INTERNET OF THINGS DATA VISUALIZATION FOR BUSINESS INTELLIGENCE

Cite as: Sima Attar-Khorasani and Ricardo Chalmeta (2022). INTERNET OF THINGS DATA VISUALIZATION FOR BUSINESS INTELLIGENCE. Big Data. Published Online: 8 Feb 2022https://doi.org/10.1089/big.2021.0200

Sima Attar-Khorasani, Ricardo Chalmeta
Grupo de Investigación en Integración y Re-Ingeniería de Sistemas (IRIS)
Departamento de Lenguajes y Sistemas Informáticos
Universitat Jaume I, 12006 Castellón, Spain.
Tel: + 34 964 728329 Fax: + 34 964 728435
al409965@uji.es, rchalmet@uji.es

Corresponding author: Ricardo Chalmeta http://orcid.org/0000-0001-5490-7802

Abstract

This study contributes to the research on Internet of Things data visualization for business intelligence processes, an area of growing interest to scholars, by conducting a systematic review of the literature. A total of 237 articles published over the past 11 years were obtained and compared. This made it possible to identify the top contributing and most influential authors, countries, publishers, institutions, papers and research findings, together with the challenges facing current research.
Based on these results, this work provides a thorough insight into the field by proposing four research categories (Technology infrastructure, Case examples, Final-user experience, and Big Data tools), together with the development of these research streams over time and their future research directions.

Keywords: Data Visualization, Business Intelligence, Big Data, Internet of Things, Industry 4.0

1. INTRODUCTION

Business Intelligence (BI) is a kind of information system for gathering, manipulating, storing and analysing raw data and transforming it into useful information for managers enabling them to make better and faster decisions and discover new business prospects.[1,2] BI systems complement corporate operational information systems such as Enterprise Resource Planning (ERP), Supply Chain Management (SCM), Customer Relationship Management (CRM), etc. and offer organizations a great potential to improve organizational efficiency. [3]

A BI system is composed of three components[4]: a technological component, which includes a broad range of analytical software for diverse organizational provisions; [5] a human component, consisting of system developers and managers with a high level of analytical skills; and the business process, which will underlie the transformation of information into knowledge. The evolution of these three components has allowed three generations of BI systems to be established.

The first generation of BI appeared in the 1990s. It was characterized by highly formatted reports developed by IT personnel using proprietary BI tools embedded into companies' desktop or client/server applications, and communicated by proprietary application programming interfaces (APIs). The second generation of BI started in the 2000s. Using data warehouses and dashboard-building tools, analysts and business users were able to access large amounts of structured and unstructured data and could create intuitive drag-and-drop reports by themselves.[6] Data were usually extracted from corporate operational computer systems and then transformed and loaded into an enterprise data warehouse and centralized metadata repository, using extraction transformation migration and loading (ETML) tools, such as Apache Spark. The next step was to cluster them in Data Marts. On-line analytical processing (OLAP) tools, such as OBIEE (Oracle Business Intelligence Enterprise Edition) and IBM Cognos were used to calculate and display indicators in user interfaces such as dashboards, spreadsheets, etc. [7] Today, the third generation of BI is characterized by (1) organizations' awareness of the possibilities of BI to create a competitive advantage and therefore of the need to develop a Data-Driven Business Strategy where data are used to create more value, to keep costs down, to drive additional sales, to engage customers more fully, and to improve process efficiency; and (2) the exponentially increasingly multi-structured data sets that companies have to analyse in an efficient manner, Internet of Things (IoT) sensors and web 2.0 tools being the major contributors. [8]

An integral part of today's generation of BI systems is IoT data visualization for business analytics processes. IoT provides massive amounts of data emitted from multiple connected sensors that, once gathered, analysed and displayed, can be used by managers to support their decision-making. Data visualization is essential in this process, since it allows managers to gain a proper understanding of the underlying patterns and results obtained by analysis algorithms. [9] Data Visualization in BI must consider the business objectives and the evolving needs of users, taking into account high-level semantics, reasoning about unstructured and structured data, and providing simplified access and a better understanding of data. [10]

However, although IoT data visualization enhances the productivity and efficiency of the business, different authors have shown that there are numerous challenges and factors that affect the use of IoT data visualization for business intelligence processes that must be identified and studied. [11] These include volume and variety of data, heterogeneity of devices, necessity of scalable and efficient storage infrastructures, data security and privacy, reliable validity of marketing segmentation, inexperienced users who do not know what type of information they want to extract from data or which would be the best type of visualization or who are wrong in their interpretation, [12] and financial management issues.

To date, no literature review has examined the IoT data visualization for BI processes. To address this gap, this paper synthesizes the body of knowledge on IoT data visualization for BI processes and establishes research categories that bring together research conducted on the basis of relevant common points. In particular, in this study the following research questions (RQs) are posed:

RQ1: Which are the top contributing authors, countries, papers, institutions and publishers in the field of IoT data visualization for BI processes?

RQ2: Is it possible to classify research papers on the basis of relevant common points?

RQ3: What are the future research necessities in the field of IoT data visualization for BI processes?

To answer the above research questions, this paper (1) carries out a systematic review of the literature on IoT data visualization for BI processes, since it is an efficient research method that allows a precise evaluation of the information published to date; [13] (2) provides a thorough insight into the field by using bibliometric analysis techniques to evaluate 237 published articles, and to identify top contributing authors, papers, countries, publishers and institutions related to the field; (3) identifies and proposes four established and emerging research categories that would encourage scholars to expand research on IoT data visualization for BI processes; and (4) identifies the future research necessities in each research category.

