

## Can Big Data Analysis Make Cities Smarter? Emerging Issues and Stakeholder Implications in Smart Mobility

**Asst. Prof. Dr. Serkan Gürsoy**

Beykoz University

**Asst. Prof. Dr. Murat Yücelen**

Yeditepe University

### Abstract

This study deals with the challenges and bottlenecks with respect to the concept of smart cities which has largely been constructed on knowledge utilization issues and challenges. Despite the abundant existent literature in this field, the effective transformation of data into knowledge which can become a source of competitive advantage is still an ongoing debate, especially due to contemporary developments in big data analysis methods, approaches and strategies. As an emerging problem, the derivation of significant meaning from big data is among popular academic research fields, as well as being a crucial industrial and policy making engagement regarding value creating mechanisms in smart cities. Therefore in this study, limitations and challenges in translating big data into valuable knowledge in academia and industries are considered within the concept of smart mobility. In an attempt to propose researchers, business firms and governmental entities a collaborative approach, a perception about emerging issues is presented for clarifying some future constructs intersecting in relevant research and applied fields.

**Keywords:** Big data, Smart cities, Urbanization, ICT, Smart mobility, Industry-Academia collaboration

### Introduction

The story of how recorded information evolved into big data begins earlier than it may be assumed. This historical evolution begins with a variety of endeavors to quantify the growth rate in records (e.g. the size of university libraries) and is currently ongoing by means of contemporary methods to utilize the latest inventory. Dictionaries<sup>1</sup> from the 1940s contain the term "information explosion"<sup>2</sup> in reference to the rapid rate of growth in the amount of information available. On the other hand, current dictionaries define big data as, extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially with relation to human behavior and interactions. While practitioners were more concerned about the range of information availability in the 1940s, in a half century's time their focus mutated into the array of information.

The basic notion here is a kind of metamorphosis from the availability of information, to its usefulness. In other words, we are now more preoccupied with the interpretation of (big) data as opposed to its acquisition. The sheer volume of the numbers provides an exemplary basis for further clarification of the issues at hand. While 100 gigabytes of data was generated each day in 1992, the same amount was created in each second by 2002. Within 10 years, around 2012, 2.5 billion gigabytes of data (equal to 10 million full capacity blue ray discs) was generated daily by people or internet of things (IoT<sup>3</sup>) and 90 percent of this data was unstructured (Dobre and Xhafa, 2014). It is expected that by 2020, 40 trillion gigabytes of data will be created every day (Gantz and Reinsel, 2012). Not surprisingly, the amount of growth in world trade volume which is 20 trillion dollars in a quarter century (\$4 trillions in 1990 to \$24 trillions in 2015), is not as rapid when compared

<sup>1</sup> Oxford English Dictionary, 1940s; earliest use found in Lawton (Oklahoma) Constitution. 1941 cited in Press, G. in Forbes.

<sup>2</sup> Abundance of information (meaningful or not) or the rapid increase in published and collected, available data.

<sup>3</sup> Inter networking devices that collect and exchange data.

with the rate of growth in data collection. According to IDC's observations and assumptions, 23% of data can be useful if tagged and analyzed and just 3% of data was tagged. Not surprisingly, only 0,5% of data was analyzed or interpreted.

The growth rate of interpreted data is not as high as the increase in unstructured data due to reasons such as the shortage of skills, models, applications, governance and technical knowledge. To briefly reiterate, in order to effectively resolve emerging business or industrial problems, big data - as a new source of contemporary global demands - needs to be interpreted or refined. Big data analytics can be supposed as a refinery of this digital source (big data). The key function of this refinery is creating value from big data by harnessing its benefits. In order to clarify the challenges in extracting this value, understanding and conceptualizing the characteristics of big data is crucial. By taking into account the variety of the challenges and opportunities concerning big data and its analytics, this paper aims at focusing on and exemplifying issues with respect to smart mobility in smart cities within the framework of conversion of data to information.

The first section of the study explores and presents the basis of big data and analytics. Then the challenges in conversion of knowledge from unstructured data to information are conceptualized within the frame of smart mobility. In this context, the interpretation of the most common and popular bottlenecks with a view on specific challenges is critical for the extraction of value from big data. Finally, implications and research directions in relevant fields are presented in the concluding section.

