An Exploratory Urban Analysis via Big Data Approach

Sudhir Allam, *Sr. Data Scientist, Department of Information Technology, USA*

Abstract— *This paper describes how big data is influencing urban development. The proliferation of big data technologies for collecting and analyzing large amounts of data presents unprecedented possibilities for effective urban planning. Therefore, this paper addresses big data in urban science and its urban development prospects. It highlights the current state of big data implementation via examples and demonstrates how urban cities, including those experiencing rapid development, may use big data to effectively address energy and transportation challenges. For many years, the problems of global climate change and development have been on the priorities of academic organizations, government agencies, and civil society organizations. Although the ramifications of environmental issues' transnational and global dimensions remain central to policymaking, administration, and governance, the growing focus is being paid to the necessity and reach of local intervention [1]. The subject has been described by local governments in terms of the so-called "triple bottom line," which defines sustainable growth as development that is socially, economically, and environmentally sustainable. The integration of social media, census data, sensors, and conventional data provides a fresh perspective for addressing contemporary urban issues holistically and inclusively. Nonetheless, the BOLD techniques are modern and have not been adapted to urban infrastructure [1]. The aim of this paper is to determine if BOLD approaches will assist in reconceptualizing urban infrastructure not just in terms of technological and organizational characteristics, but also in terms of social values.*

Keywords: *Big Data, Urban science, urban development,*

I. INTRODUCTION

Most studies in metropolitan environments have historically focused on rough statistical numbers and smaller samples [2]. Even so, during the last decade, record digitization, sensor network extension, and society's computerization have resulted in a wealth of city data with large temporal frequencies and low spatial and temporal clustering. The breakthrough of "big data" would radically alter urban research. Big data transforms a cross-section of space into living data, providing a broader and more detailed view of urban existence than was previously possible [2]. Additionally, when used in conjunction with predictive analytics, big data can allow extrapolation of outcome variables to previously unmeasured segments of the population. Nonetheless, traditional causal inference challenges persist—big data only resolves attribution concerns on its own. When paired with exogenous sources of heterogeneity, big data becomes powerful for addressing classic social science issues, such as determining the effects and dynamics of economic development and valuing urban services and policies. The use of big data is enhancing city administration. By being increasingly data-driven in their activities, cities will fine-tune policies, optimize resource distribution, and predict potential needs. Consequentially, with certain urban data interventions, the ability to forecast effects or features is useful in and of itself—many functional challenges in cities do not need causal inference explicitly [3]. Additionally, since many data-driven initiatives are scalable, expanding data collection and digitization activities throughout cities fosters entrepreneurship and innovation. This paper outlines how big data approaches are important in urban analysis and illustrates how the analysis and interpretation of urban settings can be improved.

II. RESEARCH PROBLEM

The main problem that this explorative study aims at solving is to understand how the big data approach is significant in urban analysis. The significant effect of the global economic crisis has been the widening of inequality across the globe. Infrastructure is critical in this discussion because it brings broader health, financial, and environmental impacts to the community that

goes beyond traditional cash returns estimates [4]. Cities are also faced with obstacles as well as resources to utilize a variety of static and complex databases as a result of the future exploration of big data. Techniques based on big data are opening up new avenues for developing decision-making processes for community development and management. City attributes and output factors can now be measured at higher frequencies and on finer spatial scales than ever before, thanks to new big data sources. However, big data alone would not be sufficient to answer major urban science issues [4]. The idea of an intelligent city is highly preferred because it improves the standard of life of local people, with various disciplines like smart community, intelligent transit, intelligent health services, smart parking, etc. The continued expansion of dynamic metropolitan networks is greatly harmed by real-time data collection and intelligent decision-making capability [5]. Big data is of major importance for the analysis of cities whether estimation of the visibility is required or if exogenous shocks may be associated with persons or places.

