Global Journal of Business Research
Vol. 9, No. 4, 2015, pp. 77-89
ISSN: 1931-0277 (print)
ISSN: 2157-0191 (online)

IBFR
www.theIBFR.com

DATA ECONOMY DIMENSIONS

Peter Géczy, National Institute of Advanced Industrial Science and Technology (AIST)

ABSTRACT

Data is the 'new oil'. Data has been widely regarded as the main innovation driver in information technology and several other segments of developed economies. Inherent value of data has a significant transformational power. Realization of value of data drives the rapidly expanding data-driven technology sector. Data is among the most prized assets of not only data-oriented technology companies but also governments and individuals. Data plays a central role in a growing spectrum of economic activities. Data economy has manifested its effects along a number of dimensions—particularly, trade, labor, education and government. However, to realize the full potential of data and manage its transformational power, it is necessary to understand the value of data, its influential spectrum, and enablers facilitating social and economic benefits. Despite the growing importance of data economy, there is a considerable lack of studies encompassing the associated range data oriented economic activities and their analysis. This work examines the dimensions of contemporary data economy and presents a pertinent encompassing perspective.

JEL: A10, E24, E66, F23, J21, J26, J31, K10, M15, M51, O11, O14, O15, O33, O38.

KEYWORDS: Data Economy, Trade, Labor, Education, Government, Data-Driven Innovation, Data Science, Data Engineering, Data Products, Data Protection, Privacy, Big Data, Data Management

INTRODUCTION

Digital data production has been experiencing an exponential growth over the past decade. This growth is projected to continue during the following ten years (Gantz and Reinsel, 2012; EMC, 2014). At the beginning of the twenty-first century, many advanced economies reached the digital dominance (i.e. digital data surpassed analog data). The next milestone is expected to occur in 2017—when emerging markets shall outgrow mature markets in data production (EMC, 2014).

The digital dominance has brought both benefits and challenges (McAfee and Brynjolfsson, 2012; Charlesworth, 2009). Digital data is relatively easier to store and mange than analog data. There is also a greater range of tools for digital data processing. The main tools have become digital computers, database systems and a broad range of devices in the domains of information technologies (Kroenke and Auer, 2013). Digital data is also easier to produce—given the contemporary technological advancements. A multitude of devices ranging from large-scale systems, throughout mobile devices, to miniature sensors can produce rapidly large amounts of digital data.

Ease and speed of digital data production has brought several challenges. Data is being produced at a rate faster than we can process and store it. The production outpaces the processing (Gantz and Reinsel, 2012). Contemporary digital data expansion is disproportional to the progress in data processing technologies (EMC, 2014). Data grows overly faster not only in volume, but also in diversity, complexity and other aspects.

The discrepancy between growing aspects of data (such as volume, diversity, complexity, etc.) and processing capabilities has led to big data problems (Géczy, 2014; Fan et al. 2014). The big data problem underlines the gap between the required and available processing capabilities. If such gap occurs in organizations, it needs to be addressed promptly and appropriately. Otherwise, it leads to diminishing abilities to utilize the potential of big data.

Despite challenging big data problems, organizations have been eager to realize the transformational power of data (Redman, 2015). Data can provide benefits for organizations both internally and externally (Provost and Fawcett, 2013). Suitably collected and analyzed organizational data may deliver better perspective on internal functioning of organizations. Utilizing the insights from data, organizations can effectively address the issues of operating efficiency and productivity.

Attempts at realizing economic potentials of data have initiated various commercial and technical activities (Hurwitz and Kaufman, 2015). The activities include data collection and storage, data analysis and processing, actionable knowledge extraction, design of data products and services, direct and indirect data monetization, etc. Data collected and engineered for external monetization provides opportunities for organizations to expand their revenue streams. The spectrum of economic activities is expanding as organizations are exploring innovative ways to gain strategic advantages (McCallum and Gleason, 2013). The scale and scope of data oriented economic activities are bringing about notable effects on economy.

Data oriented economic activities are playing increasingly important role in developed economies. As the importance of data rises, so does the appetites of commercial organizations and governments to collect, control, and use more data for their own benefits (Zwitter, 2014). These benefits are not necessarily aligned with the interests and rights of individual users and citizens. Many citizens around the world regard this vast data collection as an intrusion of their privacy (Cate, 2011). Governments and regulatory bodies must progressively intervene in order to establish balance that stimulates economic growth and protects the rights of citizens and individuals.

The manuscript is organized as follows. The literature review section is followed by the 'Dimensions of Data Economy' section. It presents an encompassing perspective on data related economic activities and highlights four influential dimensions of data economy. The dimensions are elucidated in the following sections. The section 'Trade Dimension' addresses the issues of value of data and its monetization. The next section, 'Labor and Education Dimensions', illuminates these two closely interrelated aspects of data economy. Shortage and high demand for data professionals affect labor dynamics and stimulate expansion of educational programs. The following section, 'Government and Regulatory Dimensions', exposes the need for regulating the forces of data economy both nationally and internationally. The presentation concludes with a concise summary of the essential points.