### Issues and Challenges in Big Data and its Analytics

The past decade witnessed a sensational growth of effort in attempts to conceptualize big data as a major ongoing challenge that arose from an unprecedented expansion in the volume, variety and velocity of stored information. These focal attributes of big data are also known as the 3Vs model actually introduced in the late 1990s by Doug Laney and published in 2001 as a research note. Briefly, the first V, **Volume**, indicates the size of the data set, and the second V, **Velocity**, implies the rate of data inflow of a non-homogenous structure. The final V, **Variety**, refers to the range of data types within multiple structured or unstructured formats (e.g. text, audio, video or even data noise collected by sensors). Beyond the basic 3Vs, practitioners' and scholars' initiatives in analyzing and handling big data propose one another V. As the fourth V, **Value** implies the significance of extracting the benefit from the accessible data sets (IDC, 2012). Efforts on the part of academia and the industry for defining big data expand its characteristics by identifying new V's. **Veracity** (Ohlhorst, 2012) implies the accuracy and trustfulness of the information within the existence of complexities, anonymities and inconsistencies. **Variability** is about the changes in the meaning of collected data. **Visualization** means transformability of data to be readable (Sivarajah, et al., 2016). Debates on characterizing big data are still incomplete. Even though the character of big data expanded from 3Vs to 7Vs, the literature is still inconsistent and weak in generating complete definitions of each V and clarifying the relationships among them. Moreover, within the existing quantity of V attributes, practitioners from industry/business (Richey, et al., 2015) or practitioners from academia still remain confused about the big data creation and sourcing issues. At this point, practitioners face different forms of solid challenges. The most popular and common challenges about creation and sourcing - needed to be resolved urgently - are about volume, variety, velocity and veracity. As it is exemplified in the introduction section, data overcrowding increases redundancy and causes a serious challenge in the smart era (Chauhan, et al., 2016). Even though cloud-based structures provide opportunities for cost-effective ways of overcoming the challenges regarding these four Vs, the constant interaction between cloud servers still comprises complexities such as structuring desirable data (Khan, et al., 2015), automatic recognition of desirable patterns (Cheng, et al., 2015), and scalability and modularity of data for context aware platforms (Li, et al., 2015).

The insufficient comprehension of these characteristics causes a lack in the utilization of big data by practitioners. Beyond the shortcomings in comprehending big data characteristics, similar ambiguity can be found in the process of analytics. This ambiguity or insufficiency about analytics is related to the techniques of capturing, storing, restructuring and integrating big data to reap its concrete benefits. The literature on big data analytics prompts such majority of engagements as the collection, storage, integration, manipulation and finally, presentation of data. The process of analysis mostly covers the range of data from unstructured form to structured form for deriving meaningful presentations. By structuring or reconfiguring raw data, the community of big data analytics practitioners (or machine learning applications) tries to figure out valuable information by using a collection of different tools including integrative analytics, data mining, statistics, artificial intelligence, natural language processing, and so forth. The most common methods about analytics can be grouped into five types as **descriptive**, **inquisitive**, **predictive**, **prescriptive** and **pre-emptive analytics**. According to a systematic literature review study provided by Sivarajah, et al., (2016), even though **descriptive** analytics supplies some useful patterns for backward looking and current situations in order to forecast the next move (**predictive**), there is a mention (in

the literature) of *required ability* to read the facts and relate these facts to relevant decision making processes. The challenge concerning this ability stems from both human resource and machine learning issues. Some authors (Hasan, et al., 2014) propose some self-algorithms (e.g. self-organizing maps) for interpreting multi-structure inputs from a variety of sources. Finally, employing **prescriptive** analysis in order to clarify reason and result effects for business optimization would work fine when all the required dimensions were included. However, most of the databases are constrained on the number of dimensions (Banerjee, et al., 2013) and provide only partial insight into complex business problems.

### Smart Mobility for Smart Cities: Basic Issues and Challenges

As commonly proposed, big data contains an enormous potential to make urban life more convenient despite the considerable issues challenges outlined in the previous section. The population density in urban life is going to rise especially due to the movement of people from pre-urban areas (Bohli, et al., 2015). Today, 30 percent of the world's economy is accounted by just 100 cities while in a decade 600 cities are projected to constitute 58 percent of the global Gross Domestic Product and accommodate 25 percent of the world's total population (McKinsey Global Institute, 2016). Almost in a quarter century, 70 percent of the world population will be moved to cities if this urbanization trend continues (Lierow, 2014). This percentage will be 86 percent for more developed regions (UN DESA, 2012). The rapid growth in city population generates some widely acknowledged problems such as congestion, pollution, social inequality, and so forth. As expected, there is a growing need for infrastructure to handle people's needs in areas such as the flow of goods, information, security, health, education, and entertainment. Within this frame, the increasing density and complexity of urban life and supply systems prompt various challenges with respect to sustainability and efficiency. In order to assure future viability and prosperity in urban areas, new technology based solutions as well as new perceptions are becoming the most crucial requirements.