III. LITERATURE REVIEW

A. What is big data?

Big data is a relatively recent concept used to characterize highly open and multi-source data, but the fact that it's so large will render interpretation very challenging. Big data is described by others as anything that is collected or tracked digitally by knowledge and communication systems such as networked sensors, "smart" computers, the internet, and social networking [5,6]. Big data is often a flexible concept, indicating various things for different markets, based on technological context and skills. Perhaps the most prevalent assumption is that it refers to vast and complicated data sets that necessitate the adaptation of existing approaches or the creation of completely new ones. The concept "large data" is often used in conjunction with forecasting computational techniques that forecast potential scenarios. Standard data analysis may be used to explain well-established urban associations (for example, a high population correlates with increased carbon emissions). However, as modern techniques of research are used, big data will uncover both established patterns and novel insights [6]. Sometimes presented in terms of length, range, and pace, the difficulties inherent in the usage of big data at the city level can be classified into three broad categories. To begin, current information management platforms and conventional data processing methods are not optimized for cleaning, measuring, mining, managing, or maintaining large data sets. Second, there is a dearth of resources at the municipal level for analyzing community-wide results. Third, city councils and their development agencies can lack the tools necessary to translate conclusions into actionable recommendations. And of the most pervasive myths about big data is the implication that the data collection must be enormous. However, certain data sets can only span a large area, as in the case of traffic data collected from users located many miles away. The amount of inputs does not have to be enormous; usually, inputs are gathered remotely from wireless networks, mobile devices, sensors, applications, video and camera streams, and embedded radiofrequency identification readers. Big data are massive, include a variety of data forms, both organized and unstructured, and grow rapidly, i.e. data is stored and analyzed automatically, usually in real-time. With this structured concept in mind, I define big data as data that is too massive to be managed by standard software programs such as Microsoft Excel, Access, or even MySQL, which is skilled in managing large data sets [7]. It's similar to pornography. In other terms, anytime you see big data, you immediately recognize it as such. Simply stated, if you are unable to manage the scale of the data in the current program, it is most likely big data. The strength of big data, on the other hand, stems from the idea that concrete trends can be extracted from billions of documents without aggregation. Previously, achieving this level of granularity in data processing was very difficult due to the time and effort required to process such a large body of data. Recent technical advancements, such as Hadoop and MapReduce, have made it significantly simpler and quicker to process and interpret large amounts of data.

B. Urban analysis and big data approach

Urban science is partly concerned with the use of disparate data sets to better explain and solve architecture and functionality issues in densely populated cities. Until recently, urban science was mostly theoretical and concentrated on classifications centered on urban constraints: purpose, scale, and patterning of device networks [8]. This methodology is evolving in lockstep with the rapid growth in evidence on how cities work. New data sources have significantly aided the effort of urban science to characterize the components of urbanization, recognize urbanization as a normal phenomenon governed by legislation, and link urban areas to broader global innovations [8]. Social networking and

crowdsourcing are constantly generating large amounts of data, which can be used in a variety of spatial contexts to further explain urban patterns such as violence, blight, and transit flows. The aggregation and review of big data will help guide both immediate urban decisions (such as where to place a stoplight) and long-term development processes for enduring urban problems (such as where to site water storage to support 20-year growth projections). When used properly, it can be an effective way to easily identify opportunities to save time, resources, and money [9]. Although uses for these databases are still being explored, researchers are gradually turning to big data to solve critical urban problems. The capacity to define and comprehend urban systems, analyze data requirements, and address urban challenges analytically in situations of many constraints is a skill set that is underrepresented in many cities. Using big data to gain a greater understanding of urban functions is a growing necessity in local government, but it frequently remains elusive outside of a laboratory environment.