LITERATURE REVIEW

Data is the most valuable asset in information and data economy (St. Amant and Ulijn, 2009). Data oriented economic activities have been progressively evolving from development of technological solutions, throughout provision of services, to maximizing the value of data. Information technologies and data services have been rapidly expanding. Data services are gaining a solid ground in the dominant economic sector of contemporary developed economies―service sector (Bryson et al., 2004).

Value of data has been rising and organizations have been realizing it (Lievesley et al., 1993). Extracting value from data has become the target for many organizations. They have started collecting vast volumes of data about their operations, suppliers and customers. Large operational data can be explored for gaining insights into functioning of organizations and increasing operational efficiency (Géczy et al., 2007 and 2008). Data analytics have been employed to attain competitive advantage for organizations (Davenport et al., 2007 and 2010). Actionable knowledge extraction methods have been utilized for improving core competencies (Laursen and Thorlund, 2010).

Collecting and managing bid data by organizations have led to various challenges (Wigan and Clarke, 2013; Klein et al., 2013; Malik, 2013; Walsh et al., 2012). Data needs to be stored and appropriately processed as it grows (Tallon, 2013). Rising demands for processing power, storage and management of big data have led to development of scalable and distributed technologies (Frischbier and Petrov, 2010). However, data and its complexity have been expanding faster than capabilities of organizations to manage it. This has led to emergence of big data problems (Géczy, 2014; Gantz and Reinsel, 2012).

Persistent big data problems have been raising concerns about viability of data oriented strategies (Jacobs, 2009). Critics of data orientation have argued that big data presents more paradoxes than potentials (Richards and King, 2013). These arguments have gained a substantial strength with the major failure of the Google Flu Trends project (Lazer et al., 2014). The project has been regarded as a prime example of the power of data. Insights derived from data should have been a more useful indicator of flu spread than the government statistics. However, the insights completely missed the largest swine flu pandemic.

Shortage of qualified data professionals has presented additional challenges for organizations in realizing the potential of data (Bakhshi et al., 2014; Manyika et al., 2011). Occupation of a data scientist has been labeled as 'The Sexiest Job of the 21st Century'—due to high demand and attractive remuneration (Davenport and Patil, 2012). However, qualified data professionals are difficult to find. They are required to have advanced multidisciplinary skills (Bakhshi et al., 2014). The skillset should incorporate advanced technical skills, soft skills, and relevant domain knowledge. Universities have seized this opportunity and have developed advanced degree programs for training new data scientists.

Data–as the central asset in data oriented economic activities–needs adequate protection and security (Garber, 2013; Tallon, 2013; Anthes 2010). The appropriate measures must extend to both organizations and individuals internationally (Cate, 2011; OECD, 1980). Organizations should keep their data protected and safe from misuse by governmental and/or commercial entities (Wei et al., 2015). Individuals should have their privacy respected and properly protected by legislative measures (Margulis, 2011; Lanois 2010). These challenges require involvement of government and regulatory bodies.

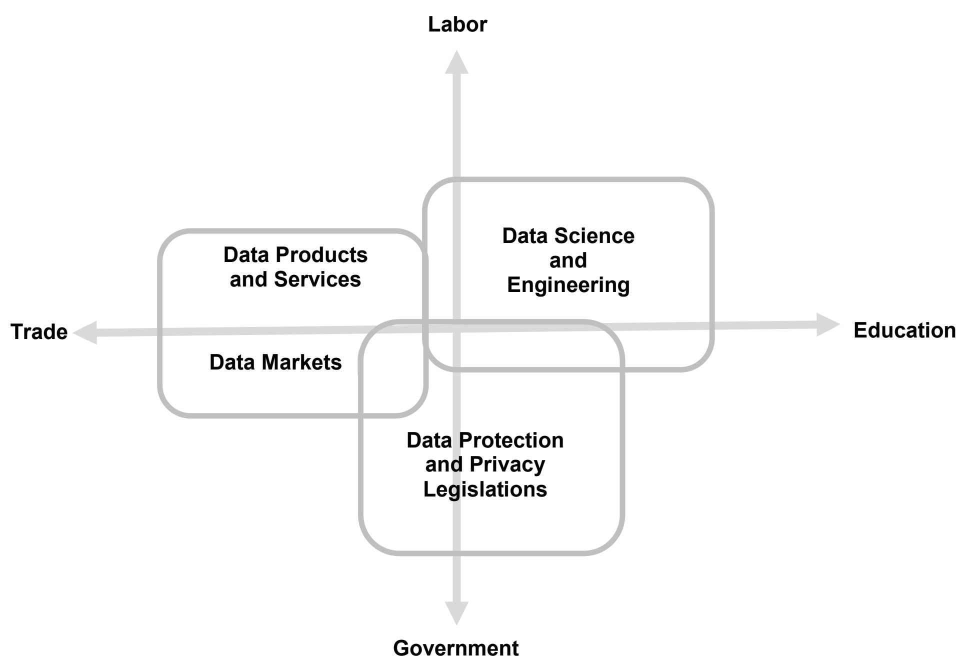
DIMENSIONS OF DATA ECONOMY

Digital revolution and rapid expansion of information technologies have brought both benefits and drawbacks. Digitization of data has allowed efficient data management and automation of data processing. However, it has also led to extensive rise of digital data over the past few decades (Gantz and Reinsel, 2012). Explosion of digital data and inadequate progress in data processing technologies have resulted in big data problems (Géczy, 2014). The big data problems have been presenting notable challenges to organizations (Buhl et al, 2013; Hunter, 2013; Klein et al., 2013; Walsh et al., 2012). Although the big data meme has provided the initial visibility of emergent data economy, several factors have played important roles.