Information and Communication Technologies (ICT) is an appropriate tool for facing these challenges in transforming traditional cities into smart cities. Each advance in ICT has a great impact on the urban life by reducing costs, facilitating new products or processes, substituting goods or travel with digits and signals (e-books, e-audio files, e-tickets, telecommunication, messaging, navigating, etc.). Within this context in the literature, a smart city is defined as a sustainable city which benefits from various ICT services, big data and domains of application. It can be briefly described as a kind of vision for urban development in definitive assets by benefiting from ICT where the data is mostly created by people or IoT. These definitive assets can vary from city infrastructures to city rules and regulations which include the water supply system of the city, transportation systems, pharmacy sharing or power plants among many others.

Although the origins of the definition dates back to the 1990s in the form of "networked cities", it actually addresses places where ICT is combined with everyday objects to shape a more prosperous and inclusive future for the citizens. Some popular definitions of a smart city from the literature are presented in Table 1. As it can be seen, defining a smart city is not easy because of the absence of a general consensus on the meaning. However, there is a wider agreement about smart city characteristics. There are six important axes or characteristics of smart cities: smart economy, smart people, smart governance, smart environment, smart living and smart mobility (Kumar, 2015; Chatterjee and Kar, 2015, cited in Chauhan, et al., 2016). With few exceptions, the applications domain is significantly parallel to these characteristics. Briefly, **smart economy** mostly refers to domains such as productivity, markets or innovation, while **smart people** refers to flexibility in social and professional areas. **Smart governance** covers aspects of participation in decision-making, public and social services, political strategies and suggests transparent governance. Environmental protection, resource management and other natural conditions for a viable city are the aspects of **smart environment**. Living healthy, improving the quality of life, culture, safety and tourism, etc. are encompassed by **smart living**. Finally, **smart mobility** refers to local and global accessibility, and a smart system of transportation.

Enabling accessibility in urban places requires having various data and knowledge integrations and operational innovations such as on demand travel opportunities, cost and time effective access opportunities, safe and secure movement opportunities and the like. Ironically, in this context ICT plays a complementary role in the demand for travel instead of substituting it (i.e. telecommunications), because it triggers the steady rise of demand for mobility of goods and people and exerts pressure on the urban mobility system. This interplay between ICT and mobility unveils a critical challenge concerning long-term sustainability. It is a big challenge when it is perceived as the convergence of digital and physical facts. As a critical and crucial characteristic, smart mobility has an unavoidable impact on all potential stakeholders expending the benefits of living in a city (Arena, et al., 2013) in various ways; environmental, social and economic. A smart city needs smart mobility for reducing air and noise pollution, traffic congestion, transportation costs, and increasing safety

and speed. As a multifaceted point, smart mobility refers to integrated systems providing efficient and flexible travel opportunities in all modes of transportation with the help of smart infrastructure, intelligent systems and operational modeling.

**Table 1: Definitions of a Smart City**

Definition	Source
<b>Smart city as a high-tech intensive and advanced city that connects people, information and city elements using new technologies to create a sustainable, greener city, competitive and innovative commerce, and an increased life quality.</b>	Bakıcı, et al. (2012)
<b>Being a smart city means using all available technology and resources in an intelligent and coordinated manner to develop urban centers that are at once integrated, habitable, and sustainable.</b>	Barrionuevo, et al. (2012)
<b>Smart community – a community which makes a conscious decision to aggressively deploy technology as a catalyst to solving its social and business needs. The real opportunity is in rebuilding and renewing a sense of place, and in the process a sense of civic pride. Smart communities are not, at their core, exercises in the deployment and use of technology, but in the promotion of economic development, job growth, and an increased quality of life. In other words, technological propagation of smart communities isn't an end in itself, but only a means to reinventing cities for a new economy and society with clear and compelling community benefit.</b>	Eger (2009)
<b>A city connecting the physical infrastructure, the IT infrastructure, the social infrastructure, and the business infrastructure to leverage the collective intelligence of the city.</b>	Harrison, et al. (2010)
<b>Smart cities of the future have a high quality of life supported by sustainable urban development policies where all residents, including the poor, can live well and the attraction of the towns and cities is preserved. They are cities that pursue sustainable economic development through investments in human and social capital, and traditional and modern communications infrastructure; and manage natural resources through participatory policies. Smart cities are sustainable, converging economic, social, and environmental goals.</b>	Thuzar (2011)