However, skills may be established at the local government level by establishing data collection needs and sources. Data collection methods and interpretation may expose trends and interactions in urban systems. It is the foundation for our interpretation of existing partnerships, such as higher commuter traffic and higher asthma rates [9]. Data helps planners to critically analyze urban traffic trends to consider which types of transport are selected for what population. It helps energy management and utilities to schedule changes in energy demand through daytime, environmental forecasts, or customer behavior. Many claim the technologies and techniques of big data would prove similar in social sciences to the invention of the microscope in natural sciences. Big data, though, represents far more than findings [10]. It's not just analyzing the cities; it's studying the structures inside the city and how they can be changed for progress. Urban Science is focused on 4 predictable approaches to scale towns. More recently, science looks past scalar processes, exploring what causes scaling trends. Although using and big data can seem daunting, especially for "emerging" cities, it may also generate distinct opportunities. Emerging cities are tiny yet fast-growing metropolitan centers, sometimes part of a megalopolis, sometimes reinventing their futures. Although with already limited time and money, the challenges presented by big data processing to maximize and improve efficiencies through city departments and infrastructures are difficult for these cities to overlook. Urban areas are often early adopters of smart technologies, but the current infrastructure can restrict them [11,12]. Due to a lack of budget, emerging cities might have missed multiple cycles of technology upgrades, but they're ready to use the sorts of insights big data may deliver.

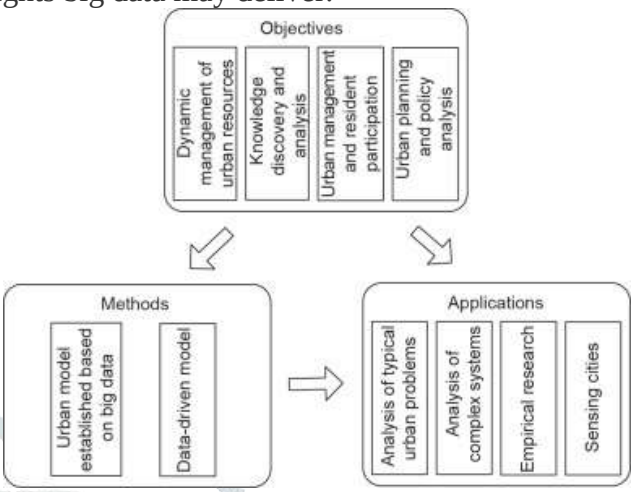

Fig i: objectives, methods, and applications of urban big data

C. Applications

i. Urban energy review

Major cities fail to consider energy use trends across markets, to reduce use in ways that benefit urban wellbeing and lead to low utility bills. Smart meters are one indicator of hard technology improvements that many local services are beginning to deploy, allowing them real-time access to data processing. Smart grid and Microgrid systems show how data is starting to be collected quickly and remotely to guide the preparation of urban electricity connections [12]. However, major hardware infrastructure projects that collect real-time data may stimulate large upfront capital spending, grid restructuring, and often neighborhood opposition to privacy loss perception. Software and frameworks are constantly being used to collect and analyze energy data for industrial and residential buildings. For example, energy usage data may be collected voluntarily by sector or by transparency mandates. Without realizing what the energy demand data reveals, it is impossible to make informed and strategic investments in both local and regional energy upgrades. Using group device technologies around residential energy use is increasing. This type of data can be used to detect energy demand patterns, which can then warn program implementation to encourage or mandate energy benchmarking [13,14]. Energy applications are gaining use and rapidly transforming how municipal government engages people around energy usage.

Using computer programs to evaluate energy usage results. Power-related results that

can be achieved using Big Data approaches and research include: 'Emissions control, reporting and verification; disaster response vulnerability detection; building and appliance energy quality standards; and using behavioral intelligence to promote energy efficiency'[13). Local government data issues include: 1) knowing what energy data is available; 2) how to access it; 3) what needs to be accomplished to make it usable; 4) what research methodology better addresses related questions, and 5) communicating data conclusions for desired outcomes [14]. While it is important to be able to obtain and interpret this type of information to make educated choices, obtaining this data will also enhance engagement in governance and preparation, as well as allow greater collaboration between agencies.

ii. Using mobility data to transform transportation systems

Energy usage is not restricted to electricity in homes but includes transport fuel consumption. Alternative transport approaches can not only boost the perceived viability of a metropolitan environment but can also favorably correlate with other metrics such as economic development and air quality. Using data to explain how regular transportation of a metropolitan community will contribute to effective preparation, tailored grant proposals, and adoption of long-term strategies. Commuter data may be collected by survey and transportation-type sensors (e.g. bicycle, truck, automobile, and rail) and evaluated to provide answers about the right transit categories to invest in. Integrated modeling is a strong data analysis technique to show how to support chosen methods of transport. For example, a model of commuting activity in a metropolitan region with fragmented jobs and pollution may investigate connections between scattered land use and long carbon-producing commutes [14]. Local government agencies also utilize data sets to consider traffic habits, use of bike and rideshare lanes, and walking behaviors. Private people and businesses may use aggregated crowd-sourced data from mobile devices to handle real-time traffic. Many of these apps merge dynamic smartphone data and mapping tools, providing channels that consumers modify and manage.

