DATA SCIENCE - EXTRACTING INSIGHTS
Open Journal

*Review*

# The Synergy of Data Science, GIS Spatial Analysis and Knowledge Management as a Path to Sustainability Insights

**Mohamed B.O. Osman**[1,2*]

[1]*Institute of Geomatics and Civil Engineering, University of Sopron, Sopron, Hungary*
[2]*GIS Department, University of Khartoum, Khartoum, Sudan*

[*]***Corresponding author***
**Mohamed B.O. Osman**
Institute of Geomatics and Civil Engineering, University of Sopron, Sopron, Hungary; GIS Department, University of Khartoum, Khartoum, Sudan;
E-mail: mohamedbabekir95@gmail.com

***Article information***
**Received:** February 15th, 2023; **Revised:** June 13th, 2023; **Accepted:** June 30th, 2023; **Published:** July 16th, 2023

***Cite this article***
Osman MBO. The synergy of data science, GIS spatial analysis and knowledge management as a path to sustainability insights. 2023; 1(1).
doi: https://doi.org/10.70705/ppp.dsei.2023.v01.i01.pp30-37

## ABSTRACT

This chapter focuses on data science, GIS-based spatial analysis, and knowledge management (KM) threesome and its potential contributions to sustainability insights. Discussion commenced with tracing the historical evolution and essen- tiality of geography for comprehending environmental and social dimensions of sustainability. Then, argumentation delved into the interplay among the three- some's domains and their contributions to sustainability achievement. A prolonged elaboration was provided on geospatial data analytics, visualization, geospatial data mining, and predictive models and their significance for extracting informative and meaningful insights. Since data science has transformed and enriched most scientific disciplines, its empowering implications on GIS were explained. Spatial analysis, therefore, occupied a central position and enabled advanced GIS technique utiliza- tion to reveal patterns, relationships, and trends in geospatial data. Furthermore, the chapter explained the interdependent relationships between GIS and KM. Integrating GIS and KM techniques has revolutionized geospatial data organization, visualization, and dissemination and enhanced applications of decision support, environmental planning, and others. The Nexus of this threesome, therefore, serves as a roadmap for solving issues of intricate spatial problems via modeling and informed decisions. The chapter stressed and concluded that the integrated fusion of data science, GIS, and KM provides a robust framework and ideal tools supporting sustainability.

**Keywords**
Informatics; Data science; GIS-spatial analysis; KM; Sustainability.

## INTRODUCTION

Space where humans exist is vital for their lives and activities. Thus, geography has always been of significance to humans. Since times immemorial, with concerns and ongoing endeavors, man has been striving to understand the phenomena and objects in his surrounding environment to enable utilization as well as prevention of hazards and disasters nature holds for the sake of a prosperous and safe life.

In this light, the stone-age hunters were anticipating the locations of the quarry, early explorers' deaths or lives were determined by their geographic knowledge, and contemporary societies would not be able to plan and perform their everyday activities without asking and finding geographic answers such as where and how far, [1].

Over these centuries of lengthy and multiple interactions, humans have known and experienced nature by benefiting from its resources and, at times, being trau- matized by the disasters it causes. This, besides developing approaches for spatial thinking, had accumulated tremendous knowledge about the surrounding objects and phenomena and composed the scientific heritage, especially in various fields of the so-called "life and applied sciences", such as geography.

Thanks to these experiences as well as the spatial thinking approach, humans have accumulated spatial knowledge that gradually varies over time and space, on the one hand, and from knowledge about objects or phenomena characteristics to their interrelationships, which all together comprise the surrounding environment, on the other hand.

For this reason, knowledge, especially geographic knowledge, starts

with and var- ies between simple and complicated forms. The simple form is the level of describing the object or phenomenon and identifying its characteristics, while the second form is detecting the interrelationships among these objects or phenomena and understand- ing how they affect each other's [2].

Historically, and prior to the technological and communication revolution, the sources of knowledge were basically confined to two sources: direct and personal observation or experience and the indirect experiments of others [2]. This caused knowledge accumulation and dissemination over time and space to become very decelerated and limited.

Aftermath, this indolent and gradual knowledge exchange had been abruptly disturbed by iconic events and great breakthroughs, such as the prosperity of the paper industry and the invention of the typing machine. These backboned and revolutionized various forms of printed press and hence dramatically changed the production and dissemination of information and knowledge, reaching to the recent situations of information and the technological revolution led and shaped by the internet.

