Big Data, Big Needs

Meeting the Societal Needs for Data Science and Cybersecurity Skills
Executive Summary

The need for data-fluent talent has become one of the definitive jobs issues in our country and globally, and touches all levels across all industries. Unlocking the power of data has the potential to drive tremendous growth for our economy, and the explosion of business-related data reveals a growing need for talent that can capitalize on the opportunities buried in the numbers. This paper explores this pervasive skills gap and the partnerships between corporate America and post-secondary education required to address the gap, based on discussions held during the “Big Data/Open Data: Inspiring Corporate & Civic Innovation” conference held March, 2016 featuring over 20 experts from Fortune 500 business, higher education and government, and hosted by University of Phoenix.

The world is creating data at a tremendous pace — over 5 zettabytes of data is available today and that figure is projected to grow to over 44 zettabytes by 2020. This data comes from a wide variety of sources including government, customers, social media, and data brokers. This wave of data presents enormous opportunities, but only for organizations that have the will and the talent to understand and utilize the data. This talent shortage is pervasive across sectors, not just in companies who are easily identified as technology companies.

How do we make sure that a diverse talent pool is ready to take advantage of these continually growing opportunities? Data scientists are needed to transform this data into actionable knowledge; however, there are not enough traditionally trained analysts available, and they lack the real-world experience necessary to quickly identify the business opportunities. This has created a need for data literacy throughout the enterprise. We must add data knowledge to the list of ‘basic skills’ for all Americans, but unlike reading and mathematics, this new technology need is changing and expanding at lightning speed.

For some positions, this is a technical skills issue — learning the tools used in data analytics and presenting the data in a way that is easy to understand. For others, the skill set is much more analytical — using predictive analytics to identify trends and then communicating a business case for change to others. All the while, security remains a primary concern to protect any individual data involved. In whatever way you examine this issue, our workforce, perhaps more than any other, will have to continue to learn on the job. The once distinct lines across technical and business skills, including management, have now blurred.

Higher education cannot meet this challenge alone. Robust partnerships between industry, education and government are necessary to unlock the power of big data. University of Phoenix is collaborating with industry partners to build solutions to help develop a pipeline of data-literate talent and working with companies to upskill their workforce with data-informed decision making. The idea of using data as a disruptor is reimagining industries every day and creating endless opportunities for those who have the knowledge and ability to harness the power of open data to invent the next big idea. Together we can create that diverse talent pool, build stronger companies and create a more robust economy.
Big Data, Big Needs
Meeting Business Needs for Data Science and Cybersecurity Skills

Thanks to growth of digital devices and the Internet, the world is creating data at an amazing rate. IDC Digital Universe reports that we have a wave of new data available to us — over 5 zettabytes of data available today and that the figure is projected to grow to over 44 zettabytes by 2020. A zettabyte is $1000^7$ bytes, or 1 trillion gigabytes. If you were in the business of storing information, you can immediately see the implications. But what are the implications for other businesses, and specifically, how do these businesses capitalize on the potential that this data brings?

This paper will examine how a company’s skills and human capital affect its ability to gain a competitive advantage from this explosion of data. These skills bridge the traditional division between business operations and information technology. Data analytics can and should happen at all levels in the business, and employees need to have the ability to identify and exploit the opportunities that are available. This quickly-widening gap in employee skills can only be filled by robust partnerships between business and education.

The conclusions of this paper are based on discussions held during the “Big Data/Open Data: Inspiring Corporate & Civic Innovation” conference held March, 2016 featuring over 20 experts from Fortune 500 business, higher education and government, and hosted by the University of Phoenix.

The Proliferation of Data in Business

The 44 zettabytes of data expected in the world by 2020 represents over 5200 GB of data for every man, woman and child on the planet. This amount of data in the world today is not only huge, it is also growing at an ever-increasing rate. By some accounts, over 90% of the data existing in the world today was created in only the past two years. The wide variety of types and sources of data helps to expose the fact that this data will impact nearly all competitive businesses, not only “technology” companies.¹
Data Sources are Growing Rapidly

The number of sources of business, government and other data is growing rapidly, as is the amount of detail that they contain. Here are a few examples of data sources that are now being utilized by industry:

**Government Databases**

Government is one of these sources of useful data. Whether being commercialized, such as the NOAA data that was used in the creation of the Weather Channel, or used to inform and influence policy, federal and state government websites offer a treasure trove of useful information. There are 80,000 databases publicly available from 75 agencies. According to Dr. Erica Groshen, Commissioner of Labor Statistics for the U. S. Bureau of Labor Statistics (BLS), the BLS now provides data through 300,000 microsites and data interfaces. The BLS is increasing the rate and accuracy of data collection through computer-assisted online collection techniques. And while agencies like the BLS continue to inform the public on the availability of information, still too few really understand what is there, how to access it and how to use it.

