressing public policy and scientific issues of the 21st century signals that disruptive innovation to the established federal survey research business is just around the corner (Christiansen, 2013). Disruptive innovations are not uncommon, of course, and many industries have faced comparable or greater challenges. Big data represents a threat to one of RTI's core competency areas as the federal survey research community faces increased competition with a "big data is faster and cheaper" value proposition. Our response thus far to leverage knowledge from organic data has been a "skunkworks" approach, using internal resources to fund several multi-disciplinary R&D teams. These teams are charged with experimenting with new methods and publishing their results in the peer-reviewed literature to establish our market position and the credibility of our methods. Our teams are exploring the use of Twitter, Facebook, and other social media platforms as a source of scientifically valid data.

Although this approach is more experimental and riskier than our traditional interdisciplinary model, it is nonetheless focused on establishing RTI's credibility with new methods of data collection and analysis. We pay for these projects through R&D funds created specifically for this purpose, allowing teams to experiment without the financial pressures of business unit revenue projections. In RTI's cost accounting system, having staff involved in these internal projects is "budget friendly" to line managers with profit/loss responsibility – it generates overhead to the business unit in the same way as externally funded research. Thus, a large component of the potential opportunity cost to the business unit is offset. However, we are fast approaching the transition point where our core competency of collecting data needs to be augmented, and eventually replaced, with a new competency of blending statistically representative (expensive) designed data with commodity (inexpensive) organic data while still drawing valid inferences suitable for public policymaking. Navigating this transition presents enormous challenges to our company.

Those challenges exist on two levels. The first level is almost existential -- how to address (not necessarily change) the fundamental view of data validity held by many RTI staff. This view insists that data must be sampled, collected, analyzed, and reported using traditional, proven statistical methods. Efforts to change this model are viewed with suspicion and deep-seated fears that the resulting data analyses will not produce consistent, reliable results. We are attempting to counter these fears by encouraging staff to publish their work in traditional journals, thereby establishing the credibility of the new methods. The second challenge is more mechanical but no less daunting. We need to learn how to merge data from disparate sources which were not originally intended to be joined together. This requires that we create different ways of thinking about how to synthesize data so that it can address the critical data integrity questions of causality and generalizability.

Our close relationships with area universities are also helping to address these challenges. Through means like visiting scholars and sabbatical-type exchange programs, we are able to access researchers working on some of the foundational issues around high-performance computing and management of voluminous data sets. RTI is a founding member of the National Consortium for Data Science, headquartered at the University of North Carolina at Chapel Hill, which also provides us access to staff and expertise from IBM, SAS Institute, General Electric, Cisco, and other large organizations with numerous relevant capabilities. These types of partnerships and relationships are helping us accelerate our transformation and refine our expertise.

Define What Clients Need – And Don't Need – From Big Data

RTI's third big data challenge is helping our clients use new, as well as existing, sources of information to envision research questions that, until recently, had been impossible to quantify. This newly developed approach will integrate designed and organic data with subject matter and research expertise to produce focused insights that can better inform policy and decision support, predict how resources can be used in a more cost-effective manner, and put information in the hands of end-users in a more efficient, customized manner.

Some long-standing federal government clients are beginning to make inroads in pursuit of these goals. For example, the National Institutes of Health announced this summer that it will fund up to $24 million per year for the next four years to establish investigator-initiated Big Data to Knowledge Centers of Excellence (National Institutes of Health, 2013). These centers are intended to help the research community use ever-larger and more complex datasets through development and distribution of innovative approaches, methods, and tools for data sharing and analysis. They will also provide training for students and researchers on data science methods. More broadly, RTI gained new insights into the big data capabilities and needs of our federal government clients at a recent event convened by the White House Office of Science and Technology Policy. Building on a $200 million Big Data Research and Development Initiative unveiled by the White House in 2012, the Obama Administration is now encouraging stakeholders including federal agencies, academia, nonprofit organizations, and state and local governments to participate in projects and initiatives that move big data from knowledge into action (Weiss & Zgorski, 2012). Key priorities include:

- Advancing technologies that support advanced data management and data analytic techniques
- Educating and expanding the data science workforce
- Developing, demonstrating, and evaluating applications of data that can improve key outcomes in economic growth, job creation, education, health, energy, sustainability, public safety, science, and manufacturing
- Fostering regional innovation.