This era is, therefore, the era of information and technological revolution,[3] which is forged and directed by the extensive and diversified range of human activities and the general advancements in all aspects of life. Awing to automation, economic growth, urbanization as well as the technological explosion, especially the advancements of communication means and computerized information technology.

These factors had drastically changed and modernized the availability and acces- sibility of information and knowledge through developing data collection, processing, and dissemination techniques. This dissemination is embodied in the obtainable and massive volumes of knowledge and databases, data warehouses, and big data archives. Abundance of these enormous data, in both levels of phenomenon and inter-phe- nomenal escorted, consequentially, urbanization and industrialization, besides the ever-growing complexities in different life aspects.

In the same context, the increasing population and their economic activities have urged the dire need to take appropriate decisions regarding development, resources utilization, society, and environmental conservation [4]. Likewise, it aroused the development of techniques to utilize and extract useful and practical-guiding knowl- edge out of these compilations of knowledge and data bases.

These techniques include, for instance, GIS spatial analysis, spatial decision support systems, and expert systems. All of these have the purposes of enabling us to simplify, model, and abstract the complex and thorny reality for better understand- ing as well as supporting data-driven and informed decisions for best and friendly practices for man and his surrounding environment alike.

## 2. A brief epistemological framework of data science and GIS-spatial analysis

As side-effects of the overwhelming digitalization, the massive advancements in computation capabilities and information and communication technologies affected approximately all disciplines and professions. It was thus argued that data science had shifted and redesigned all scientific disciplines, by means of adjusting their adopted approaches [5].

Therefore, new trends and developments in science and research emerged and caused many fields and disciplines to revisit and upgrade their methods and practices to become more IT-oriented [6]. Furthermore, the merging of geography into other fields was noticeable in many disciplines that engaged the spatial dimension in their approaches by being oriented to handle more spatially referenced information and data through the platform of geographic information systems, or GIS. This indicates growing concern, cognition and consideration of geographic and spatial perspectives.

Geography itself, even prior to the computation and information explosions, had reshaped its nature, through adapting newer approaches and transforming from the traditional descriptive geography (regional and systematic geography) to the so- called, applied geography. This added value and dimensions of practicality to geogra- phy as a problem-solving scientific discipline.

Applied geography, in the form of maps and spatial information, has served dis- covery, planning, cooperation, and various management-related practices for not less than 3000 years. For this reason, it was debated that most of the geographic knowl- edge is applicable to routine tasks [1]. Yet this seems to not be enough and satisfying to the requirements of the state-of-the-art in informatics and modern technological situations in contemporary society.

To further cope with the general advancements in all of life's aspects and the complexities of the rapidly changing world, geography has shifted its paradigms and approaches; thus, it has undergone different phases, which can generally be expressed in the aforementioned transformation from descriptive to applied geog- raphy. Of course, the quantitative revolution, digitalization, and engagement of computer graphics were the most important and influential milestones toward that transformation.

These all paved the road and led to the emergence of a newer field and branch of applied geography, the geo-informatics, that endeavors to take advantage of the advancements in information and communication technologies' infrastructure to leverage the various applications of geography. Because spatial information is paramount, we have developed tools and techniques called Geographic Information Systems (GIS) to handle geospatial information and support us with geographic knowledge. GIS enable us to collect and utilize spatial data (See Figure 1) [1].

GIS is a computer-based system that facilitates the development and use of spatial data. This is typically known as spatial analysis. There are many reasons for develop- ing, utilizing, and adapting spatial analysis methods and general GISciences, the most prominent factors are societal and technological. The societal reason is basically the need to more effectively and efficiently use the resources, alongside consideration of the environment, finite resources, and growing population. While technological reasons are the interest and necessity of applying the new tools and technology to handle these thorny issues.

GIS technology as well as its methods of spatial analysis are therefore based on geographic information science, which is supported and widely applied by various disciplines of geography, surveying, engineering, space science, computer science, cartography, statistics, and others for purposes of spatial data organization, analysis, and distribution to eventually improve life.