**Internal Enterprise Data**

Business enterprises are collecting their own stores of data on customers, inventory, suppliers, employees and others. Every sales transaction that is done online or using a credit card creates a trail of data that can connect an individual customer to the products that he or she buys. For instance, grocery stores collect data on the individual items purchased on each “basket” or sale. Not only does this reveal what the customer buys, but also what items are bought together. For instance, if a person buys lunch meat, how often do they also buy cheese in the same basket? This information is being used to determine when to offer certain coupons at checkout to entice the customer to buy an item he/she may have forgotten to buy.

**Internet and Social Media Data**

Additional data on individuals can be “scraped” from information available online on websites (such as phone numbers of employees of a company shown on a company directory) or on social media (such as employment history of individuals on LinkedIn). This information can be used to determine the probability of purchasing certain products or services. For instance, if you recently changed jobs, do you need new work clothes?

Web browsing data also can be mined for additional information. You may have noticed that searching for products on the Internet or even on websites like Amazon can result in seeing advertisements for similar products on other websites that you visit, such as Facebook.
Sensor and Partner Data

Additional data is created by devices used by individuals or businesses. Overnight shipping services such as FedEx or UPS track packages several times a day as they move from one truck to another. Cities continuously monitor the location of buses so that users can predict down to the minute when a bus will arrive at a particular location. Some airlines track the movement of luggage at each point in the process so that passengers boarding a flight can know the moment that their luggage is placed in the plane. Similar processes can also be used by industrial partners so that an auto assembly plant can monitor in real time when parts for particular cars will arrive from each supplier.

Data Brokers

Data brokers take much of the data from these sources and others and aggregate them into data lists that nearly any business can buy. Some of these are from well-known sources, such as credit agencies, voting records or driver’s licenses. Other data comes from sources that consumers may not be aware is being sold, like credit card purchases, charitable contributions, paycheck information and store loyalty cards. The result is that you can easily buy lists of individuals who have a given income, are looking to buy a certain type of car, who have children of a given age, or have a large number of contacts on social media.²

Data Analysis is Creating Opportunity in a Wide Range of Industries

The variety of data available demonstrates that big data is not just the realm of scientists or technology companies. Intelligent analysis of this data is increasingly a competitive advantage for retail, healthcare, finance or even entertainment companies.

Amazon is well-known as the world’s largest online retailer. It got to this position through the intelligent use of data analytics. Amazon grew in sales and popularity by intelligently recommending new purchases to customers. These recommendations came from a sophisticated analysis of past purchases by both that individual and by similar customers.³ It has such a large store of information and computing power that it now has positioned itself as provider of data hosting and processing services itself. Walmart has recently acquired online retailer Jet.com in order to gain additional customer data and challenge the position of Amazon.

Other industries such as insurance also find competitive advantages through data analytics. Some health insurance companies are using data to encourage healthy lifestyles, such as monitoring physical activity through employee feedback or even wearable devices.⁴

Others are using data analytics to create essentially new products and services. Lori Sherer of Bain & Company, Inc. cites Progressive’s Snapshot tool as “a great example of a data-driven product innovation.” Progressive realized that certain driving behaviors were good predictors of the total cost of auto insurance for individual drivers. Therefore, Progressive created a physical
device that users plug into their car to collect information on driving speed every second, which is relayed back to Progressive’s data centers and analyzed to find the frequency of hard braking events. This data is used to offer lower insurance rates to those drivers that exhibit good driving habits. By 2014 Progressive had collected data on over 10 billion miles of driving.\(^5\) This trove of data provides Progressive a much better understanding of its customer’s driving habits, allowing it to segment and price customers more effectively.