This paper is organized as follows: Section 2 describes the software tools and the research methodology used to perform the bibliographical analyses. Section 3 offers the findings of the bibliographical analyses. Finally, section 4 discusses the findings as well as the conclusions, research limitations and future work.

2. RESEARCH METHODOLOGY

To answer the above-mentioned research questions, a semi-structured literature review was selected because it makes it possible to understand the state of knowledge, to identify the historical evolution and to develop a research agenda. [14]

Among the various text mining and grouping methods to extract non-trivial patterns and knowledge from textual and unstructured documents written in natural language, probabilistic topic modeling can be highlighted. [15,16] A topic model is normally defined as an approach for discovering the latent information in a corpus. This method takes in account two essential aspects: (1) words can have multiple meanings; and (2) interpretations and documents may contain one or more topics. [17] One of its main applications is the classification of documents, which has received significant attention. In general, document classification problems are solved following this sequence: (1) coding each word or document with a numerical vector, and then (2) classifying documents. [18]

In this study, to provide insights into the main topics, the machine learning algorithm Latent Dirichlet Allocation (LDA) was applied instead of other topic modelling such as Latent Semantic Analysis (LSA) or Probabilistic Latent Semantic Analysis (PLSA). LDA, an unsupervised topic modelling approach, was selected because among the different methods it is the simplest and most popular, [19,20,21] it can be applied to different kinds of problems, [22] it allows to obtain good performance, [23] it can be used to quickly identify thematic clusters in large documents, [24] it can significantly reduce the workload involved in the screening phase of a systematic review, [25] and it is not necessary to have any knowledge in advance about the topics. [26]

The LDA was carried out in Python using several libraries. The core libraries include data science libraries like Pandas and NumPy for handling the data, and Gensim library, written by Radim Rehurek, an open-source NLP library intended for unsupervised topic modelling. [27]

LDA works by assuming that each document is a probability distribution of topics and each topic is a probability distribution of words from the document. The idea is that documents are "*represented as random mixtures over latent topics, where each topic is characterized by a distribution over words*". [15] LDA is based on three concepts: the corpus (the text collection), the document (one item within the corpus), and the terms (the words within a document). Then, the aim of the LDA algorithm is to infer topics from recurring patterns of word occurrence in the documents. Topics are heuristically located on an intermediate level between the corpus and the documents and can be imagined as content-related categories, or clusters. [28] Five steps were carried out to apply LDA (figure 1):

2.1. ***Building and pre-processing the corpus of documents***. The first step was to select the data sources, which in this case were Scopus and Web of Science. Although there are other popular interdisciplinary databases such as Google Scholar, the low quality data found in Google Scholar raises questions about its suitability for research evaluation. [29] Scopus and Web of Science are the main sources of bibliographic citations used for bibliometric analyses. This is mainly because (1) they are the only ones that combine both a rigorous selection process and wide interdisciplinary coverage, which make them significantly stronger than the other databases; [30] (2) they are the two world-leading and competing citation databases. Although, during the last decade, there was a significant growth of available bibliographic data sources and metrics, Web of Science (WoS) and Scopus databases (DBs) still remain the two major and most comprehensive sources of publication metadata and impact indicators. Therefore, they serve as the major tools for a variety of tasks: from journal and literature selection or personal career tracking to large-scale bibliometric analyses and research evaluation practices in all possible levels; [31] and (3) both DBs are constantly being improved due to the intense competition and notable transfer of academic activities into digital internet-based environment. [31]

The selection of articles and reviews was carried out by identifying those that had certain keywords in the title, in the keywords section or in the abstract (Table 1). The keywords used were organized in three sets: IoT, Data visualization and BI, with all the possible synonyms and terms related to them. The search was conducted in October 2020 and the results were limited to papers in English published in journals and at conferences as of 2009. The fields of research were also limited to the following areas: Business, Management and Accounting; Computer Science; Economics, Econometrics and Finance; and Engineering.

Table 1. Search strategy

Source	Keywords	Content	Period	Document	Category

Scopus	("IoT" OR "Internet-of-Things" OR "Internet of Things" OR "Internet of Everything" OR "Industrial Internet" OR "smart product" OR "smart object" OR "IIoT" OR "Industrial Internet of Things" OR "Industrial IoT" OR "I4.0" OR "I 4.0" OR "Industry 4.0" OR "Fourth Industrial Revolution") AND ("data analysis and visualization" OR "Data visualization" OR "information visualization" OR "Visualization Technique" OR "visual exploration of patterns" OR "graphic representation" OR "visual depiction of data" OR "graphic portrayal") AND ("BI" OR "Business Intelligence" OR "Business Analytics" OR "competitive intelligence" OR "BI&A" OR "enterprise strategy" OR "company strategy" OR "firm strategy" OR "managers" OR "decision making" OR "performance systems" OR "balanced scorecard" OR "enterprise knowledge" OR "company knowledge" OR "firm knowledge")	TITLE ABS KEY	2009–20	Articles Reviews	Business, Management and Accounting; Computer Science; Economics, Econometrics and Finance; Engineering areas.
Web of Science					Engineering; Computer Science; Business Economics

2.2. ***Data cleaning and pre-processing of unstructured text data.*** To generate significant topic modelling, data in the corpus was cleaned. Different methods were used to do so: [24] removing punctuation and tokenization (segmenting documents into case words); making bigrams and trigrams using the n-gram language model (two/three words frequently occurring together in the text, e.g. “decision” and “making” become “decision-making); stopping (removing punctuation and the so-called stop words, e.g. conjunctions and any meaningless words); and lemmatizing (converting word to its meaningful base form, considering the context, e.g., “assessing” and “assessment” become “assess”).