Over time, the significance of spatial analysis increased. It utilizes massive amounts of various data formats to serve a wide-range applications of the con- cerned scientific disciplines. Each discipline uses special software and hardware that suits its mission and data formats. For this, recent practices of GIS and spatial analysis need advanced algorithms and programs to enable proper and timely handling and processing of immense and interdisciplinary data. This resulted in the spatial data science, which is the branch of data science that manipulates the spatial data.

## 3. Data science essentiality for geospatial big data analytics

Per the previously discussed issue of data inflation, the new discipline of data science was founded to handle and make use of these data floods and ensure getting the maximum out of them. The arena of this newly founded field, therefore, was the data with its different formats. This subsequently resulted in a number of new connected and related concepts and terminologies, such as big data, the internet of things, machine learning, data bases and warehouses, data archives, knowledge bases, expert systems, decision support systems, and many others. Forming the subdivisions and environment of data science. The below chart, highlights the nested and con- nected advanced technologies for data analytics (See Figure 2).

Big data is one of the prominent beneficial effects and usefulness of data science; nevertheless, it is still not a clearly defined term, and it has been defined differently according to each of the technological, industrial, research, or academic viewpoints [8]. However, in general, big data is considered structured and unstructured datasets with massive data volumes that cannot be easily captured, stored, manipulated, analyzed, managed, and presented by traditional and ordinary hardware, software, and database technologies [9].

The dire needs of analytics as well as newer methods and techniques of data sci- ence emerged as a response to the current enhancements in life aspects generally and the quest for better living conditions. Accordingly, humans—via various disciplines— strive to make use of the recent advancements in ICT researches and technologies to cope with reality's complexities through data and models. For this, big data emerged and prevailed in most of science fields. Geography and geospatial sciences generally have the most stake, as reflected in "geospatial big data". It was argued that "80% of the data is of a geographic nature.

Despite criticisms regarding the societal effects of big data, which has been accused of having limited benefits for society and violating issues related to data privacy, confidentiality, and security, geospatial big data, nonetheless possesses sig- nificant potential to greatly benefit various societal applications. These applications encompass areas such as climate change, disease surveillance, disaster response, the monitoring of critical infrastructure, transportation, and more [9].

Geospatial big data processing techniques encompass a range of methodologies designed to handle and analyze vast volumes of geospatial data due to programming models and frameworks for distributed processing of geospatial data [10]. These tech- niques are vital in extracting valuable insights from geospatial data, aiding decision-making in a spectrum of various domains such as urban planning, environmental monitoring, and disaster management.

In order to be handled and processed, the available big data volumes need to be compressed without compromising data quality [8]. The parallel processing and distributed computing frameworks, therefore, came and leveraged the computational demands of geospatial big data [11]. These all go beyond the capacity of the human mind to handle.

Reading and understanding big data patterns and trends—the minimal level of benefits—is not possible without the aid of computerized methods of summariza- tion and visualization. Further benefits could be yielded by deploying knowledge discovery algorithms and data mining techniques for the extraction of more valuable and well-oriented knowledge that serves problem solving, decision support, and even strategic policy-making on the basis of deep understanding.

In the same context, visualizing this variable and complex data needs a computer- ized and graphical capability to facilitate communication of these data. The data in its real forms and row formats is not understandable and readable, and both forecasting future scenarios and making the right decisions for the present based on the available data are not possible without developing algorithms and complicated models. The human mind cannot handle the complicated calculations for the big data volumes.

This necessitated and urged recent data science to embody.

## 4. Spatial analysis in the information era: the emergence of spatial data science

The ultimate goal of science, the applied and life sciences in particular, is to support humankind and equip it with the facilities needed for a prosperous life. Science, thus, augments sustainability by employing advanced analytics and machine learning techniques to extract actionable insights from vast and diverse geospatial datasets. This analytical ability enables proactive identification of environmental trends, prediction of risks, and optimization of resource allocation strategies [12].

Historically, observation—the direct field observation—was the sole prevailing source of data, and decision-making was not complicated, so the human mind was capable of handling that. However, the informatics insurgency came to digitalize and automate most activities and life aspects. In science, data collection, processing,

storage, and visualization had been profoundly reformulated by computerization and informatics, which led to the information era.

In the Information era, spatial analysis has emerged as a crucial discipline, facilitated by data availability and advancements in technology. Spatial analysis further encapsulates this transformative trend and highlights the significant role of spatial data science in extracting insights from vast and complex geospatial datasets. Therefore, emergence of spatial analysis and geospatial sciences is featured by their capacity to address multifaceted challenges across diverse domains. See the below chart, Figure 3. It demonstrates the spatial data science emergence.

