Revista Geográfica de Chile Terra Australis, Vol.60 (2024) 102-119 ISSN 0719-9562

https://doi.org/10.23854/07199562.2024601.alciaturi

A WAY FOR EXPLAINING GEO BIT DATA THROUGH ITS CHARACTERISTICS, SOURCES AND CENTRAL SUPPORT TECHNOLOGIES

UNA MANERA PARA EXPLICAR EL GEO BIG DATA A TRAVÉS DE SUS CARACTÉRISTICAS, FUENTES Y TECNOLOGÍAS DE APOYO FUNDAMENTALES

MSc. Giancarlo Alciaturi[1], PhD. María del Pilar García - Rodríguez[2] y PhD. Virginia Fernández[3].

ABSTRACT

Big Data has significantly impacted scientific research and everyday life, resulting in solutions that align with its principles. Therefore, it is essential to propose ways that enable a basic understanding of specific fields of knowledge, such as those related to spatial domain information. Through a narrative review, this document explains the fundamentals of Geo Big Data by contextualising the characteristics, sources, and central technologies of Big Data in relation to geographic information. One of the most significant findings is that Geo Big Data helps enhance the understanding of biophysical and human dimensions of spatial matters. The technology boosts the acquisition of valuable insights and opportunities across various areas, including risk management, health, agriculture, environment, and open government. This document lays some groundwork for implementing Geo Big Data initiatives towards more specific efforts.

Keywords: Geo Big Data, Geographic Information, Geography.

RESUMEN

El Big Data ha tenido un impacto significativo en la investigación científica y en la cotidianidad. Ello ha llevado al desarrollo de soluciones que se alinean con sus principios. Dado el contexto, es esencial proponer elementos que orienten hacia una comprensión básica de campos específicos del conocimiento, como aquellos relacionados con el dominio espacial. Mediante una revisión narrativa, este documento explica los fundamentos del Geo Big Data al contextualizar en el ámbito de la información geográfica; las características, fuentes y tecnologías principales del Big Data. Dentro de los hallazgos más significativos, se determina que el Geo Big Data es una alternativa para potenciar el entendimiento de los fenómenos espaciales tanto en la dimensión biofísica como en la humana. La tecnología facilita la adquisición de valiosos conocimientos y oportunidades en diversas áreas, incluyendo gestión de riesgos, salud, agricultura, medio ambiente y gobierno abierto. Este documento sienta algunos elementos para orientar iniciativas Geo Big Data hacia esfuerzos más específicos.

Palabras clave: Geo Big Data, Información Geográfica, Geografía.

[1] Programa de Doctorado en Geografía, Universidad Complutense de Madrid, España.
[2] Departamento de Geografía, Universidad Complutense de Madrid, España.
[3] Departamento de Geografía, Universidad de la República, Uruguay.

INTRODUCTION

BD is a revolutionary approach for acquiring, processing, and generating valuable knowledge. It is mainly related to computational technologies with more significant capabilities than the standards.
From the perspective of the spatial domain information of BD, Geo Big Data (GeoBD) has previously been quoted by Goffi et al. (2020) and Brovelli et al. (2019). The meaning points to georeferenced BD, which generally provides sets with a relatively high spatial and temporal resolution for analysing complex phenomena. In this recent field, contributions come from different but traditionally complementary focuses: the human and social sciences, the Big Earth Data (BED) or a focus that connects both.

From human and social sciences, Rowe (2021) demonstrates that GeoBD offers unprecedented opportunities to transform our understanding of the social world and human behaviour. According to Thatcher et al. (2018), the most important empirical interventions are urban and smart cities, social media mapping, geoweb and digital humanitarian research, geosocial footprints, and combining demographic and social media data. Gutierrez-Puebla et al. (2016) affirm that these interventions may be developed from data surveyed via mobile devices, social media, geolocated photographs, accommodation services, bank card transactions, and intelligent transportation cards. It is important to note that the availability of information and communication technologies has promoted quoted opportunities. However, expanding broadband access remains a challenge, especially in developing countries where only 35% of the population has access to the Internet. In contrast, advantaged economies have an access rate of about 80% (The World Bank, 2023a).

Advances in Earth Observation (EO), such as the launching of new satellites, improvements in its sensor's resolution, OD policies and the increased demand for the applications, have led to innovative approaches in storing, managing, processing, and analysing remote sensing data under the Big Earth Data (BED) philosophy (Taylor & Francis, 2021). It primarily enables data-driven analysis, interpretation, and understanding of earth-based interactions. Imagery is massive, multi-source, heterogeneous, multi-temporal, multi-scalar, highly dimensional, complex, nonstationary, and unstructured (Zhu, 2019).

The advantages represent demanding computational requirements that have been principally satisfied by data cube technologies, online portals, user-oriented web services and Analysis Ready Data (ARD).
Although most BED applications are primarily associated with natural sciences or biophysical, various initiatives have successfully been deployed to understand the human dimension. Yong et al. (2022) demonstrate the capabilities of the Défense Meteorological Program Operational Line-Scan System (DMSP - OLS) to evaluate poverty in China. Using this source, Biao et al. (2022) describe the population's spatial-temporal evolution pattern and compare its agglomeration and migration characteristics in coal mine concentration areas and cities like Linfen - China.

Today's world dynamics have led academia, governments, and several organisations to focus on areas where spatial trends are well explained by integrating social and environmental dimensions. Sustainable Development Goals (SDGs), Disasters Management and Mitigation (DMM), and Environmental Epidemiology (EE) are cases. SDGs watching could be supported by remote sensing and socially integrated analysis that complements official statistics (Allen et al., 2021). EE focuses on physical, chemical, and biological agents in the environment as disease risk factors, habitually affecting large populations (Bloom, 2019). As critical points, Tran et al. (2016) highlight that environmental conditions may influence disease outbreaks.

Risk models can be deployed from data drawn from remote sensing and epidemiological, entomological, and sociodemographic data acquired via BD. For instance, Tobias et al. (2021) studied relationships between greenness and predicted SARS-CoV-2 (COVID-19) disease incidence in the United States and the United Kingdom. As the main conclusion, the authors demonstrate that higher levels of greenness assessed through the Normalised Difference Vegetation Index (NDVI) may reduce the risk of predicted COVID–19 illness incidence. Rosenblum et al.

(2021) demonstrated that BD applies to disaster DMM after floods, earthquakes, fires, and flood, fire, and explosions. The rationale is as follows: the GeoBD means varied possibilities for providing, visualising, studying, and predicting natural disasters.
A narrative review of scientific and grey literature is carried out to provide insights into recent theories and evidence from diverse viewpoints, emerging trends, and broader perspectives

This document provides a comprehensive and general framework for explaining Geo BD's characteristics, sources, and central support technologies. The last only refers to various means for analysis purposes.

The results aim to select balanced and relevant content from the social sciences, BD, and practices that connect both. However, some limitations exist, such as non-reproducibility related to the authors' influence on searching sources and analysis. This content aims to encourage further exploration and potential advancements around Geo BD.

THE CHARACTERISTICS, SOURCES, AND CENTRAL SUPPORT TECHNOLOGIES OF GEO BIG DATA

THE GEOBD CHARACTERISTICS.

BD encompasses volume, value, velocity, variety, visualisation, and veracity.

VOLUME.

The volume has been encouraged by the vast number of sources that generate data second by second. For example, Statista (2021. a) estimated the growth of data generated worldwide between 2010 and 2024 (Figure 1). The situation reveals substantial interannual increases, representing a difference close to 7,500%.