2.3. ***Model selection***. Once the corpus had been cleaned, the LDA algorithm developed by Blei and Hoffman [32] was implemented for topic modelling. This algorithm is based on on-line stochastic optimization, and learns the Dirichlet hyperparameter α document-topic density directly from the data. The Dirichlet hyperparameter β word-topic density was fixed at the value of 1/K as proposed by the topic model library that was used: gensim. [33] The number of topics K was set taking into consideration the topic coherence metric.

2.4. ***Topic validity and labelling***. To ensure the validity and interpretability of the results, a content analysis was carried out independently by the authors with the objective of checking whether the topics depict relevant research issues and if a topic should be discarded or whether two or more topics should be merged. Once the topics had been validated, a label was assigned to each of them.

2.5. ***Presentation of findings***. After the identification, validation and labelling of topics, the analysis tools available at Web of Science and Scopus were used to perform a biographical analysis using descriptive statistics. Finally, a content analysis of the papers belonging to each of the topics was carried out to identify the main research conducted, the main conclusions, and the future research directions in all the research categories.

3. FINDINGS

3.1 Initial Results

The initial search was conducted in October 2020 and resulted in 230 journal papers and 194 conference papers. The initial set of papers were filtered further by reading the titles and abstracts to check coherence with the research questions and eliminate duplications. This filtering led to the final set of 126 journal papers and 111 conference papers, giving a total of 237 papers.

Figure 2 shows the evolution in the number of publications. It has increased considerably in the last six years, and a boom can be highlighted as of 2019 and 2020.

A bibliometric analysis was carried out over these 237 papers. Tables in the appendix A report the top contributing authors (table A1), countries (table A2), institutions (table A3), papers citation (table A4), and publishers (table A5) in the field of IoT data visualization for BI processes. Findings shows that (1) there is a small number of experts in the field; (2) the distribution by countries reveals the leadership of China, followed by the United States, India and Germany; (3) the most cited papers are focused in Big Data; (4) none institution is detected that stand out significantly in terms of the number of publications and (5) three sources stand out over the others: ACM International Conference Proceeding Series with 14 publications, followed by Advances in Intelligent Systems and Computing publications, and Procedia CIRP both with 11 publications.

3.2 Papers classification

Eighteen LDA models were calculated, the number of topics K varying between 3 and 20. From these, the model with 6 topics was selected because it was the one with the highest coherence value (the coherence score increases rapidly until topic 6, and then starts to follow almost a horizontal trajectory) while least repetition in the words and better interpretable topics. Therefore, as suggested by [44,45], we selected k using a reasonable practice of evaluation among alternative values in such a way that the interpretation of the machine-generated model results becomes as easy as possible from the point of view of a human reader. Then, the validity and the interpretability of the topics were assessed and, as agreed by the authors of this paper, two topics were merged in one topic, and other two topics were merged in one topic too, since the papers discussed similar issues. The four topics are:

Topic 0, Technology infrastructure. This focuses on the technology architecture/infrastructure of the system and proposes available analysis methods, tools and technologies.

Topic 1, Case examples. This focuses on analysing and proposing solutions (ad-hoc or general) for real problems in different areas of enterprise/industry.

Topic 2, Final-user experience. This topic focuses on the final-user experience.

Topic 3, Big Data tools. This topic focuses on how Big Data tools/methods can be used in this field. It seems that Big Data is very important in this field and requires a topic of its own.

Table 2 reports the four validated topics identified along with the top 10 words and the number of papers.

Table 2. Validated topic model

Topic label	Top 10 words	No. Papers
Technology infrastructure	smart, network, energy, application, sensor, security, service, internet, cloud, communication construction, device, sensor, context, user, technology, application, real, environment, site	53
Case examples	process, design, product, production, model, technology, industry, manufacturing, management, development	48
Final-user experience	visualization, user, analysis, network, method, model, visual, set, tool, technique	61
Big Data tools	big, data, processing, tool, stream, application, storage, real, process, analytic big, business, service, analytic, data, analysis, process, technology, application, knowledge	75

Regarding the number of papers in every topic, it has been increasing every year, being topic 3, Big Data tools, the one with the most significant progress in recent years (figure 3).

3.3 State of the art in each topic

In the following, the main research lines identified in each topic are shown (figure 4).

3.3.1 Topic 0, Technology infrastructure

On considering 53 papers on the topic “Technology infrastructure”, which are primarily focused on transformation towards industry 4.0 by means of Information and Communication Technologies (ICT), two major research lines can be inferred: 1) papers that use modern technologies in order to manage energy consumption and to achieve more sophisticated monitoring methods, and 2) papers that made an effort to enhance optimization and productivity in businesses.

Research line 1: papers that use modern ICT in order to manage energy consumption and to achieve more sophisticated monitoring methods