In its generic and plain definition, spatial analysis refers to the process of exam- ining, modeling, and interpreting patterns and relationships within geographic data to gain insights and make informed decisions about the spatial distribution of phenomena. Spatial analysis further encompasses a wide range of techniques and methods used to analyze spatial data, including mapping, clustering, interpola-

tion, and geostatistics [13].

The overall goal of spatial analysis and its processes can be categorized into two levels of ultimate knowledge generation. These levels vary between the identification and classification of objects and phenomena, revealing patterns as well as under- standing the relationships among these phenomena—how objects interact and affect each other. The spatial analysis, thus, strives to model and abstract reality, particu- larly what regards geography and has spatial dimension, to enhance understanding and, subsequently, enable better decision-making and problem solving.

Thus, the complex reality and synergy of geospatial technologies, data science methodologies, and advanced analytics insisted and compelled a deeper understand- ing of spatial relationships, patterns, and trends, and was the driver the paradigm shift in spatial analysis. This caused the spatial analysis to play a crucial role in various fields such as geography, environmental science, urban planning, and epidemiology, [14].

Furthermore, due to its ability to reveal hidden patterns and trends that may not be detectable through the traditional statistical analysis, spatial analysis gained spe- cial interest among researchers and practitioners. This novel approach thus become inevitable in sophisticated applications of natural resources management, urban planning, environmental monitoring, epidemiology, and all spatial-related fields.

This is what enabled it to drive evidence-based decision-making and contributing to more sustainable and insightful policies [11].

In the contemporary context, the intricacies of life are induced by the fusion and interplay of phenomena, exemplified, for instance, in the nexus of environment, economy and society besides, the global landscape characterized by swift transfor- mations and changes, wherein the accumulated data and knowledge are combined with complicated decision-making processes. These processes, in their multifaceted nature, require the consideration of numerous dimensions simultaneously.

These facts necessitate reading the massive data volumes, of different sources and formats, available in databases, data warehouses, and obtainable knowledge bases to fully understand the current state of affairs. Spatial indexing and query optimiza- tion would be the right and ideal techniques for that, as they enable organizing and handling various data besides creating efficient spatial indexing structures and query optimization methods to accelerate spatial queries [15].

Geospatial Data Mining, in the same context—taking advantage of the immense availability of integrated and multi-sourced data—emerged to apply data mining and machine learning techniques to extract patterns and relationships from geospatial datasets [16, 17]. Especially these data, in order to give meaningful values, should be subjected to careful and rigorous analysis through well-oriented questions that extract insights and valuable knowledge for problem solving and decision-making out of those massive data volumes. Geospatial Data Mining provides an effective, efficient, and fast mechanism to do so.

## 5. Common data science processes and approaches in GIS context

The application of data science to different disciplines is inevitable in the era of technology and informatics. It was previously elaborated how data science had practi- cally affected the various scientific fields. Geography is not an exception and had been one of those affected the most. This, along with the recent merger of geography and data science resulted in the newly found branch of geography, known as geographic information systems. Some scholars went beyond that and perceived GIS as an independent discipline "geospatial data science".

Definitions of GIS unanimously agreed that GIS is basically the science of spatial data collection, processing, and visualization. All of GIS's functions have thus been deeply altered by data science. Data science has a number of manifestations in the GIS world. For instance, transforming the way spatial data collected and interpreted alongside the inclusion of spectrum of methodologies that leveraged GIS analyt- ics, machine learning, and data mining techniques to extract valuable insights from geospatial datasets.

Prior delving into the details manifesting data science role in GIS world, the fol- lowing diagram gives a quick glance and outline the GIS and data science synthase, by means of highlighting the techniques and approaches of data science GIS had adopted and their respective applications and uses (See Figure 4).

In the GIS world, data preprocessing is a foundational step involving data clean- ing, transformation, and integration to ensure the quality and consistency of geospa- tial datasets [16, 17]. This aligns with data science's emphasis on data preparation for meaningful analysis.