Even the entertainment industry can find value in big data analytics. Netflix has over 83 million subscribers who watched more than 42 billion hours of video in 2015.\(^6\) Information gleaned from viewing habits helped Netflix expand from just a distributor of content to a content creator. Using this data, it was able to predict the success of new shows before they were even filmed. For instance, it knew that a significant number of their subscribers regularly watched films starring Kevin Spacey, and also works from the director David Fincher. It also knew that the British series “House of Cards” performed well. From these three points of data, Netflix determined that an American House of Cards series starring Spacey and directed by Fincher would certainly be a hit. It ordered the entire series to be filmed at one time without the need for a pilot episode. It also uses its data to promote the series by showing different advertisements to fans of Spacey, Fincher, or other actors in the series.\(^7\)

Small businesses can also gain significant impacts from big data. A beach house rental company had data records on rentals for a number of years, but was not able to use it effectively. Through data analysis, the company was able to determine patterns in rental rates based on week of the year and the size and location of the home. Increased rental rates produced a positive return in one year. The local zoo in Tacoma, Washington was also able to use past records to accurately predict daily attendance and adjust staffing levels accordingly.\(^8\)

**Characteristics of Useful Data**

As the amount of data grows rapidly, businesses need to be able to identify and collect useful data. Dr. Groshen states that “useful data is the lifeblood of the economy,” and stresses that there are five characteristics of useful data:

- **Accurate** – The data must be correct. If inaccurate data is used to make decisions, then the decisions can have significantly negative consequences.

- **Objective** – The data should not be highly subjective. Even data that is individually subjective, such as the opinions of an individual customer, become objective data when it is averaged over a representative number of customers.

- **Relevant** – The data should be pertinent to the question at hand. Unnecessary data can make the analysis slower or less accurate.
• **Timely** – The data should represent the situations that currently exist or that can be applied to the current situation.

• **Accessible** – The data must be available for analysis. Thanks to modern tools, even unstructured data such as open-text customer feedback can often be made accessible to quantitative analysis.

The cases shown above demonstrate that data analytics are now forcing dramatic changes in the way the businesses make decisions and create new products. Sherer points to five forces ensuring that data analytics will continue to fuel innovation for decades to come: a magnitude increase of data generated; new business models that are founded on data analysis; an increase in investment in the data collection and analysis infrastructure needed to leverage data; the proliferation of multi-disciplinary math and computer science programs in education that enable employees to perform data analysis; and a generation of digital and data-enabled consumers and business people that are willing to use data effectively.

**The Impact of Data on Business Skills**

Although businesses possess an abundance of data, Lisa Dodson, Manager of Data Management Practice at SAS observed that the average organization only utilizes 10% of this data. How do we get the data and use more of it? With more data being generated every two minutes as was created from the beginning of time all the way to the year 2000, this is now the challenge that most organizations face.

**Data Analytics Requires Change from the Top**

The key is the ability to both understand the data and to put it to use by integrating it into business decisions. Doing this requires access to quality data, the knowledge and tools required to make useful conclusions from the data, and the organization and willingness to utilize the data.
During the recent Big Data/Open Data conference in Dallas, the panelists all agreed that the CEO drives the value of data in business strategy. Starting from the top is still the most successful way to integrate an effective data utilization strategy. At the end of the discussion, Ruth Veloria, Executive Dean of the School of Business at the University of Phoenix asked the panelists to give a sign of company maturity in regards to data strategy. “It’s about demystifying the data by making it more accessible and democratized in data we trust,” added Pawan Divakarla, Data and Analytics Leader at Progressive Insurance. Cortnie Abercrombie, Emerging Roles and Markets Leader, IBM Analytics and Zulfikar Sidi, Vice President-Enterprise Data & Analytics, Sabre both agreed there are three major indicators—culture, process, and technology. Is data embedded in the organization beginning at the CEO level? Are there processes to utilize the data and cultivate career paths for the data employee? Is the company innovating and enabling itself to leverage the power of big data?

Because data has changed the workforce, the panelists agreed that employees need more than data analysis abilities. “We need good storytellers with deep technical knowledge,” said Sidi. Linda Vytlacil, Vice President-Global Customer Insights, Walmart also stressed the importance of providing these employees with “a tribe, a way to learn, reward system, and a way to unlearn old behaviors.”

These changes have to flow from the top all the way through the organization. While the C-level enables the organization to utilize data, the actual discovery and implementation is initiated through all lower levels and departments. Sherer presented an example of how retailers now combine data with marketing, logistics and customer service. Many retailers use celebrity endorsements to promote products on TV, print and social media. Today these
stars are selected for their appeal to specific target audiences that the retailer wants to attract using data on the star’s previous sales, audiences and followers. These audiences are then cross-referenced to the particular products that they buy. Stores, warehouses and suppliers are alerted ahead of time to increase stocks of potential low-stock items. Following the endorsement, consumer surveys and social analytics are used to measure the effect of the celebrity on sales and attitudes.