In this research line, several papers paid special attention to develop the Iot technological infrastructure for monitoring buildings and construction sites. Zach et al. [46] presented a scalable approach to monitoring buildings and processing the data by means of data pre-processing algorithms and virtual data points. The good point about their work is that software interfaces are independent of the hardware and can be supported simultaneously, which allows batch processing for various applications. Donnal et al. [47] believed that in order to monitor, measure and control power consumption, high-performance low-cost computers should be employed to make local analysis possible. They proposed the construction of an energy box around these computers as a non-intrusive load monitor (NILM) in order to provide the end-user with visualization of power consumption and the generation of custom reports by means of an integrated scripting engine. Song et al. [48] developed an IoT platform that can monitor and refine the air conditioning operation habits in a house. Their development employed a deep belief network algorithm to analyse the data and recognize consumption patterns, a cloud server to collect the data, and visualization techniques to display power consumption. This resulted in a reduction in energy consumption. Another monitoring system, proposed by Zhang et al. [49] could recognize non-hardhat use, which is considered a crucial safety measure on construction sites. Their IoT-based proposal took advantage of sensors, RFID triggers, smartphones, a web-based application for data visualization, and a cloud server that store and retrieve the collected data. In terms of consumption management, Assad et al. [50] introduced a framework, based on the use and incorporation of virtual models, to predict the key performance indicators in energy consumption in manufacturing systems before being set up. Therefore, they use the VueOne virtual engineering (VE) tool to achieve the best energy-efficient production systems. Finally, another interesting study was carried out by Jin et al. [51] who attempted to eliminate some of the problems on a construction site, including detecting and locating errors and identifying intruders, by developing an intrusion monitoring system based on IoT. In order to develop their system, they took advantage of five key components, including radio-frequency identification (RFID) triggers, safety hardhats, a backend cloud server that would be contacted by a smartphone application, and a web-based management platform.

Other papers in the same research line focused their state-of-the-art monitoring approaches on other areas. Gao et al., [52] Kurniawan et al., [53] and Pachayappan et al. [54] proposed an IoT-based management system for farmlands that used multiple sensors to gather monitoring data belonging to the weather, soil and plant conditions such as soil humidity, pH, the wetness of leaves, etc., intending to ensure the satisfactory situation of the planting environment, improving the economic return and achieving better control over the state of the planted crops in real time. The studies by Gao et al. [52] and Kurniawan et al. [53] were published in the same year and were somewhat similar but with some differences, especially in the hardware they used. Thus, the former employed RFID and ZigBee, and the latter implemented Wemos D1 mini and Arduino. Another paper by Alam et al. [55] presented an integration of the IoT with augmented and virtual reality technologies in order to monitor and maintain the system and ensure the safety of the personnel in an extreme environment. Their emphasis was on using mobile computing equipment on the workers' side and processing real-time data. While the last four papers mentioned aimed to deal with real-time data, Tan et al. [56] managed to perform data mining and to analyse and visualize historical data on the quality of the manufactured products in order to discover and monitor the performance trend of the business. Additionally, Benedetto et al. [57] attempted to solve some consumption management problems by deploying sensors on the production line.

Research line 2: papers that use modern ICT to enhance optimization and productivity in businesses

The second research line is mostly concerned with developing the technological infrastructure to enhance business optimization and productivity. In order to attain optimization of production and productivity in the manufacturing lines of factories, Jinushi et al. [58] suggested a comprehensive system based on IoT that covers the collection, aggregation and visualization of data, to lessen the workers' responsibility and support decision-making. To make optimal decisions and improve the business in the food industry, Goti-Elordi et al. [59] combined and integrated data from different Iot sources by using Business Intelligence tools. In order to have a productive business, the discover-innovate-predict-perform-sustain (DIPPS) model based on the analysis of Iot data was introduced by Rane and Mishra. [60] Apart from that, Yu et al. [61] exploited information technology to propose a BIM-based smart management model for a construction site in order to increase productivity and efficiency. They deployed IoT, cloud computing and Big Data analytics to attend to the issues related to data analysis and storage, augmented and virtual reality, digital processing to pre-process the components before transporting them to the construction site, and three-dimensional scanning to detect probable errors in measurements. Finally, in an attempt to optimize various areas of smart sustainable cities such as freight logistics and citizens' transportation, Beneicke et al. [62] considered using analytical tools, hybrid simulation–optimization techniques such as agent-based simulation (ABS) and machine learning algorithms, to analyse Iot data in order to enhance citizens' behaviour insights and cognition.

3.3.2 Topic 1, Case examples

Topic one, Case examples, comprises all the papers that have proposed solutions (ad-hoc or general) for real problems in different business sectors and corporate areas (48 papers). Different business sectors have been considered: Energy, Health, Building/construction, Smart Cities & transportation, smart factories, Telecommunications, B2C companies/Business Shopping, Geography/Environment & Agriculture, Mining Industry, as well as other proposals suitable for enterprises in any business sector. On the other hand, the corporate functions that have been studied are: Strategic Planning. Manufacturing, Facility Management, Human Resources Management and Risk Management, and there are proposals suitable for any corporate function. Table B1 in appendix B synthesizes the Topic 1 proposals by business sector and corporate function.

In addition to Topic 1 papers, Table B1 in appendix B also includes proposals of papers that belong to other topics and have some kind of example of application. These proposals include the topic number of their paper in brackets. They have been included in Table B1 because in this way the table contains all the real applications of IoT data visualization for BI processes that have been carried out so far. Therefore, this information is gathered in one place for practitioners.

3.3.3 Topic 2, Final-user experience

The second most addressed topic in the pool of papers is "Final-user experience", with 61 papers, primarily focused on data visualization techniques and their application. These papers can be classified into three noticeable research lines: 1) those that made an effort to facilitate human–computer interactions through visualization and simulation and to improve user experience, 2) those intended to develop smart visual monitoring systems, and 3) those that attempted to develop easy-to-use visual intelligent decision-making systems.