The infusion of data science processes and approaches into the GIS context has ushered in a new era of spatial analysis and understanding. Data fusion and integra- tion are among essential data science processes in the GIS context. They allow com- bining data from diverse sources, such as satellite imagery, maps, and other forms of statistical and descriptive data, from which analysts can create enriched datasets that offer comprehensive insights into various phenomena [18]. This is particularly valu- able in disaster management, where the integration of "nearly" real-time data streams aids in tracking, monitoring and responding to natural hazards.

In recent years, along with the availability of new sensors, new ways of collecting geospatial data have emerged [7], which means the digital data collection methods introduced and accompanied by data science have shifted data collection and trans- formed it from sparse data to rich and plentiful databases. This situation had been constantly developing until the introduction of spatial data infrastructure (SDI). Data collection, likewise, had evolved and reached the recent state of availability of nearly real-time data and, at certain times, real-time data.

Awing to SDI and massive existing geospatial data, geospatial data exploration and visualization which is one of core data science practices found resonance within GIS. Interactive maps, spatial dashboards, and geo-visualizations provide insights into spatial patterns and relationships, aiding decision-making and the communication of complex information to all different audiences [19]. Furthermore, spatial cluster- ing and pattern recognition, are key techniques in data science that are extensively deployed in GIS analytics for identifying hotspots, spatial trends, and clusters in geospatial data. (See Figure 5) [20].

Moreover, spatial autocorrelation and spatial regression techniques, such as geo- graphically weighted regression, have gained prominence in both data science and GIS. In GIS, they allow the identification of spatially varying relationships among given variables, en-

hancing predictive accuracy in local contexts [13]. Another form of spatial clustering is the process of grouping spatial data points based on their similarity. Methods like K-Means clustering are applied in GIS to fragment geo- graphic data into clusters, revealing spatial patterns and helping to identify regions with similar characteristics [22]. This has frequent use in urban planning, where it aids in delineating distinct neighborhoods based on their common socioeconomic attributes.

Deep learning methodologies and Machine learning algorithms developed by data science have leveraged a wide range of GIS processes. Deep learning tools, such as Convolutional Neural Networks (CNNs), have enriched object detection and image classification within GIS domain. CNNs have proven to be effective in identifying objects and features in satellite images [23]. While machine learning facilitated predictive modeling and image classification tasks. Supervised and unsupervised machine learning methods grew in popularity as a frequent tool for land cover clas- sification, species distribution modeling, and urban growth prediction [24]. See the below diagram.

One of the key machine learning algorithms for spatial prediction and classifica- tion is Random forest. Random forest enables the creation of predictive models that account for spatial relationships, allowing for the identification of trends and patterns in complex geographical data [12]. For instance, ecologists, in order to model species distribution, utilize machine learning to foresee the spatial distribution of species based on certain environmental variables. Likewise, Machine learning algorithms

like random forest and support vector machines leverage spatial data layers to predict outcomes like land cover changes or disease outbreaks [25].

Therefore, by harnessing the power of data science, GIS practitioners can generate accurate predictions and enhance their understanding of complex spatial patterns.

Wherein large and diverse datasets are processed and analyzed to extract meaning- ful insights. This involves techniques like data clustering and spatial interpolation, which facilitate the extraction of valuable knowledge from vast amounts of geospatial information [26].

Regarding the institutional part of GIS, the integration of data science and GIS is also evident in spatial optimization problems. Likewise, location-allocation models, the classic GIS application, had important benefits from data science optimization algorithms to determine optimal facility locations and resource allocations [27].

This fusion of practices and the interdisciplinary approach enhanced the efficiency and accuracy of GIS as decision support systems.

Integration of data science processes and approaches has, consequently, become increasingly prevalent in GIS practices, ranging from data collection and prepro- cessing up to the final stage of laying out the results. These all have led to enhanced understanding and utilization of spatial information. Hence, a transformative shift in the capabilities of GIS enabled more sophisticated analysis, modeling, and visualiza- tion of geospatial information.

## 6. GIS, KM and data science: the indispensable sustainability nexus

The relationship between Geographic Information Systems (GIS) and Knowledge Management (KM) is interdependent, with GIS serving as a powerful tool to facilitate various aspects of knowledge management. And KM enhances the management and utilization of geospatial information, fostering a structured approach to acquiring, storing, sharing, and utilizing spatial knowledge. This interplay enhances the acces- sibility, organization, and utilization of geospatial information within both GIS and knowledge management frameworks. The following diagrams explain the main func- tions and processes of knowledge management and their cyclic nature respectively (See Figures 6 and 7).