In developing cities, environmental factors associated with alternative transport (such as city walking and biking) are significant. Some research showed the increased walking correlated with industrial land usage. Studies have also shown that size, mixed land use, and network access are important physical environment variables influencing bicycle use. The availability of well-connected, low-traffic streets with cycling facilities, such as bicycle lanes and bicycle boulevards, is important for transit-oriented usage of bicycles. In addition to the impacts of the physical world, research finds that cycling use is favorably affected by a welcoming social environment [15]. Data on citizens' tendency to prefer alternate transport forms can be reversed toward socioeconomic or other situational considerations, contributing to a greater understanding of means of influencing citizens to choose pedestrian and cycling paths. Mobile knowledge exchange apps are the most common and best-used medium [15]. Accessibility of smart mobile devices has enabled contextualized coordination on various spatial scales, exposing urban designers to more direct contact with the city than ever before [15]. Because of their proximity, smaller cities have a larger and more normal capacity to engage more personally with their inhabitants and technological culture. Collecting, managing, and reviewing data will allow city councils to address a particular issue in a small jurisdiction and handle and respond faster as needs shift. Small and developing cities are ready for big-data conversation. They will achieve this by opening internal data sets, encouraging the technology world to leverage them to recognize urban patterns. They should support current applications such as the Neighborland framework (https://neighborland.com) to allow people to attend group talks and get input on community development processes. Emerging cities should also develop their applications for urban problems like Bloomington does for the latest "Rent Rocket" program [16]. This application is intended to chart average rental property energy use, so tenants can make educated choices on where to reside based on the energy needed to heat, cool, and commute. Many such volunteer data systems operate, which can be effectively transferred to small cities based on their long-term economic development targets.

D. Case studies

City of New York

In recent years, an open-data and open-government revolution has hit the globe like an outbreak. Most major cities worldwide already have open data policies (Figure 2). 46 Towns and regions in the U.S. have open data policies. NYC leads the crowd by experimenting with modern technologies and resources to allow easier access to their results. As of July 2015, NYC made 1,350 data sets publicly accessible with plans to publish all of their interactive public data by 2018[17].

Other cities include Los Angeles, Chicago, Boston, and Seattle join New York City. Thanks to this widespread open data program worldwide, you will now view all manner of government data at your fingertips. The form of data opened to the public varies from a straightforward federal job tally to live video streams from every street corner of New York City.

Palo Alto, California

The city of Palo Alto, CA, which had a population of 64,403 inhabitants in 2010, has a lengthy tradition of excessive vandalism on public and private property. To address this, the city created the "PaloAlto311" software, which enables anybody with a mobile or tablet to take a snapshot of the vandalism and submit it to city authorities. Service orders may also be made by phone or via the city website [16]. Following that, users will monitor the settlement of their case in real-time, from the time it is entered and allocated before it is resolved. After its debut in June 2014, the software has assisted in the removal of over 114 instances of vandalism. Aside from stopping graffiti, Palo Alto residents have used the software to alert authorities to approximately 80 instances of unlawful dumping and approximately 40 traffic street light malfunctions [16]. They have also recorded clogged drains, asking for money in parks, and faulty electric vehicle charging stations to the area. PaloAlto311 has aided in the resolution of conflicts, inspired people, and, ultimately, increased transparency.