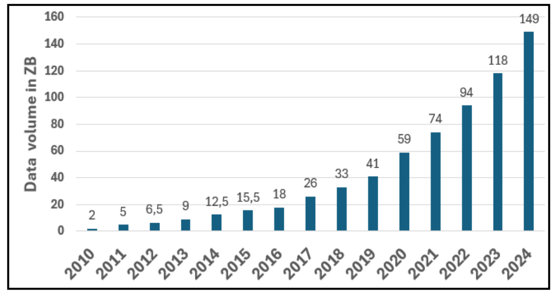

Figure 1. Data worldwide generated. 2010 – 2024. Source: Statista (ob. cit).

Measuring GeoBD merely based on volume as a distinctive characteristic can be difficult. Studies conducted by Aguirre et al. (2021) and Li & Biljecki (2019), who analysed Airbnb's spatial impact on Mexico and China, respectively, revealed that GeoBD can be obtained from databases (DBs) stored on personal computers. However, analysing these DBs requires using Data Science (DS) or Artificial Intelligence (AI) due to the large number of records. While none of the authors specifically mention the DB size, a practical exercise provides an idea about this: by consolidating all the city-level georeferenced records provided by Airbnb (2021), created a 175 MB DB that contains more than one million two hundred thousand records (Figure 2) that provide information about latitude, longitude, name of the housing provider, neighbourhood, prices, availability, and user's opinion.

Figure 2. Airbnb’s worldwide locations. 2021.

VALUE.

The actual value of GeoBD is realised when it is transformed into knowledge and used to make informed decisions. Spatial analysis techniques offer a way to unlock the value of specific data conglomerates. In this regard, artificial intelligence (AI) and data science (DS) have evolved into areas that provide knowledge for leveraging Geographic Big Data (GeoBD). These methods correspond to an evolution related to previous proposals widely used in the study of geographic space.

The value of GeoBD is critical in various situations, such as The United Nations (UN) Sustainable Development Goals (SDGs), humanitarian projects, geospatial marketing, and business intelligence. The UN (2021) exhibits how GeoBD can reveal value in various SDGs. For example, SDG 9 (Industry, Innovation, and Infrastructure) states that data acquired through GPS and mobile telephony can improve public transport and vehicular traffic in urban environments. Similarly, SDG 13 (Climate Action) reveals that satellite images, people's testimonies, and open-access data can help track deforestation processes.

Geospatial marketing and business intelligence rely on the value of GeoBD to make informed decisions. UBER Technologies Inc. (UBER, 2021) developed a platform capable of analysing vast data from its transportation service operations through mobile applications. One of the platform's objectives is to support decision-making that links the spatial component. For example, the platform can choose the best route for driving by analysing factors affecting mobility such as traffic, travel times, user demand, and available carriers.

VELOCITY

According to Kitchin & McArdle (2016. a), velocity can be divided into two types: the frequency of generation and the frequency of handling, recording, and publishing. The latter is related to the ability to collaborate in real-time with various organisations' spatial databases and perform spatial analysis tasks using AI. Velocity is relative, as the data acquisition frequency varies depending on the phenomenon being analysed. For example, concerns like automotive mobility, election reports, or financial statistics may require real-time data acquisition, while land use and coverage (LUC) dynamics or many natural sciences topics may not.

Specialised tools and techniques, such as stream processing or real-time analytics, are necessary to manage resources efficiently and operate GeoBD's velocity. For example, financial institutions like Banco Bilbao Vizcaya Argentaria (BBVA, 2023) provide data from selected countries for "Monitoring the Economy in Real Time and High Definition". Underneath this scope, users find a summary with information from this bank's economic analysis; nevertheless, academic or research institutions may ask for a complete database. This information may be helpful for investment tracking by economic sectors and regions, providing a meaningful "time advantage" to analysts and policymakers or analysing critical events such as the Covid-19 pandemic in a very granular way to design better "targeted" policies (Ortiz & Rodrigo, 2020).

About BED, current initiatives include Google Earth Engine (GEE), The African Regional Data Cube, The Australian Geoscience Data Cube, The Brazil Data Cube, Mexico's Geospatial Data Cube, and the Multi-Mission Algorithm and Analysis Platform initiative led by the European Space (ESA) and the National Air and Space Administration (NASA) are highlighted. GEE remains one of the most widespread free-access alternatives to searching and processing vast amounts of data acquired from different missions for EO (Zhao et al., 2021). For reference, a Scopus search reveals that between 2015 and 2022, almost three thousand documents were completed under GEE as a term in article titles, keywords, or abstracts. One case that stands out is Simonetti et al. (2021) because an application was developed to intuitively generate and explore Sentinel 2 multitemporal mosaics with 20 m spatial resolution and layers representing the potential changes in the land cover during a reference year. For example, a multitemporal 20-meter mosaic (Figure 3), whose coverage represents the territory of approximately 176,215 km^2 that continental Uruguay has, was virtually created in less than two minutes. The described alternative supposes significantly fewer work times than the traditional mechanisms that require downloading each scene individually to apply the processing routines later. A noteworthy resource is the Dynamic World Project (Brown et al., 2022) because it allows obtaining an open, continuous feed of LUC data in parallel with Sentinel-2 satellite acquisitions. This is a revolutionary tool as it enables the creation of global coverage and medium spatial resolution products.

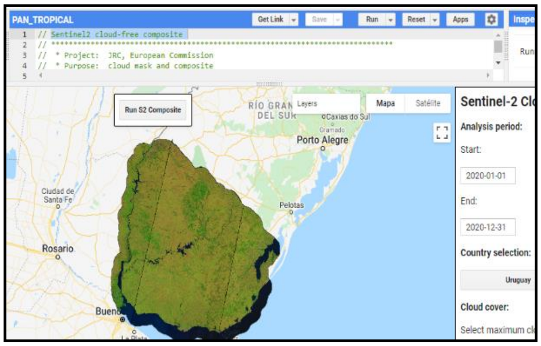

Figure 3. An example of a GEE application

Microsoft (2022) is working on the Planetary Computer Platform. When fully available, this resource will provide a multi-petabyte catalogue of global environmental datasets that will allow flexible scientific research to resolve questions through geospatial techniques. The main applications are global LUC classification, deforestation risk analysis, ecosystem monitoring, conservation planning, and forest carbon risk assessment. Regarding the processing under a license, consortia, such as the Environmental Systems Research Institute (ESRI, 2021), offer ways to process, visualise and analyse raster archives under the CC scheme through Amazon Web Services (AWS). High-Performance Computing (HPC) capabilities support these technologies.

VARIETY

Variety means data properties (structure and content) (Table 1) and using multiple data sources to solve tasks.

Properties	Type	Main content - characteristics
Structure	Structured	Has a predefined data format
	Semi-structured	Doesn't fit neatly into traditional structured models like relational databases, but still has some level of organization,
	Unstructured	Data is not arranged according to a preset data model or schema. Therefore, it cannot be stored in a traditional relational database or RDBMS. It has historically been challenging to analyse (MongoDB, 2022).
Content	Single media	Contents just one type of media.
	Multimedia	Represents BD that combines multiple media (e.g., text, images) that are utilised together.
	Graph Data	It is composed of a set of vertices and a set of edges. Real-world scenarios such as social media or GPS locations are examples.