There are three significant factors contributing to the rise of data economy: digital data dominance, unprecedented data growth, and value of data. It is estimated that at the beginning of the twenty-first century digital data outgrown analog one (EMC, 2014). Digital data has been expanding exponentially―pressing organizations to face the new data challenges and adopt data-oriented strategies. This has led to development of novel technologies, markets and economic realization of inherent value of data. These issues have commenced the data economy that has been evident along several dimensions.

Data economy has been influential in four primary dimensions: trade, labor, education and government (as illustrated in Figure 1). Along the trade dimension, contemporary data-oriented economic activities revolve around data products and data markets. The labor dimension underlines a high demand for data scientists and engineers. Presently, there is a lack of qualified labor to fill these positions. Increasing demand for data professionals stimulated various educational activities. Universities and commercial organizations have been developing novel educational programs to train new generation of data professionals. Academia has also been approaching novel data-oriented challenges from research and development angles. Growing spectrum of data-oriented economic activities has also attracted involvement of governmental and regulatory bodies. Their involvement has emerged from the need to regulate data economic activities and provide suitable grounds for adequate data and privacy protections.

Figure 1: Illustration of Four Essential Dimensions of Data Economy



Data-oriented economic activities span across four major dimensions: trade, labor, education and government. In the trade dimension, data products and data markets are the primary constituents of economic activities. In the labor dimension, the economic activities are driven by a high demand for qualified data scientists and engineers. Lack of skilled data scientists and engineers is a motivating force for educational and research activities in data science and engineering. Governmental involvement emerged from the need to regulate data-oriented economic activities and provide baselines for adequate data protection and privacy.

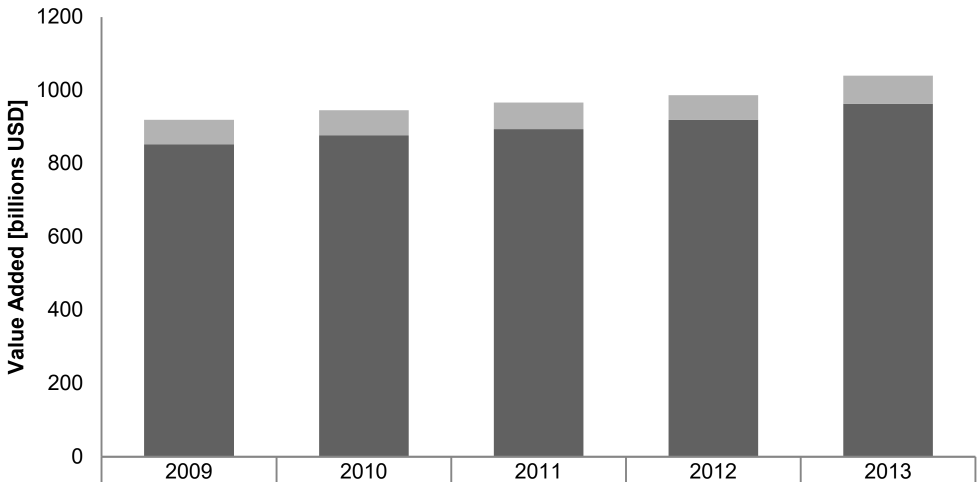
TRADE DIMENSION

Data has a value that can be monetized. Value of data is the enabler for direct or indirect monetization and underscores the trade dimension of data economy (Vertesi and Dourish, 2011). Data markets and data products are two main facilitators for monetizing data. While data markets serve as platforms for direct data exchange, data products allow for indirect monetization of data. Data oriented economic activities increasingly contribute to GDP of developed economies. In the United States, value added by data processing industries represents approximately eight percent of GDP contribution by Information and Communication Technologies (see Figure 2).

Data markets are platforms for direct monetization of data. They offer systems and tools for mediating economic transactions between one or more economic agents (individuals, businesses, governments, etc.).

Economic transactions are generally carried out in a medium of exchange—financial quid pro quo. However, there are also data markets focused on data sharing that facilitate barter transactions, i.e. limited or unlimited direct exchange of data without using a medium of exchange. Economic transactions on data markets rely on exchange of value inherent in data.