RTI is at an early stage of developing its approach to big data, and we are deliberate about how to best invest our human and capital resources. One approach we are unlikely to pursue is creating a big data architecture or platform to support current and future projects. Instead, we tend to side with conclusions of the TechAmerica Foundation report that suggest that successful big data initiatives, especially in the public sector, are tailored to a specific, narrowly defined business or mission requirement (TechAmerica Foundation, 2012). RTI's experience in this arena supports an approach that integrates new and existing data to address focused research questions. For example, a research effort among RTI, the RAND Corporation, Structured Decisions Corporation, and the Washington, DC Metropolitan Police Department analyzed text data from 911 call transcripts to generate more precise forecasts of areas at elevated risk of specific types of crime. Collected by every police department in the U.S., 911 calls for emergency services data traditionally have been used to review the efficiency of response time and, among larger police departments, to better allocate resources. RTI researchers (with expertise in criminal justice and data analysis) developed a prototype software toolkit that can routinely process calls for services data and extract key characteristics or behaviors found in each call's narrative comments. This approach makes these data more flexible in helping to identify specific, short-term changes in a given location, such as an increase in the presence of certain types of drugs or disturbances. Using keywords or themes, the software's detailed information can help police anticipate and respond more quickly to changes in certain types of criminal activities or predict more precisely which areas are likely to see an increase in criminal activity. Big data will play an increasingly important role in projects like this, as the next generation of 911 computer-aided dispatch (CAD) technology will allow individuals to submit information captured by video and text message.

At an earlier stage of development is a project being co-funded by RTI and Duke University that is evaluating the effectiveness of massive open online courses, or MOOCs. Duke launched its first MOOC in conjunction with Coursera in September 2012 (Ferrari, 2012). Since then, Duke has offered more than a dozen courses and is developing additional MOOCs that cover humanities, natural and biological sciences, social science, nursing, medicine, and engineering. Properly designed and effectively presented, MOOCs can bring the benefits of a world-class education to motivated individuals with little more than an Internet connection. Nonetheless, assessing the effectiveness of online teaching approaches is an important consideration for MOOCs to live up to their potential. Big data's evaluation methods include in-depth tracking and analysis of online student learning activities, even down to the level of mouse clicks. This analysis can be performed across input generated from thousands of students instead of from data pulled from small studies (TechAmerica Foundation, 2012). RTI and Duke will not only be working on ways to mine and evaluate MOOC results but will also be conducting interviews with dozens of the largest employers in North Carolina to assess receptivity to hiring workers who have been educated using non-traditional methods.

CONCLUSION

As the challenges discussed above illustrate, RTI's workforce needs will be transformed by the skill-set demands of big data. We anticipate a blurring of the lines among the disciplines of mathematics, statistics, computer science, and various subject matter areas. Meeting the demands of big data will require internal changes that range from education and training to executive leadership. Based on RTI's experiences over the last six months, we offer for consideration some of our lessons learned thus far.

(1) Establishing a robust communication plan around our reorganization was not sufficient. We quickly found it necessary to form a cross-organization steering committee, led by a senior executive, to involve staff and leaders who were not directly impacted by the reorganization. We did this in part to correct perceptions that not being included in one of the new groups implied not having a role to play in RTI's big data future. Subject matter experts from across the company engaged quickly and enthusiastically to help shape the steering committee agenda, a process that expended more time than anticipated in meetings and discussions. Our decision to use a "big tent" approach to ensure broad buy-in has slowed us down.