Research line 1: Papers that attempt to facilitate human–computer interactions through visualization and simulation

Within this first research line, there are several examples of papers that are mainly focused on human–computer interactions using visualization methods. Pfeffer et al. [92] believed in using interactive surfaces as a means to facilitate collaborative work, as well as virtual and augmented reality to help in problem-solving by representing a product virtually. They illustrated a comparison between possible future technologies deployable in manufacturing plants and the current methods to demonstrate how plant control is going to be changed. They believed that the integration of data collected from various steps in a product life cycle may help the internal and external stakeholders share the information collaboratively. An example of this was developed by Rubart et al. [101] They presented an interactive BI digital boardroom that would enhance user experience and promote the level of interaction between analysts and planners. They made use of a multi-display environment along with multi-touch and multi-user interaction approaches to display data visualizations. Additionally, Shao et al. [102] introduced an IoT-Avatar architectural framework based on mixed reality for human–computer interactions in an IoT system. Their system was adaptive, flexible and engaging and also included a two-way communication method between the IoT system and the representation of the virtual avatar character to deal with the bandwidth deficiency within IoT system communication. Finally, Yun et al. [103] attempted to support decision-makers in industry by proposing a novel visual human–computer interaction decision-making system based on data mining techniques. In order to evaluate the performance of the proposed method, the authors applied various data mining algorithms and assigned different values to the key parameters, and then proved that their method was robust and effective. Table 3 shows a summary of the problems that these papers are trying to solve and their proposed techniques.

Table 3. Visual human–computer interactions proposals

Problem	Represent product virtually	Stakeholders collaboration	Analysts and planners collaboration	IoT human-computer interaction	Support decision-makers
Techniques	Interactive surfaces & virtual and augmented reality	Product Data integration	Interactive BI digital boardroom	Mixed Reality	Data mining
Author	Pfeffer et al. [92]		Rubart et al. [101]	Shao et al. [102]	Yun et al. [103]

Research line 2: Papers that attempt to develop smart visual monitoring systems. These papers are focused on monitoring Smart Factories, Smart Buildings, and Health.

Regarding smart factories, Bojan et al. [93] presented a generic architecture to deal with the visualization of large-scale time-series data that can be employed in various systems, if they work with structured data. Using cloud computing in data collection, as well as sensors and actuators, they managed to implement their proposal on a farm and decreased the number of workers. In addition to this, in order to prioritize end-user attention in high-volume fast data streams, Abuzaid et al. [104] developed a search engine called MacroBase to work with fast data that classifies and explains fast data to the end-user with the help of a combination of streaming classification and data explanation techniques and would lead to an increase in performance. Moreover, in order to monitor the organoleptic properties of the food products in an Iberian ham manufacturing company, García-Esteban et al. [105] proposed a software platform called ICatador based on cloud manufacturing which consists of four collaborative agents. In this proposal, they made use of advanced visualization techniques, communication technologies, and Artificial Intelligence techniques, mainly to facilitate quality testing for the professional taster. Furthermore, to facilitate monitoring of the condition of the machines by the engineers in the manufacturing industry, Olivotti and Eilers [106] proposed a visualization technique that exploits sensor data, with the aim of detecting the reasons underlying anomalies, optimizing maintenance efforts, and increasing the availability of machines. Additionally, Iftikhar et al. [107] used machine learning techniques to produce a solution for real-time analysis and dynamic visualization of the sensor and ERP data in smart manufacturing companies in order to detect possible future faults. Finally, Gu and Gao [81] proposed a visual particle system in digital twin capable of predicting and monitoring industrial production processes and would succeed in reducing production costs. Their system is helpful in terms of the limitations of current digital twin systems regarding simulation and is capable of simulating more complex objects such as gas and fluid.

With regard to smart buildings, some authors contributed to the transition from regular buildings to smart buildings by employing visualization along with other technologies. Fraternali et al. [108] developed a platform in which they visualized the data gathered from smart meters and sensors to display energy consumption patterns in an understandable way for a wide range of users and to give them adaptive recommendations regarding energy saving while encouraging them to collaborate. Their proposal also took advantage of gamification methods to promote awareness and make durable behaviour changes towards sustainable energy consumption. To improve current visualization techniques in a building that are deficient in interaction and immersiveness, Carneiro et al. [109] presented an approach based on IoT by integrating augmented reality (AR) technologies and smart buildings, and provided effective interactive AR visualizations for the occupant to monitor the energy consumption and learn about the interconnection of the building system. Additionally, Kazado et al. [110] mentioned that Building Information Modelling (BIM) is a virtual presentation of a building that displays the exchange, management and communication of data about the building, but it is unable to represent the real-time information related to the performance. Therefore, they introduced an add-in program to integrate BIM and real-time data collected from the existing building sensor technology, within three approaches, and developed a data platform to visualize of indoor environmental parameters. Furthermore, Ceccarini et al. [111] took advantage of IoT and demonstrated a case study of a smart campus in which the heterogeneous data gathered from the sensors is visualized and contributes to a more efficient use of the campus premises.

As regards health issues, Yang et al. [112] introduced a system called VisOSA to monitor patients suffering from a chronic disorder called Obstructive Sleep Apnea (OSA). VisOSA is a web-based application that allows patients to assess their health condition and physician staff to monitor their patients both individually or in a group. Additionally, Lupión et al. [65] and Marques and Pitarma [113] worked on health-related concerns in smart homes. To monitor the health condition of elderly or disabled people in smart homes, Lupión et al. [65] developed an IoT platform with the help of sensors that send the data to the cloud, and then the data are analysed via activity recognition algorithms. Finally, the results would be visualized through a web-based system and necessary alerts or notifications would be issued for doctors or a user's family. Marques and Pitarma [113] presented an approach based on IoT to monitor the environmental noise in a building, since serious health issues can originate from noise pollution. Then they visualized the collected data through web software to assist decision-makers with taking appropriate measures. Finally, to develop a health monitoring system, Elouni et al. [114] proposed integrating a Remote Health Monitoring Systems (RHMS) that uses multi-agent technology with machine learning approaches to deal with the temporal aspect of real-time health data, to extract knowledge from collected data, and to predict the patients' state.