Figure 2: GDP Value Added by Data Processing and Information-Communication Technologies (ICT)



	2009	2010	2011	2012	2013
Data	66.5	69.5	72.7	67.3	77.5
ICT	852.4	876.3	894.1	919.1	962.5

Information-Communication Technologies (ICT) and Data Processing economic activities increasingly contribute to GDP of the U.S. During the five-year period 2009-2013, data processing economic activities represented approximately 8% of ICT value added to GDP. [Source: U.S. Bureau of Economic Analysis (BEA), Release Date: November 13, 2014]

Value of data is largely determined by data providers (Chatain, 2011). Presently, there are no commonly agreed upon data valuation methodologies. However, a number of quantitative and qualitative factors are taken into account. For instance, volume and richness of data, completeness or incompleteness of data, accuracy or inaccuracy of data, level of contamination or noisiness of data, timeliness of data, etc. Data has also a significant potential for indirect value creation.

Prominent potential for indirect value creation by data is evident in several ways (Manyika et al., 2011). For instance, transparency creation: making relevant data easily accessible across otherwise separate departments can sharply reduce search and processing time, and enhance productivity; segmentation: large and complex organizational data can be suitably segmented according to desired characteristics for improved decision making; customization: specific user characteristics or product features observed in data enable tailoring of products and services precisely to meet those needs. Value created by data is in symbiosis with value of products and services utilizing given data.

Data products and services facilitate monetization of indirect value creation. That is, data is not monetized straightforwardly, but rather products and services built upon data are monetized. Data is symbiotically linked to products and services in value creation. Data products and services would not exist without data. Furthermore, the products and services built using data can deliver greater value than data alone (Chatain, 2011; Lievesley et al., 1993). Products and services encapsulating data also allow for value creation in domains where data alone would have little or no value (Hearn and Pace, 2006). There are two main categories of data products and services—depending on whether the consumer is an individual user or an organization (or government). Products and services oriented towards organizations and governments are generally monetized with financial quid pro quo. That is, organizations and governments purchase the products and services. Although this monetization model is also used for data products and service oriented towards individuals, a significantly greater trend in monetization utilizes cross subsidization and freemium practices (Seufert, 2014).

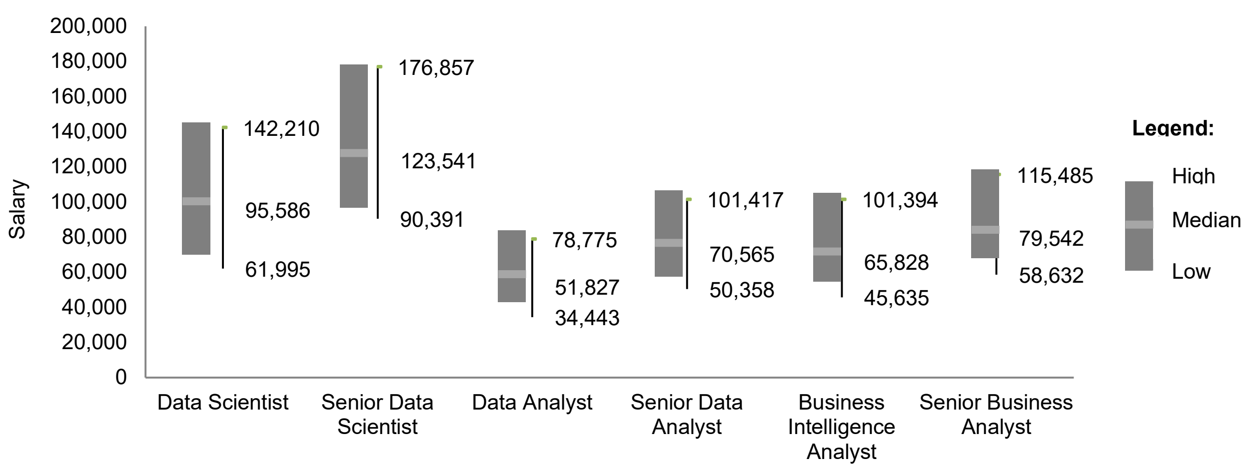
LABOR AND EDUCATION DIMENSIONS

Growing adoption of strategic data utilization by organizations has been creating an increasing demand for qualified labor. Exponential rise of data and rapid emergence of big data issues speeded up the demand for skilled data professionals (Bakhshi et al., 2014). Educational systems have not been prepared for such a situation and have not been able to produce sufficient supply of qualified data professionals in a desirable time. Data talent is difficult to produce—taking years of training. Hence, the strong demand for skilled labor has not been matched by an adequate supply.

Increasing gap between the demand and supply sides of skilled labor has brought forward several issues: labor shortage, labor migration, rise of wages, and boom of education programs for training new generation of data professionals. Presently, the main issue is a shortage of skilled data professionals in advanced economies. The shortage is expected to last during a five-to-ten year period. For instance, in the United States—based on the trends projected from 2008, it has been estimated that the demand for data professionals could exceed the supply being produced on current trends by 140,000 to 190,000 positions. Moreover, the demand for data talent in the United States could be 50 to 60 percent greater than its projected supply by 2018 (Manyika et al., 2011).