Table 1. Geo BD variety dime and main characteristics

Several organizations have confirmed that multiple data sources can solve complicated spatial problems. For example, the United States Census Bureau (2022) and the United Nations Development Programme China (2016) operate GeoBD sources with official statistics to gain new insights beyond what the government and private sector can provide. In another instance, Vu et al. (2021) used freely available products like Sentinel 2 imagery and Open Street Map to create land use and cover maps for Ho Chi Minh City districts. They aim to use these data types to help cities in developing countries. In a third example, Sadiq et al. (2021) used satellite imagery and social media data to detect and understand human activity during flood events. They used AI models to process geolocated text and photos from Twitter and Sentinel 1 classification. Represa (2020) developed a methodological framework for air quality assessment that uses data from the Moderate-Resolution Imaging Spectroradiometer (MODIS) and surveys. The author validated the framework using three case studies in Spain and Argentina, including one without regular air quality monitoring.

VISUALISATION.

GeoBD poses new challenges for visualisation since good practices require programming and cartographic design skills to communicate a message that is likely to be quickly updated. These trends led Robinson and others (2017) to suggest the advent of a new paradigm in cartography because the diffusion of the message may require traditional cartography or advanced geovisualization techniques. For the first option, a vast amount of data must be an object of statistical synthesis like that provided by reducers (e.g., minimum, maximum, mean, median and standard deviation.) to aggregate data over time and space (GEE, 2023). As an example of traditional cartography, there is the map (Figure 4) developed by Bhimala et al. (2020) based on data from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor acquired during the period 2000 - 2016. The authors performed the Mann-Kendall statistical analysis to identify India's NDVI trends (p).

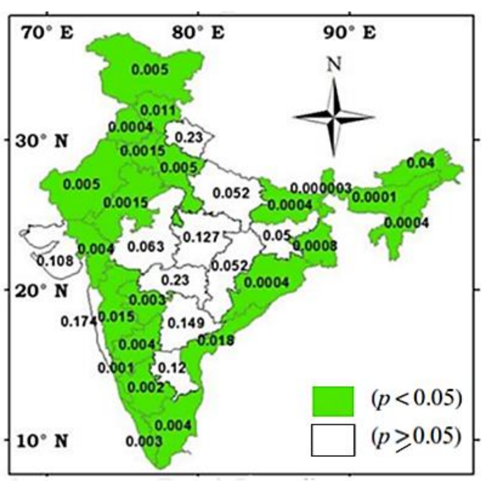

Figure 4. NDVI trend in India. 2000 – 2016 Source: Bhimala & others (ob. cit).

Geovisualization may involve dynamic cartography or web applications like dashboards. These tools enable the representation of dense and updatable datasets related to circumstances that connect with rapid decision-making scenarios.

A dashboard is a web-based user interface that organises and presents context-specific data to a broad constituency of users that may, such as experts, policymakers, politicians, and civil society (Young et al., 2021). The tool helps analyze human dynamics, especially those related to cities or matters of significant importance to society. It also promotes more efficient, fair, sustainable, and resilient communities through various shifts in thinking and practices. (Kitchin & McArdle, 2016. b). As a significant instance, Johns Hopkins University (2021) created a COVID-19 dashboard (Figure 5) to disseminate global statistics as a tool for authorities to take life-saving actions and mitigate the pandemic's effects.

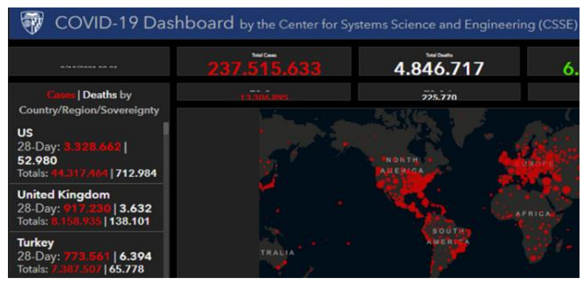

Figure 5. COVID–19 Dashboard Source: John Hopkins University (ob. cit).

There are visualisation engines available, such as deck.gl (https://deck.gl/) or carto (https://carto.com/) that enable users to produce outstanding results by creating preferred layers or utilising a flexible architecture to address specific requirements. In Figure 6, using the deck.gl services and data from DATA.GOV/UK official statistics, it is possible to analyse personal injury road accidents in Great Britain as a factor related to United Kingdom Road Safety. The platform mentioned can be easily configured to match user preferences and needs.

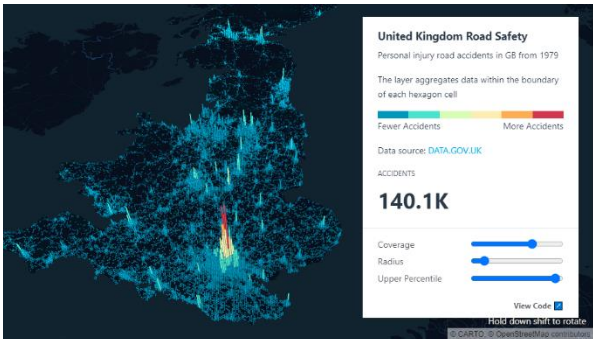

Figure 6. Personal injury road accidents are represented through the GeoBD visualiser. Great Britain, 1979 to 2022 Source: https://deck.gl/

VERACITY.

Veracity depends on sources, applications, and user-defined requirements (Ardagna et al., 2018; Ramasamy & Chowdhury, 2020). It can be measured by fulfilling the conditions of accessibility, accuracy, cohesion, confidentiality, credibility, lineage, and real-time analysis (Table 2).

Attributes	Characteristics
Accessibility	It is the quality and possibility of accessing data according to available technologies. OD is valuable in defining how scientific data may be published and re-used without price or permission barriers. OD policies enable the chance of accessing ARD or not for analysis.
Accuracy	Accuracy refers to the degree to which spatial data conforms to the correct value or standard. It has structural characteristics that involve geometrical, thematic, and temporal properties. Geometrical accuracy can be relative, absolute, horizontal, vertical, or gridded. Thematic accuracy relates to the quantitative or qualitative correctness of the data. Temporal accuracy refers to the exactness of a time measurement or the relevance of a data acquisition set.
Credibility	Credibility involves evaluating the transparency of data sources, methodologies, and processes to ensure spatially referenced datasets' trustworthiness, reliability, accuracy, and derived insights.
Lineage	Data lineage refers to tracking data flow over time, clearly understanding where it originated, how it has changed, and its definitive destination within the data pipeline. It provides a data record throughout its lifecycle, including source information and transformations. It enables users to observe and trace different touch points along the data journey, allowing organisations to validate for accuracy and consistency (IBM, 2022. a).

Attributes	Characteristics
Real-time analysability	It refers to the ability to analyse data efficiently and timely in real time. It involves processing and analysing data as it is generated without significant delays, enabling decisions or actions based on the most recent available information. In summary, it is about the capability to analyse data as it is generated instantly. For instance, access to georeferenced election data is critical in assessing the transparency of an electoral process (Open Election Data Initiative, 2022) because it makes it possible to add visibility to redistricting, streamline election management, encourage voter participation, or identify non-regular.

Table 2. Geo BD veracity properties and characteristics (cont)

THE GEO BD SOURCES.

Huang and Wang (2020) state that around 80% of BD holds spatial location data, indicating that most of these records are suitable for GeoBD practices. In this context, it is crucial to recognise the role of Open Data (OD) policies since they allow scientific data to be published and re-used without monetary cost or permission barriers. Commercial firms provide a range of products and services, including pre-packaged data and information products, data acquisition services, and customized data processing services.
In general, there are three primary sources (Figure 7): data generated by experts, citizens, or institutions (DEI), sensor systems and machines (SSM), and biometric operations (BO).