Table 4 summarizes the problems that these papers are trying to solve and their proposed techniques.

Table 4. Proposals for monitoring improvement using visualization methods

Smart Factories	**Problem**	Improve productivity	Quality control	Indicators development	Maintenance		Reduction of production costs
	Techniques	Cloud computing & sensors and actuators	Artificial Intelligence, Cloud computing, Machine learning	Streaming classification and data explanation techniques	An IoT platform & a web-based system	Machine learning	Simulation, Data conversion
	Author	Bojan et al. [93]	García-Esteban et al. [105]	Abuzaid et al. [104]	Olivotti and Eilers [106]	Iftikhar et al. [107]	Gu and Gao [81]
Smart Buildings	**Problem**	Energy saving		Management of indoor environmental parameters		Efficient usage of the University premises	
	Techniques	Gamification methods	IoT & augmented reality	Building Information Modelling & IoT		IoT & a web-based system	
	Author	Fraternali et al. [108]	Carneiro et al. [109]	Kazado et al. [110]		Ceccarini et al. [111]	
Smart Health	**Problem**	Monitor patients suffering from a chronic disorder		Health-related concerns in smart homes		To predict the patients' state	
	Techniques	Web-based application		IoT platform, Cloud computing, a web-based system, and an activity recognition algorithm	IoT platform & a web-based system	multi-agent technology with machine learning	
	Author	Yang et al. [112]		Lupión et al. [65]	Marques and Pitarma [113]	Elouni et al. [114]	

Research line 3: Papers that attempt to develop easy-to-use visual intelligent decision-making systems

Concerning the third research line in this topic, almost all the papers in our pool mentioned supporting decision-makers in some way, though some of them dedicated their work to providing support for decision-making. Teong et al. [115] aimed to discover a way to enable decision-makers to explore data and gain deeper insights from visualized data. They showed that interactive visualization can do the trick and proved the effectiveness of their proposal by applying it to an airline to predict flight delays. Additionally, Hingant et al. [116] offered an enhanced intelligence system called HYBINT that would assist decision-makers with watching over their all-important instruments. They supplied their system with cyber and physical heterogeneous data, and then the output of the analysed data would be represented through visualization techniques in a single visualization space. They applied their work to a real environment and proved that it can enhance situational awareness. Moreover, Ltifi et al. [117] and Alves et al. [118] managed to combine data mining techniques with data visualizations to enhance decision-making. Ltifi et al. [117] claimed that valuable patterns can be extracted from data by employing Decision Support Systems (DSS) based on data mining. To transit from these patterns to knowledge, they proposed a generic approach to help decision-makers take advantage of this knowledge, by using a common visual analytics process. They applied their proposal to a medical case in the Intensive Care Unit and proved its feasibility. Finally, an interactive visual analytics approach called PlanningVis was proposed by Sun et al. [119], which allows the exploration and comparison of production plans. By integrating an automatic planning algorithm with interactive visual explorations, PlanningVis can facilitate the efficient optimization of daily production planning as well as support a quick response to unanticipated incidents in manufacturing.

Table 5 summarizes the problems that these papers are trying to solve and their proposed techniques

Table 5. Proposals for easy-to-use visual intelligent decision-making systems

Problem	To gain deeper insights from visualized data	To identify security vulnerabilities of computer networks	To automatically explore and visually extract insights	Decision support through geographic information systems visualizations	To develop a systematic analytics approach for production planning
Technic	Interactive Visualization	Analytical tools and advanced representation techniques, Web Services	Data Mining		Linear programming and heuristic algorithms
Author	Teong et al. [115]	Hingant et al. [116]	Ltifi et al. [117]	Alves et al. [118]	Sun et al. [119]

3.3.4 Topic 3, Big Data tools

The topic "Big Data tools" is one of the most addressed topics in the academic literature (75 papers, 32% of all papers analysed). It is possible to classify them into two main research lines: 1) Papers that introduce a comprehensive framework of Big Data analysis that mostly covers from IoT data acquisition to knowledge visualization. 2) Papers with a more focused view that discuss the application of visualization techniques on Big Data. In both categories, the authors are mostly seeking to achieve intelligence, and sometimes they manage to integrate other technologies such as cloud computing, Machine learning, Augmented reality, virtual reality, etc. to overcome deficiencies like dealing with heterogeneous data, real-time data, rapid change of data, storing huge amounts of data, etc.

Research line 1: papers that introduce a comprehensive framework of Big Data analysis