Rise of wages for skilled data professionals has been a natural consequence of contemporary labor shortage (King and Magoulas, 2015). Qualified data scientists have been in high demand by organizations aiming to realize value of data. Senior data scientists have been even scarcer. In the United States, it has been projected that there is a need for 1.5 million additional managers who can realize the value of data, ask the right questions and effectively use the results of data analysis (Manyika et al., 2011). A data scientist and a senior data scientist positions have a greater remuneration than other data related positions. Median salaries across the United Sates for a data scientist and a senior data scientist positions are 95,586 and 123,541 dollars, respectively. This is in contrast with median salaries for other data related professionals—such as analysts—whose wages range from 51,827 to 79,542 dollars (see Figure 3).

Figure 3. Annual Salary Ranges of Data Oriented Professionals in the United States



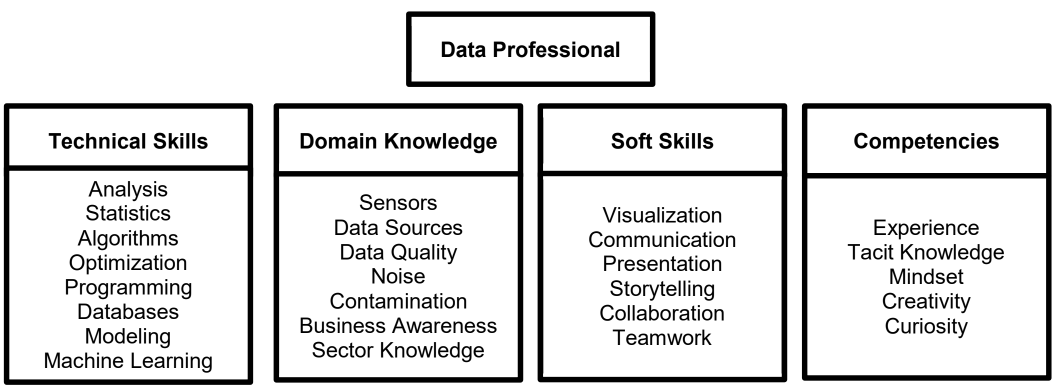
Averaged annual salary ranges for data oriented professionals across the United States. The highest salaries are for a senior data scientist and a data scientist positions. The median incomes for data science positions are higher than those of other data related positions. [Source: PayScale Inc., Updated: December 11, 2014].

Attractive remunerations and shortages of data professionals notably contributed to labor migration. Organizations aiming to substantiate strategic advantages of data and insights derived from it have been in need to fill the positions quickly (Bakhshi et al., 2014). Trained data professionals, in addition to their shortage, have been difficult to find. Hence, the positions have been filled with candidates providing a reasonable fit. Since data oriented work requires analytical and quantitative skills, trained engineers, mathematicians and physicists have been a suitable fit. This has led to migration of mathematicians, physicists and engineers to data related positions. Despite the migration, the shortage of suitable professionals for data oriented positions has remained (Manyika et al., 2011).

Data professional positions require a demanding skillset. Skillsets vary among economic sectors, organizations and particular positions (Bakhshi et al., 2014; Bruch, 2014; Harris et al., 2013; Davenport and Patil, 2012). Generally, the skillset can be categorized into four distinguishable groups: technical skills, domain knowledge, soft skills, and competencies (Figure 4). Technical skills represent analytical, quantitative, programming and modeling skills. Domain knowledge underlines understanding of data sources, their qualitative aspects and limitations, and business environment. Soft skills are abilities to present and communicate analytical insights, and work in collaborative environments. Competencies encompass abilities to approach complex questions and tasks in an effective manner and explore potential solutions from various angles.

The desired skillset of a data professional is significantly multidisciplinary (as illustrated in Figure 4). The outlined four pertinent sets of skills are difficult to acquire and may take years of training even for individuals with background in computer sciences, engineering or quantitative disciplines. The pressing need for training of data professionals, and creating a new segment of labor force, has led to establishment of new educational programs and curricula at universities.

Figure 4. Skillset Profile of a Data Professional



Four desired sets of skills of a data professional are: technical skills, domain knowledge, soft skills, and competencies. Each group contains a representative list of skills among a greater and diverse list of skills. Depending on economic sector, organization and position the preferred skillsets may vary.

Data professionals need to be well educated. They need a number of advanced abilities that require education at a graduate level and/or even higher. Such education and training responsibilities are well suited for academic and research environments. Hence, universities have been establishing new higher degree programs in data science and data analytics. Rapid rise of educational activities, curricula and research into data science has been originating from high demands for data professionals and realization of importance of data in the economy.

Educational programs aimed at training new data professionals range from intensive one-year courses to comprehensive two-year higher degree programs. For instance, the University of California, Berkeley offers a comprehensive multidisciplinary two-year professional degree program the Master of Information and Data Science (Data Science at Berkeley, 2015). The curriculum is divided into three segments: foundation courses (e.g. research design, data exploration, data storage and retrieval, machine learning, visualization), advanced courses (e.g. legal and ethical considerations, scalability), and a synthetic capstone course. Stanford University and Harvard University provide data science training as a part of their Master's programs in statistics (Data Science at Stanford, 2015; Data Science at Harvard, 2015). The curricula are more analytically and technically oriented.