(2) The data analytics talent shortage in other organizations is impacting us sooner than expected. RTI's corporate headquarters location of Research Triangle Park, North Carolina, is attracting new firms with a focus on analytics, and we have experienced voluntary departures from younger career staff who are realizing significant compensation increases. We are responding to these market forces, of course, but we clearly underestimated the degree to which our talent base would be targeted. We have increased our internal R&D funding and commissioned additional special projects aimed at keeping our top talent fully challenged and engaged.

(3) Our corporate information technology group, traditionally more of an "order taker" than a "business consultative" group, is being assessed and realigned under a new CIO. RTI's big data agenda is driving much of the assessment and revealing that our IT staffing mix may not be optimal for a big data world. We currently manage our own primary and secondary data centers, and will likely move toward outsourcing to private cloud vendors. Roles like "system administrator" or "data center manager" may need to be replaced by

"vendor management" or "business unit liaison." Such changes will not only impact our job titles but will also ripple into our cost accounting and charge-back methodologies, thus adding to the complexity of our transformation. IT's interconnectedness with the entire organization should not be underestimated.

As RTI girds for the big data revolution, we recognize that the challenges we face are similar in some respects to those facing our clients. That is, big data will usher in significant changes both to our organization and the clients we serve. We recognize the urgency of capturing and formulating insights from big data at a time when they can enhance optimal decision making, both for our organization and for our clients.

REFERENCES

Anderson N. 2013. Ex-Dodgers owner McCourt gives Georgetown $100 million to launch public policy school. *The Washington Post*. September 18. http://www.washingtonpost.com/local/education/ex-dodgers-owner-mccourt-gives-georgetown-100-million-to-launch-public-policy-school/2013/09/17/0e19e3dc-1f96-11e3-8459-657e0c72fec8_story.html (Accessed April 1, 2014)

Christensen CM. 2013. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Reprint edition. Harvard Business School Press, Cambridge, MA.

Comstock J. 2013. Ringful Health launch app to help patients compare drugs. Mobilehealthnews, http://mobihealthnews.com/25058/consumer-reports-ringful-health-launch-app-to-help-patients-compare-drugs/ (Accessed April 1, 2014)

Davenport TH, Patil DJ. 2012. Data scientist: The sexiest job of the 21st Century. *Harvard Business Review* 90(10): 70 – 76.

Ferrari E. 2012. Duke to offer free courses on internet. *Duke Today*. July 16. https://today.duke.edu/2012/07/coursera (Accessed April 2, 2014)

Groves RM. 2011. Three eras of survey research. *Public Opinion Quarterly* 75(5): 861– 871.

Kerr S. 2013. Georgetown receives $100 million to create new public policy school. Press Release, Georgetown University, Sept. 18. http://www.georgetown.edu/content/mccourt-school-press-release.html (Accessed April 2, 2014)

Manyika J, Chui M, Brown B, Bughin J, Dobbs B, Roxburgh C, Byers AH. 2011. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation (Accessed April 1, 2014)

National Institutes of Health. 2013. NIH commits $24 million annually for big data centers of excellence. Press Release, National Institutes of Health. July 22. http://www.nih.gov/news/health/jul2013/nih-22.htm (Accessed April 1, 2014)

Rooney B. 2012. Big data's big problem: Little talent. *The Wall Street Journal*. April 29. http://online.wsj.com/news/articles/SB10001424052702304723304577365700368073674 (Accessed April 2, 2014)

TechAmerica Foundation. 2012. Demystifying big data: A practical guide to transforming the business of government. TechAmerica Foundation. October 3. http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf (Accessed April 2, 2014)

Weiss R, Zgorski LJ. 2012. Obama administration unveils "big data" initiative: Announces $200 million in new R&D investments. Press Release, Office of Science and Technology Policy, Mar. 29. http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf (Accessed April 2, 2014)

TIM J. GABEL

Executive Vice President of Social, Statistical, and Environmental Sciences
RTI International
E-mail: tjg@rti.org

CATHY TOKARSKI

Senior Health Policy Writer
RTI International
E-mail: ctokarski@rti.org