Within the first research line, the papers mostly pivot around two business sectors, namely smart factories and transportation industry. As far as smart factories are concerned, there was an evolution from conceptual proposals to frameworks to support specific situations (table 6). Therefore, to accomplish an intelligent factory through the implementation of Industry 4.0 and Big Data, Shafiq et al. [86] presented a conceptual all-inclusive framework that comprises four stages, namely real-time data capture from sensors, PLCs, etc.; data standardization and formalization; semantic analysis; and real-time visualization of key performance indicators (KPI) through a GUI dashboard. Similarly, Liu and Jiang [42] introduced a comprehensive cyber-physical system architecture for the shop floor which would assist in intelligent manufacturing. Their proposed framework covered configurational and operational delicacies from data collection and interconnection between the entities to industrial Big Data analysis and acquisition of knowledge, which would help intelligent decision-making. In addition to that, Campos et al. [120] investigated the characteristics of data and Big Data and highlighted how manufacturers may implement Big Data analytics and technologies, such as data mining, as well as data visualization techniques in their organizations in order to convert the data into information and manage their assets. Additionally, Jung et al. [121] suggested a Big Data analysis framework in which data is collected from three different stages: the distribution stage in which products are distributed, the customer usage stage in which products are used by consumers, and the A/S stage in which products are repaired by repair shops. Data is then analysed and visualized and the analysis output is subsequently handed over to the companies to assist them with improving efficiency at each stage. Moreover, Yu et al.[82] proposed a comprehensive architecture to be used in IoT-based smart factories to assist with fault detection and predictive maintenance. Several technologies such as Apache Spark, OPC Collector, transformation protocols, and encryption methods have been used to produce this manufacturing Big Data ecosystem. Unlike many other papers addressing the same subject, they paid attention to data security issues and guaranteed data security. Similarly, Godinho and Piedade [122] proposed a framework called ACCEPT that was designed for the collection, storage, analysis and visualization of shop floor data through a web application.

Table 6. Big Data frameworks for smart factories and transportation industries

Smart Factories	Define the steps to carry out from data capture to knowledge generation and visualization	Analyze how Big Data technologies	Define a specific framework to make an elastic	Define a specific framework to assist with fault detection	Define a specific framework to assist

			can be used in every step	strategy for new products	and predictive maintenance	production control
	Shafiq et al. [86]	Liu and Jiang [42]	Campos et al. [120]	Jung et al. [121]	Yu et al.[61]	Godinho and Piedade [122]
Transport	Energy-efficient transportation	Complex traffic problems	Maritime traffic management	Air traffic management & energy-efficient transportation		Urban transportation systems
	Xiong et al. [123]	Zhang et al. [49]	Soares et al. [124]	Luo et al. [125]		Khan et al. [126]

With regard to transportation industry, different frameworks has been proposed to solve different kinds of traffic problems (table 6). A new comprehensive framework for Intelligent Transportation System (ITS) was developed by Xiong et al. [123] that took advantage of cloud computing to integrate and share Big Data collected from satellites and helicopters, aka Space-Air-Ground Big Data, with the aim of accomplishing accurate and energy-efficient transportation in real time. Zhang et al. [49] introduced a framework for Distributed Collaborative Urban Traffic Big Data (DCUTBD) system which takes advantage of cloud computing to provide a collaborative platform to share multidimensional traffic data, software and resources among the entities, as well as visualization of complex traffic problems. The authors applied their proposal to a real city to approve its efficiency and feasibility. Aside from these, concerning marine transportations, Soares et al. [124] proposed an agile and comprehensive framework that acquires heterogeneous data streams from various sources such as maritime and marine sensors within IoT infrastructures, integrates and processes the data employing Semantic Web Technologies, and discovers knowledge to help the ships and reveals unusual events. Furthermore, Luo et al. [125] presented a comprehensive Big Data analysis framework that covers collection of data, processing the data and data mining, and data visualization, by integrating Big Data analytics into cyber-physical systems (CPS) to help decision-makers in various areas. They proved its practicality and versatility through two practical cases in power grids and aircraft. Finally, in order to achieve smart transportation management, Khan et al. [126] presented a framework capable of integrating heterogeneous dynamic Big Data gathered from various sources in urban transportation systems. The authors employed data mining and machine learning models and covered all the necessary steps to deal with Big Data, that is, data acquisition, storage, analysis and visualization, together with real-time monitoring and forecasting, with the aim of assisting decision-makers.

Research line 2: papers that discuss the application of visualization techniques on Big Data

The second research line is mostly concerned with the application of visualization techniques to Big Data collected from various sources. The application has been focused in two areas: industrial Big Data and geographical Big Data. To deal with industrial Big Data, Xu et al. [127] designed and implemented a visual analytic system to investigate tremendous amounts of data generated in the assembly lines of factories. Their system would assist with monitoring the performance of assembly lines in real time and the investigation of historical data to uncover anomalies and deficiencies and to aid with finding the reasons for them. For the specific case of agricultural industries, Wu et al. [91] proposed a multi-dimensional information visual analysis approach for market sales Big Data which employs a density-based clustering algorithm that would expel excessive data and just keep the effective information, aiming at providing the agricultural industries with a clear vision about market status and trends, and assisting with smart decision-making. Moreover, an interactive visual analytics system was designed by Wu et al. [128] to allow monitoring of the equipment in a factory in the process industry and thus avoid unplanned downtime and unnecessary routine maintenance. They deployed advanced analytical algorithms and intuitive visualization designs to provide a semi-supervised approach to monitor the condition of the equipment. In addition, Qian et al. [129] proposed a versatile architecture to perform 3D visualization on the shop floor Big Data collected through sensors and IoT technology in real time, with the aim of showing the real-time state of the shop floor comprehensively, thus enhancing production efficiency and reducing production costs. In order to evaluate the behaviour of the Advanced Driver Assistance System (ADAS), Priyadarshini et al. [130] proposed an interactive GUI to perform the analysis and visualization of the tremendous amount of data collected from sensors and vehicles in the automotive industry. Finally, Redondo et al. [83] aimed to handle tremendous amounts of data in industrial companies to support the continuous monitoring of machines and extract knowledge and patterns through a visualization technique called Hybrid Unsupervised Exploratory Plots (HUEPs), which combines Exploratory Projection Pursuit (EPP) and Clustering methods. They applied their proposal to the practical case of an automotive industry sector to test its ability to predict failures.