GOVERNMENT AND REGULATORY DIMENSIONS

Realizations of economic and social potentials of data require governments and policy makers to appropriately respond to barriers and enablers. Progressive policies and regulatory frameworks should maintain a suitable balance between the needs and rights of individuals and organizations in data economy (Joseph and Johnson, 2013; Garber, 2013; Kuner et al., 2012). There is also a pressing need for international coordination of legal and regulatory frameworks (Cate, 2011).

Important areas where policies should play vital roles are data protection, security, and privacy (Garber, 2013; Lanois, 2010). While data protection and security pertain to the needs of both organizations and individuals, privacy protection addresses the basic rights of individuals. Data is the primary creator of value in data economy. It needs to be protected and secured. Adequate protection and security of data is in the best interest of countries as well as organizations (Wei et al., 2015; Anthes, 2010). Apt legislations and policies should be enacted to enhance data protection and set appropriate security standards.

Privacy protection is a pertinent domain where governmental and regulatory interventions are necessary (Margulis, 2011; Acquisti and Grossklags, 2005). Self-regulations by commercial organizations have been a major failure. Moreover, they present threats to data economy (Radinsky, 2015). Majority of the developed economies have deployed legal measures to protect privacy of individuals. Unfortunately, the United States is noticeably lacking behind—having presently no federal legislation regulating online privacy and data protection. On the other hand, the European Union has the most progressive legislative frameworks for protections and regulations of online data and privacy.

In the European Union, the privacy is a basic human right—per se, it is extensively regulated and enforced. Article eight of the European Convention on Human Rights (ECHR) stipulates it. Protection of personal data and privacy is further regulated by the Data Protection Directives (EUDP). The legislations include several directives (95/46/EC: The protection of individuals with regard to the processing of personal data and on the free movement of such data; 97/66/EC: The processing of personal data and the protection of privacy in the telecommunications sector; 2002/58/EC: The processing of personal data and the protection of privacy in the electronic communications sector; 2006/24/EC: The retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks).

Data protection and privacy are regulated in an increasing number of countries. Multinationals are required to comply with legal requirements in the legislative regions of their business. The legal requirements may notably vary depending on the legislative regions and sensitivity of data. Organizations must implement necessary measures and management practices in accordance with the effective legislations. Spectrum and diversity of legislative issues present numerous challenges.

Internationally accepted regulatory frameworks need to be established and adopted (Cate, 2011). The Organization for Economic Cooperation and Development has attempted to approach this task by issuing the 'Guidelines on the Protection of Privacy and Transborder Flows of Personal Data' (OECD, 1980). Unfortunately, presently only few countries, apart from the European Union, have adopted legislations by adhering to the core set of principles provided in the OECD guidelines (Garber, 2013).

CONCLUSIONS

Data economy has been gradually growing over the past several years and has been projected to continue growing at an increasing pace. The growth has been fueled by the exponential expansion of data and its impact on organizations and societies. Emergent data economy has been evident in several domains. Notable effects have been along the dimensions of trade, labor, education and government.

Trade dimension of data economy reflects realization of value of data. Data has a value that can be directly or indirectly monetized. Direct monetization of data is realized on data markets by quid pro quo transactions. Value of data is largely determined by organizations participating in the transactions. Presently, there are no globally accepted valuation methodologies for data. Numbers of qualitative and quantitative factors are influencing the value of data.

Indirect monetization of data is realized via data products and services. Products and services that are built on data exemplify symbiotic value creation. The created value is greater than the value of data itself. Data products and services also facilitate access to markets that would otherwise be unreachable. Data related economic activities increasingly add value to the gross domestic product of developed economies. In the United States, the added value represents approximately eight percent of the information and communication technologies sector.

Labor and education dimensions of data economy underline the transformational power of data on human resources. Organizations have been striving to realize the economic potential of data. In order to do so, they need qualified professionals that are able to manage data lifecycle, extract actionable knowledge from data, and transform it into values for both organizations and consumers. Such professionals are hard to find. There has been a shortage of skilled data professionals in developed economies. The demand for data talent in the United States is estimated to be 50 to 60 percent greater than its projected supply by 2018. Shortage of data professionals has led to labor migration, rise of wages, and development of new educational programs for training a new generation of data professionals.

A skillset profile of a data professional incorporates several advanced skills that require years of training. Generally, the desired skillset can be divided into four groups: *analytical and technical skills* (e.g. data analysis, statistics, modeling, machine learning, etc.), *domain knowledge* (e.g. data sources, data limitations, business awareness, etc.), *soft skills* (e.g. communication, visualization, presentation, teamwork, etc.), and *competencies* (e.g. experience, creative mindset, curiosity, etc.). Academic and research environments are the best suited for training data professionals. Increasing numbers of universities have been offering advanced degree programs in data science, analytics, and business intelligence.

Government and regulatory dimensions of data economy underscore the necessity of balancing the needs of organizations and the right of individuals in realizing the economic and social potentials of data. Policy makers should enact progressive regulatory frameworks and legislations. Adequate data and privacy protections should stimulate the data economy and enhance citizens' trust. Moreover, there is also a pressing need for international coordination of policies.