These functions require employees throughout the organization to manage potentially huge amounts of data. The data is no longer just the property of a central IT organization. More importantly, the different business functions are in better positions to know the correct questions to ask.

Data Scientists are in Short Supply

There is a critical shortage of educated data scientists & cybersecurity specialists, from certificates to Ph.D’s

The problem is compounded by a lack of enough trained data scientists to fill all the needs for data analytics across the country. The lack of skilled IT and data analytics talent has become the headline of the shortage of STEM workers in the U.S. Recent surveys reported by Forbes point out that four in ten (43%) companies report their lack of appropriate analytical skills as a key challenge, but only one in five organizations has changed its approach to attracting and retaining analytics talent. The recruiting website Indeed.com currently lists over 50,000 job openings in big data and IT. Salaries for junior level data scientists top $90,000 with experienced professionals over earning $200,000. As a result of the scarcity of
data scientists, 63 percent of the companies surveyed are providing formal or on-the-job training in-house. “One big plus of developing analytics skills among current employees,” says the report, “is that they already know the business.” These companies are also doing more to train existing managers to become more analytical (49%) and train their new data scientists to better understand their business (34%). Still, half of the survey respondents cited turning analytical insights into business actions as one of their top analytics challenges.9

Citizen Data Scientists Combine IT and Business Skills

Business advantage is now obtained by extracting business value from data. No longer are business decisions based just on the intuition of a few senior employees, but instead are based on the logical analysis of large sets of data. This requires a new skill set for a widening number of employees.

The need for employees with a combination of technical and business skills combined with the shortage of classically-trained data scientists has now led to the rise of the “citizen data scientist.” A recent Gartner report defines a citizen data scientist as “a person who creates or generates models that leverage predictive or prescriptive analytics but whose primary job function is outside of the field of statistics and analytics.” It could be a line-of-business role, a business analyst, or a member of the business intelligence or IT team. The defining trait is that statistics and analytics are secondary in the role.10
Organizationally, data analysts and data scientists can occur at any level. Nearly all employees lack the skills not only to analyze this amount of data, but also to simply be able to understand the limits of the technology and the results presented. This leads to skills gaps that must be addressed systematically throughout the organization.

This is supported by studies by Tata Consultancy Services (TCS) that show that the business functions of sales and marketing are expected to be the biggest users of data analytics, while finance and logistics are expected to see the largest ROI on analytic investments. In fact, the surveys from the study showed that the biggest challenge faced by companies implementing Big Data initiatives was getting different business units to share information across silos.

While it is imperative that data analysis occur throughout the organization in order to gain the highest ROI, there is not a single method of management of the effort that meets the needs of all. Sherer described advantages and disadvantages of centralized and de-centralized management of data analysis personnel as shown below, with centralized efforts running a risk of being disconnected from business units, whereas de-centralized efforts can have difficulty in effectively attracting and utilizing talent.

---

**Potential Issues Facing Different Models of Data Analytics Management**

- **Centralized**
  - Risk of “Science Experiments” without business impact

- **Decentralized**
  - Risk of redundancy & incrementalism; hard to attract sufficient talent

- **Embedded COE**
  - Complicated matrix to manage effectively

---

*Image credit: University of Phoenix*
Panelists at the Big Data/Open Data conference agreed that a blend is required to best take advantage of both talent and opportunity. “It depends on how much progress we’re making on deploying analytics,” stated Vytelacil, “Data analysis is very collaborative in nature. You have the process of ‘breathing in’, centralizing people closer to data, which is followed by ‘breathing out’, regrouping and moving back into business and applying insights and codifying decision processes into tools.”

Reimagining Talent Management in the Era of Big Data

How do we make sure that the wave of data doesn’t overwhelm us? How do we make sure that a diverse talent pool is ready to take advantage of the anticipated career opportunities in this field? How do we take advantage of the big data available to us now and in the future?

Data scientists are needed to transform this data into actionable knowledge, and the organization needs to attract, train, deploy and manage these talents in a sustainable way.

Dr. Joe Hill, Chief Technologist for Analytics at Hewlett Packard states that “Big data was 2.0. We are now at 3.0, where rapid insights are providing business impact. If big data and data analytics are going to be anything more than temporary fads, then they have to act like adults. They need to be operationalized. They need to be embedded in the business to have impact. So they need to be managed and monitored.” He concludes that “In business, ‘real-world’ means everyday business processes and decisions. So operational analytics is about bringing magic to everyday processes. It is how we empower the data-driven enterprise to derive value from all the relevant data in order to produce superior business outcomes.”