Knowledge management prospects, therefore, are obvious in the contemporary world of GIS. GIS and Knowledge management have recently had a two-way con- nection; on the one hand, GIS borrowed approaches and techniques from KM for its practices as well as taking the applications and uses of KM and applying them to spatial data. On the other hand, KM takes advantage of GIS as a platform to integrate and utilize multi-sourced data in different formats, all for one unified and oriented purpose. This is encapsulated and briefed in the below table, then the details later (See Table 1).

The most prominent knowledge management methods and techniques adopted by GIS encompass a variety of approaches that, together, aim at effectively captur- ing, organizing, and utilizing geospatial information. To mention and elaborate on some:

1. Geospatial data integration that makes use of knowledge management practices and guides the categorization, tagging, and metadata documentation of this data for effective data organization. KM allows GIS to put together and integrate various geospatial data, for instance, spatial and attributive data, from diverse sources and formats to create an inclusive and coherent knowledge base [30]. This will enable comprehensive spatial analyses, allow users to consider a wide range of variables as well as facilitate well-oriented goals and problem-solving out of them.

2. Geospatial data cataloging and metadata management, i.e. implementing meta- data protocols would lead to standardized documentation, hence easier manage- ment of geospatial data, and further enhance data discoverability and usability [31]. In other words, easier and faster running of queries. Thus, efficient meta- data management and documentation ensure geospatial data's proper utilization along with enhancing data discoverability, reusability, and all other aspects of data quality.

3. Visualization and spatial insights, GIS —thanks to its super powerful tools, like interactive maps —provides visual representations of spatial data, making com- plex information more understandable. These visualizations put more than one variable together, which aids comparability and enables knowledge managers and stakeholders to grasp spatial relationships, patterns, and trends.

4. Spatial analysis documentation, as knowledge management is a tremendous tool to capture analytical processes. This enables GIS users to replicate spatial analyses, ensuring transparency and facilitating knowledge sharing among researchers.

5. Spatial data infrastructures (SDIs), establishing SDIs would enable enforc- ing data quality measures while handling and interacting with data. It further facilitates seamless sharing, discovery, and access to geospatial data across organizations and sectors [32]. Additionally, SDIs combat redundancy in data and result in efficient storage control.

6. Development of geoportal and web-linked data, the main pillar of nowadays web GIS is creating and operating geoportals as centralized platforms for accessing and sharing geospatial information to foster collaboration and knowledge dissemina- tion [33]. Semantically-related data sharing is similarly a web-based. Deploying web-based semantic data sharing technologies revolutionized geospatial data distribution and enhanced the interoperability of and among geospatial data [34].

Knowledge management applications and uses borrowed and adopted to serve in the environment of GIS involve processes and techniques leveraging GIS to achieve its ultimate goals through effectively utilizing the geospatial information. These involved key applications and prominent uses include, among many:

1. Decision support: GIS-based knowledge management aids decision-making by pro- viding spatial insights into knowledge management strategies. KM processes guide the identification of relevant data layers and spatial variables for informed deci- sion support. Through allowing stakeholders to analyze and visualize data to make informed choices in fields like urban planning and disaster management, [35].

2. Collaborative mapping: collaborative platforms offered by GIS enable users to contribute and share data and insights, fostering crowd-sourced mapping efforts for various purposes like community engagement and environmental monitor- ing. This is obvious when taking advantage of parallel computing. GIS platforms offer collaboration by allowing multiple users to access, process, and share geospatial data and analyses simultaneously. While KM principles facilitate the creation of workflows and protocols for collaborative data sharing, [36].

3. Scenario planning: GIS-driven knowledge management aids in scenario planning by modeling different spatial outcomes and assisting in policy development and resource allocation. KM strategies and workflows help capturing the rationale behind these models, documenting assumptions and methodologies for future reference and analysis.

4. Environmental monitoring: GIS utilizes methods and frameworks originated by the knowledge management discipline to support environmental management through constant environmental monitoring, enabling authorities and decision- makers to monitor and track changes, reveal trends over time, and make in- formed conservation decisions.