The presented study reflects a contemporary state of developed economies—particularly, the United States. It provides a pertinent perspective on major drivers and effects of data economic activities. Many of them will arise in developing economies later. Developing economies have a significant opportunity to learn and adequately prepare for the issues they may face. Managing a beneficial transition to data economy is the target of future study.

REFERENCES

Anthes, G. (2010) “Security in the Cloud: Cloud Computing Offers Many Advantages, but Also Involves Security Risks,” *Communications of ACM*, vol. 53(11), p. 16-18.

Acquisti, A., Grossklags, J. (2005) “Privacy and Rationality in Individual Decision Making,” *IEEE Security and Privacy Magazine*, vol. 3(1), p. 26-33.

Bakhshi, H., Mateos–Garcia, J., Whitby, A. (2014) “Model Workers: How Leading Companies are Recruiting and Managing Their Data Talent,” *Nesta, Royal Statistical Society*, UK, July 2014.

Bruch, L. (2014) “The Must-Have Skills You Need to Become a Data Scientist,” *Smart Data Collective*, http://smartdatacollective.com/lburtch/282561/must-have-skills-you-need-become-data-scientist, November 26, 2014 (Accessed: April 5, 2015).

Bryson, J. R., Daniels, P. V., Warf, B. (2004) “*Service Worlds: People, Organizations, Technologies*,” Routledge, New York.

Buhl, H. U., Raglinger, M., Moser, F., Heidemann, J. (2013) “Big Data,” *Business & Information Systems Engineering*, vol. 5(2), p. 65-69.

Cate, F. H. (2011) “A Transatlantic Convergence on Privacy?,” IEEE Security and Privacy, vol. 9(1), p. 76-79.

Charlesworth, A. (2009) “*The Digital Revolution*,” Dorling Kindersley.

Chatain, O. (2011) “Value Creation, Competition, and Performance in Buyer-supplier Relationships,” *Strategic Management Journal*, vol. 32(1), p. 76-102.

Data Science at Berkeley (2015). http://datascience.berkeley.edu/ (Accessed: April 5, 2015).

Data Science at Harvard (2015). http://www.extension.harvard.edu/courses/data-science (Accessed: April 5, 2015).

Data Science at Stanford (2015). https://statistics.stanford.edu/academics/ms-statistics-data-science (Accessed: April 5, 2015).

Davenport, T. H., Harris, J. G. (2007) “*Competing on Analytics: The New Science of Winning*,” Harvard Business School Press, Boston.

Davenport, T. H., Harris, J. G., Morison, R. (2010) “*Analytics at Work: Smarter Decisions, Better Results*,” Harvard Business School Press, Boston.

Davenport, T. H., Patil, D. J. (2012) “Data Scientist: The Sexiest Job of the 21st Century,” *Harvard Business Review*, October 2012.

EMC (2014) “The EMC Digital Universe study,” www.emc.com/leadership/digital-universe/index.htm (Accessed: April 10, 2015)

European Convention on Human Rights (ECHR). Convention for the Protection of Human Rights and Fundamental Freedoms as Amended by Protocols No. 11 and No. 14, http://conventions.coe.int/treaty/en/Treaties/Html/005.htm (Accessed: April 5, 2015)

European Union Data Protection (EUDP). http://ec.europa.eu/justice/data-protection/index_en.htm (Accessed: April 5, 2015).

Fan, J., Han, F., Liu, H. (2014) "Challenges of Big Data Analysis," *National Science Review*, vol. 1(2), p. 293-314.

Frischbier, S., Petrov, I. (2010) "Aspects of Data-Intensive Cloud Computing," *LNCS 6462*, Springer-Verlag, p. 57-77.

Gantz, J., Reinsel, D. (2012) "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East," *IDC iView*, December 2012.

Garber, L. (2013) "Security, Privacy, Policy, and Dependability Roundup," IEEE Security & Privacy, vol. 11(3), pp. 6-7.

Géczy, P., Izumi, N., Akaho, S., Hasida, K. (2007) "Knowledge Worker Intranet Behaviour and Usability," *International Journal of Business Intelligence and Data Mining*, vol. 2(4), p. 447-470.

Géczy, P., Izumi, N., Akaho, S., Hasida, K. (2008) "Enterprise Web Services and Elements of Human Interactions," *Business Information Systems* (W. Abramowicz and D. Fensel, Eds.), Springer-Verlag, p. 263-272.

Géczy, P. (2014) "Big Data Characteristics," *The Macrotheme Review*, vol. 3(6), p. 94-104.

Harris, H. D., Murphy, S. P., Vaisman, M. (2013) "*Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work*," O'Reilly, Sebastopol.

Hearn, G., Pace, C. (2006) "Value-creating Ecologies: Understanding Next Generation Business Systems," *Foresight*, vol. 8(1), p. 55-65.

Hunter, P. (2013) "Journey to the Centre of Big Bata," *Engineering & Technology*, vol. 8(3), p. 56-59.