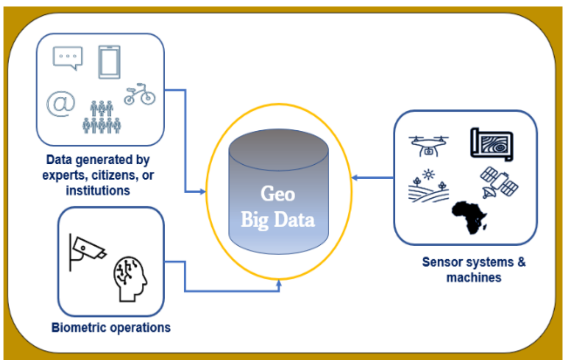

Figure 7. GeoBD sources

Data generated by experts, citizens, or institutions have been powered by the Internet of Things (IoT) and the massive availability of devices equipped with satellite navigation systems (GPS, GLONASS, BeiDou, or Galileo). Those elements are critical to creation through the web, social networks, and mobile monitoring technologies coupled on aeroplanes, ships, cars, motorcycles, bicycles, skateboarding, cellphones and even those intended for animal watching. The advancement in mobile technologies has led to the emergence of "Big Data Mobile" as defined by Cheng et al. (2018). To assess its scope, it can be noted that 60% of the global population was using the Internet in 2020 (The World Bank, 2023b) and around 4.2 billion people are active social media users (Statista, 2021b). For instance, Twitter had approximately 326 million users in mid-2020 (Zola et al., 2020). Beyond these figures, there is no global representativeness because data from only seventy-eight countries out of two hundred and sixty-four were available. In addition, the information from the African continent was very scarce because only the Ivory Coast, Egypt, Kenya, and Morocco reported the statistics.

The institutions, experts or citizens continuously feed the databases during a process that provides spatial location. The first concentrates on necessary elements to address public or private institutions' goals, missions, and mandates. The "expert" profile is typically associated with the technical or academic staff of the government or scientific community. The "citizens" contribute without a significant reference level for the analysed fact or phenomenon. This kind of participation has led to the rise of volunteered geographic information (VGI) and citizen science (CS). VGI does not require coordination among the individuals making these contributions. Records come from mobile devices, social media, geolocated photographs, web accommodation services, bank card transactions, and intelligent transportation cards (Gutierrez-Puebla et al., 2018). Pfoser (2016) mentions that the VGI may be differentiated between explicit and implicit content. The first choice is produced purposefully in the desired form (for instance, the OpenStreetMap's road network). In contrast, implicit content is derived from different-purpose information and is often embedded in social media contributions. CS refers to the participation of non-scientific stakeholders in the scientific process. At its most inclusive and innovative, it incorporates citizen volunteers as partners in the scientific process, including determining research subjects, questions, methods, and means of disseminating outcomes. As reasonable steps, Zook et al. (2017) suggest practising ethical distribution, maintaining the contributors' condition of anonymity, respecting the rules of conduct established by the user community, and identifying unreliable sources and recognising when there are more convenient solutions than this technological approach. These steps generate chances for the successful completion of CS initiatives and, as examples, are the efforts of Gacutan et al. (2022) and Weeser et al. (2021). First, this approach's feasibility is to monitor the spatial distribution pattern of collected solid waste (plastics) over a significant section of the approximately 34,000 km of Australia's coastline (Figure 8) between January 2009 and December 2019. In the second experience, the adequate quality of the data provided by some communities of the Sondu - Miriu River Basin (3450 km^2) in western Kenya was assessed to correct the lack of records inherent to the variation of the levels of flooding in different points of the water network.

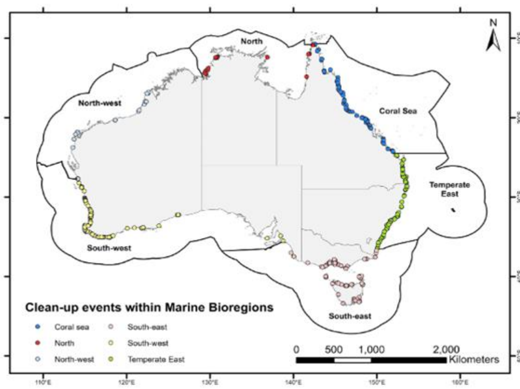

Figure 8. Clean–up events within marine bioregions – Australia. Source: Weeser & others (ob. cit)

DEI has the potential to play a vital role in measuring the ecological footprint, which is a widely accept ed method of evaluating the impact of human activities. There are also sources available to improve our knowledge of the animal kingdom, such as Movebank Solving Animal Tracking's BD problem portal (Max Planck Institute of Animal Behaviour, 2021). As of October 2021, users could access 3.2 billion locations of different species.

SENSOR SYSTEMS AND MACHINES.

SSMs are electronic devices like nanoscience technologies, in situ instruments, or remote sensing satellites. Their common attributes are the high degree of automation for survey tasks and the relatively short time that elapses from capturing to accessing data.

Nanoscience and in situ instruments generate valuable understandings from biophysical variables measuring, especially in air pollution or water quality analysis as a critical element of life quality in urban environments (Li-Minn & Kah, 2016). As a relevant situation, Ong et al. (2014) performed predictions of environmental monitoring data using open sensors that measure wind speed, wind direction, temperature, illuminance, humidity, and rain. Ying (2019) employed an online monitoring system to measure water quality parameters surveyed from sensors in Minhang District, China.

Regarding remote sensing, different agencies or corporations with relevant experience usually provide this kind of DB since these entities have a technical profile that gives credibility to their products. Remote sensing imagery is the broadest and most accessible source from SSM. An example of the former is represented by NASA or the European Space Agency (ESA), which are organisations that continuously grant free access to products with continuous and systematic coverage that meet the needs of an essential sector of the scientific community. Currently, the paradigm of non-compatibility between high temporal and spatial resolution is overtaken, and products with these advantages may be delivered at the customer's request at a monetary cost. Relevant cases include the Planet Labs Inc. and Iceeye consortiums: the former has more than one hundred and fifty satellites equipped with optical sensors, with collective capacity for daily revisits, and includes the SkySat, RapidEye, and Dove constellations. The latter stands out for having small-sized nanosatellites (mass close to 5 kg and size 10 cm x 10 cm x 10 cm) placed in low orbits of approximately 500 km. Iceeye provides high spatial resolution SAR data for revisit periods as short as a few hours (Iceeye, 2021). Two Ground Range Detected (GRD) sub-scenes (Figure 9) are a free access example (Iceeye, 2021.a): respectively acquired on April 5 and 6, 2021, cover part of the port of Rotterdam - The Netherlands, with a spatial resolution of 8.43(y) m and 5.31(x) m. The imagery reference system is WGS 84 UTM zone 32N. SAR data is effectively captured day and night and in all weather conditions. That characteristic enables the persistent monitoring of sea, land, and hard-to-reach locations.