5. Emergency response: Knowledge management in a GIS environment assists emergency response in real-time by providing up-to-date geospatial data for incident management. The synthesis of GIS and knowledge management would also enable the creation and proper operation of early warning systems and em- power preparedness for emergencies.

6. Spatial query and retrieval: GIS enables spatial queries for data retrieval. KM techniques assist in designing efficient query systems and indexing approaches for quick and efficient access to relevant geospatial information.

For these reasons, one can state that the integration of geographic information systems (GIS), knowledge management (KM), and Data Science has given rise to a transformative paradigm in the realm of sustainability. This three-way relationship represents an indispensable nexus that offers holistic solutions to complex environ- mental and societal challenges (See Figure 8).

## 7. Conclusion and future directions

This chapter was founded on demonstration and argumentation crowded of cited evidences in order to prove and support of the argument "synergy of data Science, GIS spatial analysis, and knowledge management is a path toward sustainability insights". In a world characterized by the exponential data growth due to the rapidly changing states and interplay phenomena. GIS, KM and data science nexus emerged to enable establishing of groundworks for understanding those interconnect and shared attributes, and offer a transformative approach for unraveling complex sustainability challenges and the need for informed decision-making.

To Highlight these, the chapter dived into the evolution of data accumulation from historical milestones to the modern digital era. Then it probed the epistemological foundations of Data Science and its profound influence on various scientific domains. Moreover, strived to uncover the manifold manifestations of data science processes approaches deployed in GIS practices ranging from data collection, preprocessing, statistical analysis, to machine learning algorithms and visualization techniques, each one has been examined in the context of sustainability research.

Being in a current world of informatics, essentiality of data science in the realm of geospatial big data analytics, likewise, had been explored in the context and roles of computational analysis in harnessing vast data for sustainability sake. This in response, subsequently, led the emergence of spatial data science field and its contributions to understanding spatial relationships and patterns. Through, their tremendous techniques that enhance comprehension, modeling, and prediction.

Concluding with the syntheses of GIS, knowledge management, and data science forms an inevitable nexus for achieving sustainability agendas. This integration of geospatial sciences and technologies, enabled data organization, knowledge represen- tation, and collaborative decision-making is explored in the context of sustainability insights.

## REFERENCES

[1] Bolstad P. GIS fundamentals: A first text on geographic information system. In: Manual of Geospatial Science and Tech-

nology. 5th ed. White Bear Lake, Minnesota, USA: Eider Press; 2016

[2] Eltayeb G. Geography, Environment and Development: The Fundamentals and Roles. Sana'a, Yemen: Dar Elhikma Elymaniya; 1995

[3] Osman BT. Geographic Information Systems, GIS. Khartoum., Sudan: Open University of Sudan; 2006

[4] Osman BT. Decision making support and spatial analysis in geographical information systems. Geographic Letters Series, 58, Kuwait geographic society & Geography Department. Kuwait: Kuwait University; 2003

[5] Scheider S, Nyamsuren E, Kruiger H, Xu H. Why geographic data science is not a science. Geography Compass. 2020;14(11):1-15. DOI: 10.1111/gec3.12537

[6] Osman BT. GIS applications in science and engineering research: Case study of the graduate College University of Khartoum. Sudan Geographical Journal (SGJ). 2017;1(1):1-2. Khartoum, Sudan

[7] Li S, Chen J, Liu C. Overview on the development of intelligent methods for mineral resource prediction under the background of geological big data. Minerals. 2022;12:616. DOI: 10.3390/min12050616

[8] Chen X, Li J, Yu L, Zhao Y. Compressing and exploring massive GPS trajectory datasets. ISPRS Journal of Photogrammetry and Remote Sensing. 2015;105:13-24

[9] Li S, Dragicevic S, Castro FA, Sester M, Winter S, Coltekin A, et al. Geospatial
big data handling theory and methods: A review and research challenges. ISPRS Journal of Photogrammetry and Remote Sensing. 2016;115:119-133

[10] Li Z, Li J, Wu F, Tian Z. A MapReduce-based parallel spatial join algorithm for processing big geospatial data. ISPRS Journal of Photogrammetry and Remote Sensing. 2014;98:119-132

[11] Li Z, Li J, Wu F, Tian Z. Parallel computing for processing massive geospatial data using MapReduce. Computers, Environment and Urban Systems. 2017;64:225-237