Hurwitz, J. Kaufman, M. (2015) "*Cognitive Computing and Big Data Analytics*," Wiley, Indianapolis.

Jacobs, A. (2009) "The Pathologies of Big Data," *Communications of ACM*, vol. 52(8), p. 36-44.

Joseph, R. C., Johnson, N. A. (2013) "Big Data and Transformational Government," *IT Professional*, vol. 15(6), p. 43-48.

King, J., Magoulas, R. (2015) "*2014 Data Science Salary Survey*," O'Reilly, Sebastopol.

Klein, D., Tran-Gia, P., Hartmann, M. (2013) "Big Data," *Informatik-Spektrum*, vol. 36(3), p. 319-323.

Kroenke, D. M., Auer, D. J. (2013) "*Database Processing: Fundamentals, Design, and Implementation (13th Ed.)*", Prentice Hall.

Kuner, C., Cate, F. H., Millard, C., Svantesson, D. J. B. (2012) “The Challenge of 'Big Data' for Data Protection,” *International Data Privacy Law*, vol. 2(2), p. 47-49.

Lanois, P. (2010) “Caught in the Clouds: The Web 2.0, Cloud Computing, and Privacy?” *Northwestern Journal of Technology and Intellectual Property*, vol. 9(2), p. 29-49.

Laursen, G. H. N., Thorlund, J. (2010) “*Business Analytics for Managers: Taking Business Intelligence Beyond Reporting*,” Wiley, Indianapolis.

Lazer, D., Kennedy, R., King, G., Vespignani, A. (2014) “The Parable of Google Flu: Traps in Big Data Analysis,” *Science*, vol. 343(6176), p. 1203-1205.

Lievesley, D., Ross, S., Higgs, E. (1993) “Increasing the Value of Data,” *BLRD Reports*, vol. 6112, p. 205-218.

Malik, P. (2013) “Governing Big Data: Principles and Practices,” *IBM Journal of Research and Development*, vol. 57(3/4), p. 1-13.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A. H. (2011) “Big Data: The Next Frontier for Innovation, Competition, and Productivity,” McKinsey Global Institute, May 2011.

Margulis, S. T. (2011) “Three Theories of Privacy: An Overview Privacy Online,” *In Privacy Online (2011)*, p. 9-17.

McAfee, A., Brynjolfsson, E. (2012) “Big Data: The Management Revolution,” *Harvard Business Review*, October 2012.

McCallum, Q. E., Gleason, K. (2013) “*Business Models for the Data Economy*,” O’Reilly, Sebastopol.

OECD (1980). OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. https://www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm (Accessed: April 5, 2015).

Provost, F., Fawcett, T. (2013) “Data Science and its Relationship to Big Data and Data-Driven Decision Making,” *Big Data*, vol. 1(1), p. 51-59.

Provost, F., Fawcett, T. (2013) “*Data Science for Business: What you Need to Know about Data Mining and Data-Analytic Thinking*,” O'Reilly, Sebastopol.

Radinsky, K. (2015) “Data Monopolists Like Google Are Threatening the Economy,” *Harvard Business Review*, March, 2015.

Redman, T. C. (2015) “Overcome Your Company’s Resistance to Data,” *Harvard Business Review*, March, 2015.

Richards, N. M., King, J. H., (2013) “Three Paradoxes of Big Data,” *Stanford Law Review*, vol. 66(1), p. 41-46.

Seufert, E. B. (2014) “*Freemium Economics: Leveraging Analytics and User Segmentation to Drive Revenue*,” Morgan Kaufmann.

St. Amant, K., Ulijn, J. M. (2009) "Examining the Information Economy: Exploring the Overlap between Professional Communication Activities and Information-Management Practices," *IEEE Transactions on Professional Communication*, vol. 52(3), p. 225-228.

Tallon, P. P. (2013) "Corporate Governance of Big Data: Perspectives on Value, Risk, and Cost," *Computer*, vol. 46(6), p. 32-38.

Vertesi, J., Dourish, P. (2011) "The Value of Data: Considering the Context of Production in Data Economies," *In Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, p. 533-542.

Walsh, R. O., Callaghan, R., Yoffou, S., Hughes, R. (2012) "Big Data Is a Solution - So Where's the Problem?" *Cutter IT Journal*, vol. 25(10), p. 6-12.

Wei, K., LaSalle, R., Cooper, T. (2015) "If Data Is Money, Why Don't Businesses Keep It Secure?," *Harvard Business Review*, February 2015.

Wigan, M. R., Clarke, R. (2013) "Big Data's Big Unintended Consequences," *Computer*, vol. 46(6), p. 46-53.

Zwitter, A. (2014) "Big Data Ethics," Big Data & Society, vol. 1(2), p. 1-6.

BIOGRAPHY

Dr. Peter Géczy holds a senior position at the National Institute of Advanced Industrial Science and Technology (AIST). He can be contacted at: AIST, 2-3-26 Aomi, Koto-ku, Tokyo 135-0064, Japan. E-mail: p.geczy@aist.go.jp.