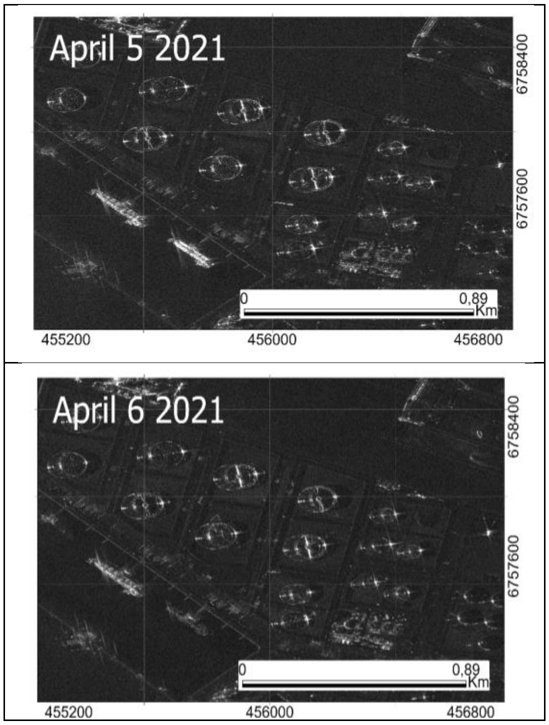

Figure 9. An example of high spatial high temporal resolution SAR imagery. Source: https://www.iceye.com/lp/example-daily-coherent-gtr-dataset-port-of-rotterdam

BIOMETRIC OPERATIONS.

Biometric operations (BO) survey an individual's physiological or behavioural characteristics with an identified spatial location. It includes gait, facial expressions, galvanic skin response, and palm or iris patterns. Crampton (2019) argues that BO assurance connects geographically distant actors and reveals new ways of value because it forms an essential part of the intelligent city movement to make cities healthier and more efficient. In addition to other proposals, it could monitor employee well-being or detect shopper frustrations at long check-out lines.

Data is mainly created and reserved for institutions responsible for citizen security, defence functions, and the fight against fraud in financial operations. Shanghai's brain (Deutsche Welle, 2021) is one relevant instance because the Chinese state gathers massive amounts of data. Regarding geolocated tools, the availability of more than two hundred and ninety thousand video surveillance cameras is emphasised for continuously monitoring strategic situations through implementing AI. As an example of digital cartography, Figure 10 shows the location of ordinary circumstances (yellow points) or the development of activities classified as irregular (red points). Concerning the last, cases are cited: a construction worker who omitted a security measure by not correctly carrying physical security implements, an individual who disposed of solid waste in the wrong place, or an improperly parked car.

Figure 10. An example of BO for detecting irregular activities in Shanghai – China. Source: Deutsche Welle (ob. cit)

BO encompasses challenging matters such as geoprivacy, which pertains to safeguarding the privacy of location-based data, mitigating the risk of location spoofing, avoiding locational surveillance, and addressing the issue of data bias associated with location-based information.

THE GEOBD CENTRAL SUPPORT TECHNOLOGIES.

Proper technologies allow the efficient processing of data. Previous sections identified some valuable resources supporting Geo BD procedures: HPC, AI, and DS. Often, because the last two share some technical foundations, it is challenging to establish concise theoretical or practical limits between them. That allows different authors under scientifically accepted criteria to refer to the discipline of their choice. Performance Computing (HPC) capabilities support these technologies.

HIGH-PERFORMANCE COMPUTING.

HPC uses supercomputers to solve various scientific, business, or governmental approaches (United States Geological Survey, 2022). This power overcomes the intrinsic barriers to the reduced execution of processes and long calculation times, making it possible to manage queries and analyse large vector or raster datasets. Now, achieving in minutes what would have previously taken days is possible. The technology lies in the simultaneous use of more than one central processing unit (CPU) grouped into computing nodes (composed of a processor or a group of processors) and a memory block (IBM, 2022. b). That advantage makes it easier for organisations to focus on developing, implementing, and disseminating resources rather than merely solving informatics issues. Depending on the infrastructure, HPC could perform in cloud computing (CC), a proprietary infrastructure (in-situ), or a hybrid model. The CC enables access to virtual computing, network, storage, and processing resources through a remote provider managing data storage and servers. In the second option, the organisation must afford the monetary costs of acquiring and updating the physical and logical equipment. The hybrid solution satisfies the requirements of organisations that own infrastructure and requires some of the benefits of HPC. This research focuses on CC instances because it could be one of the most accessible alternatives for the scientific community or organisations that do not have a supercomputer infrastructure.

ARTIFICIAL INTELLIGENCE.

Although the term AI is recurrent in the BD era, it is worth noting that its applicability in the digital geography context started at least in the early 1980s. As arguments, Couclelis (1986) demonstrated that AI and geography are brought together within a broad context involving critical issues of theory, epistemology, and the scientific method. Estes et al. (1986) evaluated classification techniques to automate satellite image processing. Campbell and Roelofs (1984) report the scientifically documented first attempts to automate imagery analysis and their applications in remote sensing.

In general, AI is the ability of machines to execute algorithms, learn from data, and use what they learn to make decisions similarly to a human. As a result, the proportion of errors is significantly minimised compared to those executed from a human capacity. For BD processing, Campesato (2020) suggests that AI encompasses the following sub-areas: Machine Learning (ML), Deep Learning (DL), Natural Language Processing (NLP), Reinforcement Learning (RL), and Deep Reinforcement Learning (DRL). The first three are relevant to this study because they strongly connect to GeoBD processing, while the analysis of the rest is beyond this research context. For the GeoBD, Suthaharan (2016) states that ML involves the execution and development of mathematical models and algorithms that learn from large volumes of data to classify them under a supervised or unsupervised scheme. First, the classes are defined through training data that will be the reference for the algorithms. For an unsupervised approach, various statistical criteria fix limits between classes. Based on computational models of interconnected artificial neural networks (CNNs), DL Learning generates valuable information, especially for automated predictive analytics. Instead of executing predefined algorithms (Lecun et al., 2015), the operation performs pattern recognition through multiple layers (input, hidden, and output) that configure the basic parameters and train the system to learn by itself. Input layers evoke neurons that take in input data (for example, an image or data table).

NLP focuses on the interaction between computers and human language to implement activities relevant to the translation between languages, extract text information, summarise documents, and detect "inappropriate" terms for specific target audiences.

Contextualising AI into geography, contributions from Janowicz et al. (2022) and Zurita (2023) refer to geographic artificial intelligence (GeoAI) as the integration of geospatial studies and AI. Belonging to NLP, geoparsing focuses on identifying the spatial location of certain situations through texts alluding to place names or spatial entities from sources such as travel blogs, real estate advertisements, and social networks (Cadorel et al., 2021). However, instead of geometries, the challenge is to process texts that may have ambiguities like the use of different languages, the same language variations, the existence of several places with the same name, colloquial vocabulary, misspellings, or multiple possible interpretations of location names. In this way, Gao (2021) affirms that there has already been an increasingly collaborative GeoAI, which refers to using AI in human and social geography and environmental remote sensing. Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP) have made significant contributions to this area.

For instance, Lavalin and Downs (2020) found that ML has many applications in geolocated data. It is used to study the production and distribution of goods, population genetics, human disease patterns, tourism forecasting, detecting military patterns, international politics, agricultural production, transportation, and measuring the human sense of place.
From a satellite imagery digital analysis, ML has helped advance research on diversity, biogeography, and abundance of environmental features, management of water resources, climate modelling, tackling climate change, predicting landslide, hurricane wind risk models, melting ice sheet effects, marine species geographic predictions, and social-ecological data.

DL also has significant potential for digital GeoAI studies (Grekousis, 2019), particularly in urban geography (urbanisation, environment and socioeconomic), LUC changes in non-urban contexts, and remote sensing performing image analysis through CNNs. It is very effective in practice but requires extensive data and computing power. About NLP, Molina-Villegas et al. (2021) developed the first geoparser system for entity recognition and disambiguation of documents written in Mexican Spanish. Berrang-Ford et al. (2021) propose a solution for systematically identifying and mapping the climate change and scientific health literature produced in English between January 1, 2013, and April 9, 2020.