Source: Adapted from Albino, et al. (2015)

Big data and big data analytics are crucial in feeding the systems that enable smart mobility to overcome the threat of rising transportation needs. However, some of the issues in big data for smart mobility emerge in parallel with transportation problems. According to a White Paper by U.S. Department of Transportation (2014), one of the challenges is the comprehensive understanding of big data pertaining to transportation that needs to be promoted at the inception of a process, despite the existence of technical and occupational obstacles in data capture, data management and analysis. Secondly, there is a lack of confidence about the specific form of utilization and benefits, such as transportation system monitoring and management for better predictions of impending conditions, imminent traffic flow breakdowns, weather and travel behavior and so forth. Moreover, traveler centered transportation strategies are required for understanding personalized travel demand and developing the correct tactical resolution. This might become available if the variety of traveler data describing traveler profiles and priorities are integrated. In addition, all the challenges with respect to traveler centered solutions are based on real time data flow and communication. Third, existing big data research is not sufficient when compared with the eminence of demand and data. As it is mentioned before, the lack of skills in big data analytics herewith emerges again as an insufficient research activity. Fourth, understanding and capitalizing on big data requires broader, non-traditional and non-transportation stakeholder engagement. Utilizing big data for smart mobility purposefully

may be expected from someone from outside of the transportation community, such as someone from the fields of statistics, machine learning, demand management, customer profiling or marketing. Finally, cost or resource implications are uncertain. More likely, developing applicable big data solutions mostly requires some investment in in-house training and/or outside consulting services, such as those associated with mining data to identify patterns and profiles, developing predictive algorithms, and incorporating the algorithms into existing expert systems. The task of facing these challenges has begun to be frequently outsourced to third parties who serve value extracted from big data. The dark side of this activity lies in the dependency for obtaining value from the outside source.

### Conclusion, Implications and Future Directions

While in some aspects the world is expanding at an exponential scale (data creation, world trade, urbanization, industrialization and pollution, travel and tourism, world population, ecological issues and sustainable practices), the world is also becoming more and more complex through advances in ICT which implies interconnectivity. With the volume of data growing incessantly, the variety and intensity of stakeholder needs and demands are also continuously increasing. In this paper, issues pertinent to the utilization of big data in improving the future of smart cities were presented with a view on the current and popular trends in academy and industry. It can be concluded that the development and implementation of effective smart city applications based on big data analysis is not an option but is a mandatory and urgent ordeal for which industry-academia collaboration is crucial. Additionally, authorities must have implemented solutions to trigger merge of interdisciplinary knowledge and skills both from academia and industry. Up to date awareness about future technologies and prospect seems to be critical for academia while up to date awareness about big data methodologies, techniques and applied knowledge is required by industry. On the other hand, shared understanding, reciprocity and skill acquisition between academy practitioners and industry practitioners can be the most critical issue within a dynamic and early stage research field.

Smart mobility was deemed to be an essential aspect of smart cities where big data analysis, if executed successfully, can provide groundbreaking outcomes with respect to cost and time management and the optimization of supply chains with a prospect on the improvement of the quality of life in urbanized locations. Based on the overwhelming speed by which data is created by multiple sources, interdisciplinary collaboration in big data analysis is essential, in theoretical and applied contexts alike.

On the other hand, in parallel with the advances in ICT and increasing complexities posed by urban lifestyles, new challenges emerge while some of the existing issues remain unresolved. The main issues that require attention were identified in areas such as the need for new skill acquisition, motivation for indulging in big data analysis, confidence in the achievement of beneficial results, and the need to update and improve physical and technical infrastructures. Other issues in the context of big data and smart cities which deserve to be investigated in more detail in the future include the development of efficient methodologies for industry-academia collaboration in the field of big data analysis, and the improvement of innovation literacy, which entails central actions in the form of resource allocation, education and business intelligence.