Dubuque, Iowa

Dubuque, Iowa collaborated with IBM Research in 2009 to concentrate on involving people as allies in the city's effort to become smarter and more prosperous. As of May 2013, the city was using 23,000 smart meters to track service and transportation data, which is shared with residents to allow for informed decision-making. Users may view their water intake in 15-minute intervals and make informed usage changes [17]. This capability also assisted in water conservation, with consumers reporting a 6-7 percent decrease in demand – as well as an 8-fold rise in leak detection [17]. The city also employs smart gas and electric meters, which have resulted in a reported 7-11 percent reduction in electricity usage since 2011[18]. Dubuque also gave 1,000 volunteers a mobile app that captures and monitors activity trends in the region. This information will be used to modify the bus system's routes and seven pickup times to attract more passengers. In the future, such an application may be used to locate parking or the best routes across the capital. Innovative methods to data-driven problem solving, such as those found in Dubuque, may result in cost savings.

IV. THE FUTURE IN THE UNITED STATES

Big technology, in conjunction with the open data revolution, creates new resources for the public and non-profit industries in the United States. Big data analytics tools are becoming more widely accessible for free, and the obstacles to accessing, collecting, and analyzing big data are diminishing throughout the coming era. Previously, creating a personalized big data infrastructure necessitated assembling a squad of computational scientists, network developers, and computing managers. There is no need to contend with the difficult activities of setting up physical infrastructure, upgrading suitable applications, and optimizing the device for optimum efficiency thanks to database-as-a-service (DBaaS) or cloud database [18]. Both back-end management activities are handled by service suppliers, allowing customers to concentrate solely on data collection and problem-solving.

V. ECONOMIC BENEFITS TO THE UNITED STATES

Emerging cities have many options when they consider connections to and usage of big data to understand patterns and associations - and, finally, to address connectivity and energy problems. A long-term vision will be created by creating a roadmap that describes a holistic community data and technology strategy. Big data in the context of productivity results usually refers to administrative income data accessible at fine spatial frequencies (and sometimes disaggregated temporal frequencies as well). Administrative Internal Revenue Service (IRS) details, for example, contain income information at the address stage. Data from the Longitudinal Business Database offers address-level information on firm production and sales. Other outlets, such as credit card firms, may have disaggregated statistics on urban company transactions through time and distance. New York became great in the nineteenth century because its harbor and water-borne entry through the American continent offered enormous productivity benefits not just to merchants, but also to producers who profit from lower shipping costs[20].

Partnerships with the public and private sectors will pave the way for this, benefiting all sectors. Seattle, WA partnered with Microsoft on the Seattle 2030 District initiative to create energy-efficient smart buildings and reduce

power use. Via the Chicago, Open Data portal, the city of Chicago, Illinois, collaborated with Allstate to use data-driven technologies to enhance city services. When more cities implement these forms of policies and collaborations, other cities may gradually follow suit.

VI. CONCLUSION

This paper summarized how big data can be used to improve urban analysis through an integrated existence of urban networks. The use of data plays an important role in the collective growth of spatial design and therefore urban technology. In a data-rich society, this case study will help in community design and policy growth. People will create more visible links with data and gain more visibility into the pervasive role of technology and data systems in the community. This also aids in the diagnosis of patterns of interactions and outcomes among hard and soft infrastructure, such as resource usage (and overuse), policy execution, infrastructure and wealth distribution, dispute, and directional shift. Furthermore, this allows a transition from 'outside-in' thought to 'inside-out' thinking, allowing neighborhoods to create resources and deliverables as resourcefulness.

Local infrastructure and big data research would continue to be critical for developing new approaches to prosperous futures. Even though the shortcomings of big data have been generally acknowledged in the scholarly literature, we will suggest that big data provides significant possibilities for unpacking urban infrastructure. Big data is not a panacea for all urban ills, and further analysis is needed to assess the possibilities and problems that lie ahead. It is essential to consider the utility of big data and sustainable planning methods concerning the importance of urban resources. This case study has the potential to shape future policies in ways that would improve the lives and well-being of all city dwellers. Each community has an excess of data, and it is critical to consider how big data can offer a deeper view of cities and how they work. Big data debates are useful in the sense of urban infrastructure since the present state of conceptualizing urban infrastructure is mostly focused on economic principles rather than how it connects various industries or resources.