DATA SCIENCE.

Data Science (DS) is a field of computation that relies on mathematics and statistics to extract meaningful insights from Big Data (BD). It follows a five-stage procedure that includes framing the problem, collecting, cleaning data, analysing it, and visualising it. DS is a multidisciplinary approach that is useful in both business and science. It empowers organisations with the data they need to optimise their processes for maximum efficiency and revenue generation. In science, it helps to develop protocols to achieve specific goals and get good results.

Artificial intelligence (AI) and Big Data (BD) are crucial components of the DS process. Thus, much of the state-of-the-art theory and research focuses on these technologies. For example, Menoyo (2021) and Pierson (2017) explain statistics and AI in detail in their DS proposals. Erman et al. (2022) highlight the importance of AI in producing high-value, high-quality, relevant, and trusted products that reflect the evolving needs of Canada's society and economy. Spatial Data Science (SDS) is a subset of DS that focuses on the characteristics of spatial data and extracting deeper insights using a comprehensive set of analytical methods and spatial algorithms (ESRI, 2022). Singleton and Arribas-Bel (2021) explain the relevance of Geographic Data Science (GDS), which combines Geography's long-standing traditions and epistemologies with recent advances in DS. While these fields are relatively new, they hold great promise for the future.

KEY FINDINGS

The revision shows that GeoBD helps integrate and understand intricate relationships between biophysical and human dimensions of spatial decision phenomena such as risk management, health, agriculture, environment, and open government. Free access policy tools are essential to utilising Geo BD technologies.

Adopting GeoBD technology can bring economic benefits to businesses and industries. Major global corporations such as Amazon, Google, and Microsoft actively provide geospatial solutions, while companies like UBER and Airbnb are developing their geospatial tools to optimise their operations. This highlights the growing importance of geospatial technology in driving business efficiency and competitiveness.

GeoBD practices typically require dense datasets that meet specific spatial and temporal requirements. These datasets are processed using AI algorithms, focusing on Data Mining (DM) or Data Science (DS). Cloud computing (CC) enables faster and more effective processing, emphasising the importance of scalable infrastructure for handling large geospatial datasets.

GeoBD emphasises the value derived from sources that meet conditions of veracity, not just the absolute data storage size. This highlights the importance of data quality and reliability in geospatial analysis and decision-making processes.

Emerging sources like DEI complement traditional sources like population census to understand socioeconomic phenomena. Novel approaches like sentiment analysis offer new insights into the human dimension, which may be challenging to evaluate through traditional methods alone. Citizen science is gaining importance, although concerns about its reliability remain.

While BO offers tremendous potential to address societal challenges and improve decision-making, its widespread use also raises significant social implications related to geoprivacy, location spoofing, locational surveillance, and data bias. It is essential to approach this source's collection, analysis, and application carefully, considering ethical, legal, and social implications to ensure that it benefits society.

Advanced visualisation techniques are critical in promoting result-sharing through scientific research or social media. While dynamic cartography and dashboards are commonly used for visualisation, traditional maps remain valuable for effectively representing complex spatial and temporal phenomena.

CONCLUSIONS

In conclusion, GeoBD significantly advances the integration and understanding of the complex relationships between biophysical and human dimensions.

The technology offers valuable insights and opportunities across various sectors, including risk management, health, agriculture, environment, and open government.

Effective use of GeoBD relies on dense datasets meeting specific spatial and temporal requirements, processed using AI algorithms, and supported by proper infrastructure. Moreover, GeoBD emphasises the importance of data quality and reliability in deriving meaningful insights.

The emergence of new data sources and approaches and advanced visualisation techniques further enhance our understanding and use of GeoBD.

While this study provides essential insights into GeoBD, further research, such as comprehensive reviews, is needed to deepen our understanding and contribute to specific domains. Continued exploration and innovation in GeoBD will be crucial in addressing complex spatial phenomena.

REFERENCES

Aguirre, M. G., Guerrero, N. S. T., & Guerrero, L. M. T. (2021). The influence of Airbnb on hotel occupancy in Mexico: a Big Data Analysis (2007-2018). Cimexus, 16(3), 9-32.

Airbnb. (2021). Inside Airbnb. http://insideairbnb.com/get-the-data.html. Accessed September 15, 2021.

Allen, C., Smith, M., Rabiee, M., & Dahmm, H. (2021). A review of scientific advancements in datasets derived from big data for monitoring the Sustainable Development Goals. Sustain Sci 16 (5), 1701-1716. DOI: 10.1007/s11625-021-00982-3.

Ardagna, D., Cappiello, C., Samá, W., & Vitali, M. (2018). Context-aware data quality assessment for big data. Future Generation Computer Systems, 89, 548-562.

BBVA. (2023). Monitoring The Economy in Real Time And High Definition, https://www.bbvaresearch.com/en/special-section/bigdata/ Accessed March 30, 2023.

Berrang-Ford, L., Sietsma, A., Callaghan, M., Minx, J., Scheelbeek, P., Haddaway, N., Haines, A., & Dangour, A. (2021). Systematic mapping of global research on climate and health: a machine learning review. The Lancet Planetary Health, 5 (8), 514–525. https://doi.org/10.1016/S2542-5196(21)00179-0.

Bhimala, K., Rakesh, V., & Prasad, K. (2020). Identification of vegetation responses to soil moisture, rainfall, and LULC over different meteorological subdivisions in India using remote sensing data. Theoretical and Applied Climatology, 142, 987–1001. https://doi.org/10.1007/s00704-020-03360-8.

Biao, L., Wang, J., Zhaorui., Jing, Z., Niu, H., & Yang, M. (2022) Agglomeration and migration of population in resource-based cities: evidence from DMSP/OLS and NPP/VIIRS nighttime lights, International Journal of Remote Sensing, DOI: 10.1080/01431161.2022.2152759.

Bloom, M. (2019). Environmental Epidemiology. Encyclopedia of Environmental Health (Second Edition), Elsevier, 419–427, https://doi.org/10.1016/B978-0-12-409548-9.10635-9.

Brovelli, M., Boccardo, P., Bordogna, G., Pepe, A., Crespi, M., Munafò, M., & Pirotti, F. (2019). Urban Geo Big Data. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XLII-4/W14, 23–30. https://doi.org/10.5194/isprs-archives-XLII-4-W14-23-2019, 2019.

Brown, C., Brumby, S., & Guzder-Williams, B. (2022). Dynamic World, Near real-time global 10 m land use land cover mapping. Sci Data, 9, Article 251. https://doi.org/10.1038/s41597-022-01307-4.

Cadorel, L., Blanchi, A., & Tettamanzi, A. (2021). Geospatial Knowledge in Housing Advertisements: Capturing and Extracting Spatial Information from Text. Proceedings of the 11th Knowledge Capture Conference. Association for Computing Machinery, New York. https://doi.org/10.1145/3460210.3493547.

Campbell, W., & Roelofs, L. (1984). Artificial Intelligence applications concepts for the remote sensing and earth science community. Proceedings IEEE 1984 IX Pecora Symposium on Spatial Information Technologies for Remote Sensing Today and Tomorrow, 232–242.

Campesato, O. (2020). Artificial Intelligence, Machine Learning, and Deep Learning. Mercury Learning & Information.

Cheng, X., Fang, L., Yang, L., & Cui, S. (2018): Mobile Big Data. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-96116-3.