### References

- [1] Albino, V., Berardi, U., & Dangelico, R.M. (2015). Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, 22(1), 3-21.
- [2] DOI: 10.1080/10630732.2014.942092
- [3] Arena, M., Cheli, F., Zaninelli, D., Capasso, A., Lamedica, R., & Piccolo, A. (2013, October). Smart mobility for sustainability. In *AEIT Annual Conference 2013* (pp. 1-6).
- [4] Bakici, T., Almirall, E., & Wareham, J. (2012). A smart city initiative: The case of Barcelona. *Journal of the Knowledge Economy*, 2(1), 135-148.
- [5] Banerjee, A., Bandyopadhyay, T., & Acharya, P. (2013). Data analytics: hyped up aspirations or true potential. *Vikalpa: The Journal for Decision Makers*, 38(4), 1-11.
- [7] Barrionuevo, J.M., Berrone, P., & Ricart, J.E. (2012). Smart cities, sustainable progress. *IESE Insight* 14, 50-57.
- [8] Bohli, J.M., Skarmeta, A., Victoria Moreno, M., Garcia, D., & Langendorfer, P. (2015).
- [9] SMARTIE project: Secure IoT data management for smart cities. *Proceedings of the International Conference on Recent Advances in Internet of Things (RIoT)*, Singapore,
- [10]

- [11] (pp. 1-6).
- [12] Chatterjee, S., & Kar, A.K. (2015, August). Smart Cities in developing economies: a literature review and policy insights. In *Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Kochi, (pp. 2335-2340).
- [13] Chauhan, S., Chauhan, S., Agarwal, N., Agarwal, N., Kar, A. K., & Kar, A. K. (2016). Addressing big data challenges in smart cities: a systematic literature review. *info*, 18(4), 73-90.
- [14] Cheng, B., Longo, S., Cirillo, F., Bauer, M., & Kovacs, E. (2015). Building a big data platform for smart cities: experience and lessons from Santander. In *Proceedings of the International Congress on Big Data (BigData Congress)*, New York, NY, (pp. 592-599).
- [15] Dobre, C., & Xhafa, F. (2014). Intelligent services for big data science. *Future Generation Computer Systems*, 37, 267-281.
- [16] Eger, J.M. (2009). Smart growth, smart cities, and the crisis at the pump: A worldwide phenomenon. *I-Ways*, 32(1), 47-53.
- [17] Gantz, J., & Reinsel, D. (2012, December). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the Far East. IDC – EMC Corporation.
- [18] Available online: <http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf> (Accessed January 7, 2016).
- [19] Harrison, C., Eckman, B., Hamilton, R., Hartswick, P., Kalagnanam, J., Paraszczak, J., & Williams, P. (2010). Foundations for Smarter Cities. *IBM Journal of Research and Development*, 54(4), 1-16.
- [20] Hasan, S., Shamsuddin, S. M., & Lopes, N. (2014). Machine learning big data framework and analytics for big data problems. *International Journal of Advances in Soft Computing and its Applications*, 6(2), 1-14.
- [21] Khan, Z., Anjum, A., Soomro, K., & Tahir, M.A. (2015). Towards cloud based big data analytics for smart future cities. *Journal of Cloud Computing*, 4(1), 1-11.
- [22] Kumar, T. V. (2015). E-governance for Smart Cities. In *E-Governance for Smart Cities* (pp. 1-43). Singapore: Springer.
- [23] Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6, 70.
- [24] Available online: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (Accessed April 12, 2017).
- [25] Li, Y., Dai, W., Ming, Z., & Qiu, M. (2015). Privacy protection for preventing data over-collection in smart city. *IEEE Transactions on Computers*, 65(5), 1339-1350.
- [26] Lierow, M. (2014). B2City: the next wave of urban logistics.
- [27] Available online: [www.supplychain247.com/paper/b2city\\_the\\_next\\_wave\\_of\\_urban\\_logistics](http://www.supplychain247.com/paper/b2city_the_next_wave_of_urban_logistics) (Accessed April 20, 2017).
- [28] Ohlhorst, F. (2012). *Big data analytics: Turning big data into big money*. New York: John Wiley & Sons.
- [29] Richey, R.G., Morgan, T.R., Lindsey, K., Adams, F.G., & Autry, C.W. (2015). Global managerial perceptions of big data strategy in supply chain management. Council of Supply Chain Management Professionals (CSCMP) white paper, Research Strategies Committee, Chicago, IL.
- [30] Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263-286.
- [31] Thuzar, M. (2011). Urbanization in SouthEast Asia: Developing Smart Cities for the future?. *Regional Outlook*, 96-100.
- [32] United Nations Department of Economic and Social Affairs, Population Division (2012). *World urbanization prospects, the 2011 revision: Highlights*. New York: United Nations.
- [33] Available online: <http://esa.un.org/unup/Documentation/highlights.htm> (Accessed April 12, 2017).
- [34] U.S. Department of Transportation White Paper (2014). *Big Data's implications for transportation operations: An exploration*.