[12] Gao S, Tang W, Li Q , Yang J. Data science and urban sustainability: A review. ISPRS International Journal of Geo-Information. 2019;8(1):23

[13] Fotheringham AS, Brunsdon C, Charlton M. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Chichester, England: John Wiley & Sons Ltd; 2003

[14] Openshaw S, Abrahart RJ. Geographical information systems and spatial data analysis in the social sciences.
GeoJournal. 2000;50(2):213-224

[15] Zhang R, Lin H, Liu Z, Shekhar S. G-tree: An efficient and scalable index for spatio-temporal data on multicore processors. IEEE Transactions on Knowledge and Data Engineering. 2012;24(5):903-917

[16] Shekhar S, Evans M, Tan X, Zhang H. Advances in geospatial big data handling. IEEE Transactions on Knowledge and Data Engineering. 2014;27(6):1510-1523

[17] Shekhar S, Evans M, Tan X, Zhang H. Advances in Geospatial Big Data Handling: Tools and Methods. Boca Raton, Florida, United States: CRC Press; 2014

[18] Mennis J, Hultgren T. Intelligent GIS: Location decisions and strategic planning. In: The GIS Book. Boca Raton, Florida, United States: CRC Press; 2006. pp. 245-263

[19] Khan O, Couclelis H. GIS and the "spatial turn" in information technology. Geographical Information Science. 2006;12(4):1-11

[20] Duque JC, Anselin L, Rey SJ. The max-p-regions problem. Journal of Regional Science. 2012;52(3):397-419

[21] Kaliski S, Mani A. How GIS and Machine Learning Work Together. California, USA: ESRI Redlands; 2019. Available from: https://www.esri.com/ content/dam/esrisites/en-us/ about/ events/media/UC-2019/technical- workshops/tw-6165-494.pdf

[22] Arbia G. A Primer for Spatial Econometrics: With Applications in R. London, UK: Palgrave Macmillan; 2014

[23] Zhang Z, Zhang J, Shen HT. Rapid object detection in satellite images using convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing. 2016;54(11):6555-6565

[24] Bui TD, Lofman O, Revhaug I. Machine learning applications for geomorphological mapping using GIS data and remote sensing images: A review. Geomorphology. 2019;327:91-102

[25] Wang S, Zhang J, Li Z, Li J. Spatial prediction and accuracy assessment of land use and land cover changes using remote sensing and GIS techniques.Remote Sensing. 2016;8(11):907

[26] Li Z, Li J, Wu F, Tian Z. A comprehensive review of geospatial big data management. ISPRS International Journal of Geo-Information. 2020;9(7):399

[27] Church RL, Murray AT. Business Site Selection, Location Analysis, and GIS. Hoboken, New Jersey. USA: John Wiley & Sons, Inc.; 2009

[28] COMPASS INORMATICS. 2023. Available from: https://compass. ie/services/gis-and-data-analytics/ [Accessed: August 16, 2023].

[29] Andreev I. Knowledge Management. VALAMIS; 2022. Available from: https:// www.valamis.com/hub/knowledge- management, https://www.valamis.com/ hub/institutional-knowledge

[30] Sui D, Goodchild MF. The convergence of GIS and social media: Challenges for GIScience. International Journal of Geographical Information Science. 2011;25(11):1737-1748

[31] Craglia M, Annoni A, Proïa F. GSDI Cookbook: Developing Spatial Data Infrastructures. Joint Research Centre of the European Commission. Redlands (United States of America): ESRI Press; 2008

[32] Rajabifard A, Feeney ME. Spatially enabling government: Opportunities, barriers and future directions. Land Use Policy. 2013;31:574-583

[33] Liang SH, Yang CF, Liu WL, Chen KC. Development and evaluation of a cloud-based geospatial information platform. ISPRS International Journal of Geo-Information. 2015;4(4):1982-2001

[34] Janowicz K, Hitzler P, Adams B, Kolas D, Vardeman C,

Henson C. On the strategic value of linking ontologies and linked data for the geospatial sciences. Semantic Web. 2018;9(4):545-563

[35] Rinner C. Neogeography: The challenge of augmenting the geospatial web. Geographic Information Sciences. 2003;9(2):129-138

[36] Dalkir K. Knowledge Management in Theory and Practice. London, UK: Routledge; 2011. DOI: 10.4324/9780080547367