Couclelis, H. (1986). Artificial Intelligence in Geography: conjectures on the shape of things to come. The professional geographer, 38(1), 1-11, doi: 10.1111/j.0033-0124.1986.00001.x.

Crampton, J. (2019). Platform biometrics. Surveillance & Society, 17(1), 54–62.

DECKGL. (2023). United Kingdom road safety. https://deck.gl/examples/hexagon-layer. Accessed October 30, 2022.

Deutsche Welle. (2021). China - Surveillance State or Way of the Future? https://www.youtube.com/watch?v=7gSU_Xes3GQ Accessed October 17, 2021.

ESRI. (2021). What is ArcGIS Velocity? https://doc.arcgis.com/es/iot/get-started/what-is-arcgis-velocity.htm Accessed November 18, 2021.

ESRI. (2022). The rise of spatial data science. https://www.esri.com/about/newsroom/podcast/the-rise-of-spatial-data-science/ Accessed October 30, 2022.

Erman, S., Rancourt, E., Beaucage, Y., & Loranger, A. (2022). The Use of Data Science in a National Statistical Office. Harvard Data Science Review, 4(4). https://doi.org/10.1162/99608f92.13e1d60e.

Estes, J., Sailer, C.,& Tinney, L. (1986). Applications of artificial intelligence techniques to remote sensing. The Professional Geographer 38 (2), 133–141. https://doi.org/10.1111/j.0033-0124.1986.00133.x.

Gacutan, J., Johnston, E., Tait, H., Smith, W., & Clark, G. (2022). Continental patterns in marine debris revealed by a decade of citizen science. Science of the total environment 807, Article 150742.https://doi.org/10.1016/j.scitotenv.2021.150742.

Gao, S. (2021). Geospatial Artificial Intelligence (GeoAI). Geography: Oxford

University Press. DOI: 10.1093/OBO/9780199874002-0228.

Google Earth Engine. Reducer overview. (2023). [en línea]. Disponible en: https://developers.google.com/earth-engine/guides/reducers_intro [Accedido el 20 de marzo de 2023].

Grekousis, G. (2019). Artificial neural networks and Deep Learning in urban geography: a systematic review and meta-analysis. Computers, Environment and Urban Systems, 74, 244–256. https://doi.org/10.1016/j.compenvurbsys.2018.10.008.

Goffi, A., Bordogna, G., Stroppiana, D., Boschetti, M., & Brivio, P. (2020). A scalable synthesis of multiple models of Geo Big Data interpretation. Journal of Software Engineering and Applications, 13 (06), 104–128. DOI: 10.4236/jsea.2020.136008.

Gutiérrez-Puebla, J.; García-Palomares, J.C.; Salas-Olmedo, M.H. (2016). Big (Geo) Data en Ciencias Sociales: Retos y Oportunidades. Revista de Estudios Andaluces, 33 (1), 1-23. http://dx.doi.org/10.12795/rea.2016.i33.01.

Gutiérrez – Puebla, J. (2018). Big Data y nuevas geografías: la huella digital de las actividades humanas. Documents d'Anàlisi Geogràfica, 64(2), 195-217. https://doi.org/10.5565/rev/dag.526.

Huang, B., & Wang, J (2020). Big spatial data for urban and environmental sustainability. Geospatial Information Science, 23 (2), 125–140. https://doi.org/10.1080/10095020.2020.1754138

IBM. (2022. a). What is data lineage? https://www.ibm.com/topics/data-lineage. Accessed November 1, 2022.

IBM. (2022. b). What is high–performance computing (HPC)? https://www.ibm.com/topics/hpc. Accessed October 30, 2022.

Iceeye. (2021). Iceeye SAR Data. https://www.iceye.com/sar-data Accessed December 25, 2021.

Iceyeye (2021.a) https://www.iceye.com/lp/example-daily-coherent-gtr-dataset-port-of-rotterdam. Accessed December 25, 2021.

Janowicz, K., Sieber, R., & Crampton, J. (2022). GeoAI, counter-AI, and human geography: A conversation. Dialogues in Human Geography, 12(3), 446–458. https://doi.org/10.1177/20438206221132510.

John Hopkins University. (2021). Coronavirus Resource Centre: COVID-19 Dashboard by the Center for Systems Science and Engineering. https://coronavirus.jhu.edu/map.html Accessed December 11, 2021.

Kitchin, R., & McArdle, G. (2016.a). What makes Big Data Big Data? Exploring the ontological characteristics of 26 datasets. Big Data & Society, 3(1). https://doi.org/10.1177/2053951716631130.

Kitchin, R., & McArdle, G. (2016.b). Urban data and city dashboards: Six key issues. 10.31235/osf.io/k2epn.

Lavalin, A., & Downs, A. (2021). Machine learning in geography–Past, present, and future. Geography Compass, 15(5), doi:10.1111/gec3.12563.

Lecun, Y., Bengio, Y. & Hinton, G. (2015). Deep Learning. Nature, 521, 436–444. https://doi.org/10.1038/nature14539.

Li, J., & Biljecki, F. (2019). The implementation of big data analysis in regulating online short-term rental business: a case of Airbnb in Beijing. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 4th International Conference on Smart Data and Smart Cities. IV-4/W9, 79—86. https://doi.org/10.5194/isprs-annals-IV-4-W9-79-2019.

Li-Minn, A., & Kha, S. (2016). Big Sensor Data Applications in Urban Environments, Big Data Research, 4, 1-12. https://doi.org/10.1016/j.bdr.2015.12.003.

Max Planck Institute of Animal Behaviour. (2021). Movebank. https://www.movebank.org/cms/movebank-main Accessed January 20, 2021.

Menoyo, D. (2021). Fundamentos de la ciencia de datos. Editorial Universidad de Alcalá.

Microsoft. (2022). https://planetarycomputer.microsoft.com/ Accessed January 15, 2023.

Molina-Villegas, A., Muñiz-Sanchez, V., Arreola-Trapala, J., & Alcántara, F. (2021). Geographic named entity recognition and disambiguation in Mexican news using word embeddings. Expert Systems with Applications, 176, Article 114855. https://doi.org/10.1016/j.eswa.2021.114855.
MongoDB. (2022). Unstructured data. https://www.mongodb.com/unstructured-data. Accessed September 15, 2022.

Ong, B., Sugiura, K., & Zettsu, K. (2014). Dynamic pre-training of Deep Recurrent Neural Networks for predicting environmental monitoring data. 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 2014, 760-765. doi: 10.1109/BigData.2014.7004302.

Open Election Data Initiative. (2022). Election data guide. https://openelectiondata.net/en/guide/introduction/ Accessed December 15, 2022.

Ortiz, A., & Rodrigo, T. (2020). High-Frequency data at BBVA Research. https://www.ecb.europa.eu/pub/conferences/shared/pdf/20201016_tracking_the_economy/Alvaro_Ortiz_and_Tomasa_Rodrigo.pdf Accessed March 20, 2023.

Pfoser, D. (2016). Crowdsourcing Geographic Information Systems. In: Liu, L., Özsu, M. (eds) Encyclopaedia of Database Systems. Springer, New York, NY. https://doi.org/10.1007/978-1-4899-7993-3_80607-1.

Pierson, L. (2017). Data Science for Dummies, John Wiley & Sons.

Ramasamy, A., & Chowdhury, S. (2020). Big Data Quality Dimensions: A Systematic Literature Review. Journal of Information Systems and Technology Management. 17. https://doi.org/10.4301/S1807-1775202017003.