In addition to data scientist careers, the pervasive nature of data has raised the bar for the general population. It is estimated that the number of ‘citizen data scientists” will grow five times faster than the actual jobs in this space. We must add data knowledge to the list of ‘basic skills’ for all Americans but, unlike reading and mathematics, the world of technology is changing and growing at lightning speed.

For some jobs this is a technical skills issue. Those jobs require learning the current technology and software that is broadly used in data analytics and becoming proficient with inputting and presenting the data in a way that is easy to understand. For other jobs the skill set is much more analytical—being able to use predictive analytics to identify trends and model future behavior based on prior trends. Embedded in these jobs is the ability to communicate in language as well as numbers, to be persuasive while allowing the data to make the case, using negotiating skills to stress the importance of specific indications and making a business case to influence thinking. Technology has made the data more understandable to a broader base in companies so that there has been a blurring of the lines between job categories that has created a need for data analytics and manipulation in new business lines.
Data scientists also need to be aware of privacy issues, and methods to protect sensitive data they may be using. Tim Patrick, Dean of Technology, University of Phoenix College of Information Systems and Technology said, “Because of technology and data collection advancements, individuals are monitored more closely, which can be frightening. It is important for organizational leadership to be in compliance because they are not always doing that. This can create a crisis.” Organizations should rethink what sources they are using and how it affects individuals. We have a responsibility when it comes to data - to bring a story to the forefront in an ethical manner. Patrick posed a powerful thought, “What happens when we know everything that is going on?” Big, open data will continue to affect society at a growing pace, and all parties involved must be aware of its effects.

Companies engaging in big data projects will therefore need to employ individuals specialized in data security, in addition to providing data security skills to many in their general workforce. There is a critical skills shortage in cybersecurity, and as personal data becomes more prevalent these jobs are going to continue to grow. These jobs need people who have critical thinking skills and who can think outside the box to match the way that hackers are thinking, using multiple entry points to gain access to protected data sets.

In whatever way you examine this issue, talent recruitment and retention remains a growing concern for employers. Previously, employers poached talent from one another, but today more are looking to new providers to both help recruit new talent and grow their incumbent workforce. The data-driven workforce —perhaps more than any other — will have to continue to learn on the job. The once distinct lines across technical skills throughout businesses, including management, have now blurred. As software changes and systems are retired, employers need employees who can continue to learn new skills. Industry recognized credentials, new integrated degrees and customized real time training will become the norm.
Talent management must recognize the need to provide skills training throughout the enterprise. However, this should not be a one-size-fits-all approach. Some employees, particularly more recent hires, may already possess a certain amount of data fluency – they only need to be “upskilled.” Others need a more comprehensive approach to data training – the need to be “skilled.”

This need for data-fluent talent has become one of the definitive jobs issues in our country and globally, and touches all levels in all industries. Higher education cannot meet this challenge alone. Robust partnerships with industry will help meet it – cross cutting needs and more intimate partnerships with specific firms to assess their own maturity as a company taking full advantage of the promise of using data. Embedded in this is the change management capabilities of moving through that maturity process, the need for faculty and programs to be as current and agile as possible and for all to understand that this is a world of continuous skill development that must be synchronized with the industry clock.

The challenge to meet business demand for big data and cybersecurity talent has become a critical issue for CEO’s, CIO’s and CTO’s at organizations across America. These leaders need to take action, but also need help at all levels of their organizations and in education and government. This work can only offer the greatest potential if we do it with partners — education, corporate, government, non-profit — all looking for creative solutions to complex problems by harnessing the power of data. Together we can create that diverse talent pool, develop stronger companies, and build a more robust economy.
Enhancing Data Talent with the University of Phoenix

In today’s environment, the skills required for data analysis are merging the jobs of traditional business and information technology. People across the organization need fundamental knowledge of analytics. Data analytics is not just an IT problem anymore. The leadership to acquire and use these skills must start at the top. Business leaders must look at their organization and decide who to skill (those who currently have few data skills) and who to upskill (those who have dated analysis skills or limited experience). How will you address these issues in your enterprise? To learn more about solutions to fill talent gaps, contact the Employer Partnership Group at 866.955.5515.

About Employer Partnership Group

Employer Partnership Group is a division of University of Phoenix that specializes in working with public, private and government organizations to provide talent management solutions for employers designed to create a higher performing workforce.