References

1. H. Zhang and H. Wang, "Distributed subdata selection for big data via sampling-based approach", *Computational Statistics & Data Analysis*, vol. 153, p. 107072, 2021.
2. B. Qin and S. Qi, "Digital transformation of urban governance in China: The emergence and evolution of smart cities", *Digital Law Journal*, vol. 2, no. 1, pp. 29-47, 2021.
3. N. Nelson, "Smart urbanism not smart cities: takeaways from Bike City Amsterdam", *Urban Notes*, 2021.
4. K. Kourtit, P. Elmlund and P. Nijkamp, "The urban data deluge: challenges for smart urban planning in the third data revolution", *International Journal of Urban Sciences*, vol. 24, no. 4, pp. 445-461, 2020.
5. H. Jiang, S. Geertman and P. Witte, "A Sociotechnical Framework for Smart Urban Governance", *International Journal of E-Planning Research*, vol. 9, no. 1, pp. 1-19, 2020.
6. A. Honarvar and A. Sami, "Towards Sustainable Smart City by Particulate Matter Prediction Using Urban Big Data, Excluding Expensive Air Pollution Infrastructures", *Big Data Research*, vol. 17, pp. 56-65, 2019.
7. H. Jiang, "Smart urban governance in the 'smart' era: Why is it urgently needed?", *Cities*, vol. 111, p. 103004, 2021.
8. M. Hu and M. Pavao-Zuckerman, "Literature Review of Net Zero and Resilience Research of the Urban Environment: A Citation Analysis Using Big Data", *Energies*, vol. 12, no. 8, p. 1539, 2019.
9. A. Gessa and P. Sancha, "Environmental Open Data in Urban Platforms: An Approach to the Big Data Life Cycle", *Journal of Urban Technology*, vol. 27, no. 1, pp. 27-45, 2019.
10. A. Pandey and R. Pandey, "Data Modeling and Performance Analysis Approach of Big Data", *SSRN Electronic Journal*, 2019.
11. H. Shafizadeh-Moghadam, "Improving spatial accuracy of urban growth simulation models using ensemble forecasting approaches", *Computers, Environment and Urban Systems*, vol. 76, pp. 91-100, 2019.
12. Y. Kang and E. Kim, "Differences of Restorative Effects While Viewing Urban Landscapes and Green Landscapes", *Sustainability*, vol. 11, no. 7, p. 2129, 2019.
13. L. Quitzow and F. Rohde, "Imagining the smart city through smart grids? Urban energy futures between technological experimentation and the imagined low-carbon city", *Urban Studies*, p. 004209802110059, 2021.
14. W. He, "Research on the Integration of Urban Traffic and Big Data", *Urban Transportation & Construction*, vol. 6, no. 2, p. 38, 2020.

15. S. Rehm, S. McLoughlin and G. Maccani, "Experimentation Platforms as Bridges to Urban Sustainability", *Smart Cities*, vol. 4, no. 2, pp. 569-587, 2021.
16. K. Clarke and J. Johnson, "Calibrating SLEUTH with big data: Projecting California's land use to 2100", *Computers, Environment and Urban Systems*, vol. 83, p. 101525, 2020.
17. S. Bibri, "The anatomy of the data-driven smart sustainable city: instrumentation, datafication, computerization and related applications", *Journal of Big Data*, vol. 6, no. 1, 2019.
18. N. Rodrigues, M. Vale and P. Costa, "Urban experimentation and smart cities: a Foucauldian and autonomist approach", *Territory, Politics, Governance*, pp. 1-19, 2020.
19. K. Kourtit, "City intelligence for enhancing urban performance value: a conceptual study on data decomposition in smart cities", *Asia-Pacific Journal of Regional Science*, vol. 5, no. 1, pp. 191-222, 2021.
20. K. Kourtit, P. Nijkamp and S. Suzuki, "Are global cities sustainability champions? A double delinking analysis of environmental performance of urban agglomerations", *Science of The Total Environment*, vol. 709, p. 134963, 2020.