Represa, N. (2020). Elaboración e implementación de una propuesta metodológica para la evaluación y gestión de la calidad del aire mediante el enfoque de la ciencia de datos. Doctoral thesis Universidad Nacional de La Plata – Argentina & Universidad Politécnica de Valencia - Spain. https://doi.org/10.35537/10915/94380.

Robinson, A., Demšar, U., Moore, A., Buckley, A., Jiang, B., Field, K., Kraak, M., Camboin, S., & Sluter, C. (2017). Geospatial big data and cartography: research challenges and opportunities for making maps that matter. International Journal of Cartography, 3, 32–60. https://doi.org/10.1080/23729333.2016.1278151.

Rosenblum, A., Wend, C., Akhtar, Z., Rosman, L., Freeman, J., Barnett, D. (2021). Use of Big Data in Disaster Recovery: An Integrative Literature Review. Disaster medicine and public health preparedness, 17, e68. https://doi.org/10.1017/dmp.2021.332.

Rowe, F. (2021). Big Data and Human Geography. In: Demeritt, D. and Lees L. (eds) Concise Encyclopaedia of Human Geography. Edward Elgar Encyclopaedias in the Social Sciences series.

Sadiq, R., Akhtar, Z., Imran, M., & Ofli, F. (2022). Integrating remote sensing and social sensing for flood mapping, Remote Sensing Applications: Society and Environment, 25, Article 100697.https://doi.org/10.1016/j.rsase.2022.100697.

Statista. (2021. a). The volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2024. https://www.statista.com/statistics/871513/worldwide-data-created/ Accessed November 18, 2021.

Statista. (2021. b). The global digital population as of January 2021. https://www.statista.com/statistics/617136/digital-population-worldwide/ Accessed November 18, 2021.

Simonetti, D., Pimple, U., Langner, A., & Marelli, A. (2021). Pan-tropical Sentinel-2 cloud-free annual composite datasets, Data in

Brief, 39, Article 107488. https://doi.org/10.1016/j.dib.2021.107488.
Singleton, A., & Arribas-Bel, D. (2021). Geographic data science. Geographical Analysis, 53(1), 61-75. https://doi.org/10.1111/gean.12194.

Suthaharan, S. (2016). Machine Learning Models and Algorithms for Big Data Classification. Springer.

Taylor & Francis. (2021). Big Earth Data: an open access journal. https://www.tandfonline.com/action/journalInformation?show=aimsScope&journalCode=tbed20 . Accesed May 29, 2021.

Thatcher, J., Eckert, J., & Shears, A. (2018). Thinking Big Data in Geography: New Regimes, New Research. The University of Nebraska Press. https://doi.org/10.2307/j.ctt21h4z6m.

The United Nations (2021). Big Data for Sustainable Development. www.un.org/en/global-issues/big-data-for-sustainable-development Accessed January 20, 2022.

The United Nations Development Programme China. (2016). The living standards dimension of the human development index. https://www.undp.org/sites/g/files/zskgke326/files/migration/cn/UNDP-CH-PR.-Publications-Measuring-Poverty-with-Big-Data-in-China-reduced.pdf. Accessed March 15, 2023.

The United States Census Bureau. (2022). Big Data. https://www.undp.org/sites/g/files/zskgke326/files/migration/cn/UNDP-CH-PR.-Publications-Measuring-Poverty-with-Big-Data-in-China-reduced.pdf. Accessed March 15, 2023.

The World Bank. (2023.a). Connecting for Inclusion: Broadband Access for All. https://www.worldbank.org/en/topic/digitaldevelopment/brief/connecting-for-inclusion-broadband-access-for-all. Accessed March 30, 2023

The World Bank. (2023.b). Individuals using the Internet. https://data.worldbank.org/share/widget?indicators=IT.NET.USER.ZS. Accessed January 7, 2023.

Tobias, A., Hashizume, M., Honda, Y., Sera, F., Ng, C. F. S., Kim, Y., Roye, D., Chung, Y., Dang, T, Kim, H., Lee, W., Íñiguez, C., Vicedo-Cabrera, A., … Gasparrini, A. (2021). Geographical Variations of the Minimum Mortality Temperature at a Global Scale: A Multicountry Study. Environmental epidemiology (Philadelphia, Pa.), 5(5), e169. https://doi.org/10.1097/EE9.0000000000000169.

Tran, A., Kassié, D., & Herbreteau, V. (2016). Applications of Remote Sensing to the Epidemiology of Infectious Diseases: Some Examples. Land Surface Remote Sensing,295 – 315, https://doi.org/10.1016/B978-1-78548-105-5.50009-8.

UBER (2021). Uber Newsroom. https://www.uber.com/es-CO/newsroom/company-info/ Accessed December 10, 2021.

United States Geological Survey. (2022). What is high–performance computing? https://www.usgs.gov/advanced-research-computing/what-high-performance-computing Accessed November 10, 2022.

Vu, T., Vu, N., Phung, P., & Nguyen, L. (2021). Enhanced urban functional land use map with free and open-source data. International Journal of Digital Earth 14 (11), 1744–1757. https://doi.org/10.1080/17538947.2021.1970262.

Weeser, B., Gräf, J., Njue, N., Cerutti, P., Rufino, M., Breuer, L., & Jacobs, S. (2021). Crowdsourced water level monitoring in Kenya's Sondu-Miriu Basin—Who Is "The Crowd"? Frontiers in earth science (8). https://doi.org/10.3389/feart.2020.602422.

Ying, L., Shihu, S., Hongyu, W., Xin, Z., & Qi, Y. (2019). Big Data Analysis of Water Quality of Secondary Water Supply. Procedia Computer Science, 154, 744-749. 10.1016/j.procs.2019.06.116.

Yong, Z., Li, K., Xiong, J., Cheng, W., Wang, Z., Sun, H., & Ye, C. (2022). Integrating DMSP-OLS and NPP-VIIRS Nighttime Light Data to Evaluate Poverty in Southwestern China. Remote Sensing, 14(3), 600. http://dx.doi.org/10.3390/rs14030600.

Young, G., Kitchin, R., & Naji, J. (2021). Building City Dashboards for Different Types of Users, Journal of Urban Technology, 28:1-2, 289–309, DOI: 10.1080/10630732.2020.1759994.

Zhao, Q., Yu, L., Li, X., Peng, D., Zhang, Y., & Gong, P. (2021). Progress and Trends in the Application of Google Earth and Google Earth Engine. Remote Sensing, 13(18), 3778. MDPI AG. http://dx.doi.org/10.3390/rs13183778.

Zhu, Y. (2019). Geospatial semantics, ontology and knowledge graphs for big Earth data, Big Earth Data, 3 (3), 187–190, DOI: 10.1080/20964471.2019.1652003.

Zola, P., Ragno, C., & Cortez, C. (2020). A Google Trends spatial clustering approach for a worldwide Twitter user geolocation. Information Processing & Management, 57 (6), Article 102312. https://doi.org/10.1016/j.ipm.2020.102312.

Zook, M., Barocas, S., Boyd, D. Crawford, K., Keller, E., Gangadharan, S., Goodman, A., Hollander, R., Koenig, B., Metcalf, J., Narayanan, A., Nelson, A. & Pasquale, F. (2017). Ten simple rules for responsible big data research. PLOS Computational Biology, 13(3), Article 1005399. doi:10.1371/journal.pcbi.1005399.

Zurita, R. (2023). GeoArtificial intelligence https://www.utwente.nl/en/digital-society/research/data/research-area/geoai/#contact Accessed March 30, 2023.