With regard to geographical data, Bornschlegl et al. [131] introduced an approach to analyse heterogeneous car-to-cloud data through comprehensive visualization so that they can detect anomalies. Aside from these, Kang et al. [132] proposed methods for visualization and spatial-temporal analysis of Big Earth data based on Keyhole Markup Language (KML) and employing Cesium, which is a java library for creating 3D virtual Earth and 2D maps in a web browser and would help reveal the correlation between dimensions and periodic trends within the Big Earth Data. Moreover, Bouloukakis et al. [133] proposed an interactive data visualization system for smart cities to transit from static IoT data visualization to interactive IoT data visualization. To accomplish this system, they took advantage of advanced user interaction techniques and Virtual Reality and overcame difficulties with data complexity and heterogeneity. Moreover, Lock et al. [134] developed an application for visualizing real-time and historical Big Data on the city transportation that applies Augmented Reality and would help with assessing the transportation performance and the results would be beneficial for a wide range of people, namely decision-makers, city planners and citizens.

3.4 New Research Agenda

Technology infrastructure:

The area of research on this topic involves approaches for integrating technologies and making them operate together properly [135], which also raises the need for standardization and improved interoperability [39]. For example, there is a need to find approaches to integrate IoT, Industrial Information Integration Engineering (IIIE), 5G and Blockchain [136]. Concerning IoT, a common consensus on the IoT standards, as well as a clear functional view of different components of IoT operating together are required [135]. Additionally, within an IoT environment, interoperability between various parts such as networking, devices, syntax, semantics and platform should be enhanced through a protocol gateway service [135]. In terms of big data, deploying programming languages such as Python and R in big data analytics should be explored [137]. Aside from these, fog computing was introduced in order to eliminate latency issues caused by transferring data between the cloud and the application in latency-sensitive applications such as healthcare applications. However, it also brought about other requirements that need to be addressed, like the business model, security, privacy and scalability [138].

Case examples:

Although a huge amount of research has been conducted in this field, there is a need to integrate technologies like IoT, Artificial Intelligence, etc. into different business areas in order to achieve a comprehensive transition from traditional to modern and smart. As far as buildings are concerned, more research should be carried out on the integration of Artificial Intelligence (AI) with Building Information Modelling (BIM) in order to make the BIM modelling process less time-consuming [139]. Moreover, the use of Cloud BIM in off-site construction is required in order to enable the stakeholders to share the project data [139]. Additionally, special attention must be paid to enhancing the exploitation of power consumption datasets by using machine learning algorithms that can contribute to reducing energy consumption and shifting to a more sustainable and energy-efficient environment, like using generative adversarial networks (GANs) to improve the quality of collected data, and deep learning models to identify consumption anomalies [140]. Also, in order to monitor energy consumption, deploying more cost-effective hardware to transmit and process data is important, as is the use of IoT sensors and smart meters to help achieve data accuracy and real-time data collection and analysis support. While end-users' power consumption data can be used to extract usage patterns, privacy must be protected. As for other areas, a promising research direction is the deployment of big data by means of drones, UAVs, and satellites to assist with the prediction of natural disasters like floods, bushfires, etc. in order to take necessary measures in advance [137]. Moreover, to monitor pollution, mobile crowdsensing can be used to collect the related big data [141]. Finally, more emphasis should be placed on the hardware and algorithms required to employ Big Data on the whole lifecycle of a product [34].

Final-user experience:

Visualization is one of the technologies that has emerged to directly assist users in various areas and is considered a hot area of research [142], although there are still challenges that need to be addressed. To name but a few, visualization techniques should be equipped with context awareness to be able to visualize the data according to the situation. Also, it is better if the reasons underlying the recommendations delivered by a visualization system are transparent to the users. Additionally, more efforts should be focused on integrating virtual reality techniques into visualization systems in order to make the system interactive by accomplishing the third dimension [142]. Aside from these, regarding energy consumption, effective visualization techniques can be employed to allow the end-users to be aware of their consumption behaviour [140]. Regarding the healthcare system, custom visualization and dashboard panels with more details would effectively help to enhance patient–physician interactions [67].

Big Data tools:

Big Data technologies have facilitated business digitalization, although more efforts are necessary for many related areas. To name just a few instances, firstly, with the growth of heterogeneous big data, some challenges arise, including the storage overhead on the servers, and to overcome this challenge many scholars have suggested using distributed storage like cloud computing [143,144,38,138]. Despite the benefits, it also gives rise to other challenges like the need to take care of data integrity, accountability, availability and authenticity [145], and above all security and privacy concerns [144]. Other than that, the lack of efficient algorithms for querying big data in a cloud environment has led to scalability issues and delayed responses. The second promising research direction is to branch Big Data out into various business fields like the tourism industry to address the problem of over-tourism and to increase overall productivity [146] and more importantly in healthcare [147], such as helping to design predictive systems for early detection of the diseases [39]. The third research direction concerns the approaches to performing effective and thorough analyses and mining of collected data [148], which can be achieved through various approaches, such as by using the power of machine learning [148] and deep learning along with data fusion mechanisms [149]. Finally, an increasing amount of attention has recently been paid to the big data collected through mobile crowdsensing, although more research is required to reveal the power of crowdsourced smartphone-based measurements. Furthermore, obtaining consistent and reliable results from smartphone sensors is still a major concern [150].

4. CONCLUSION

Although IoT data visualization for BI processes enhances the productivity and efficiency of the business, different authors have shown that there are numerous challenges and factors that must be identified and studied in this field.

To advance in this line of knowledge, in this paper a bibliographical analysis of the literature on IoT data visualization for BI processes published since 2009 has been carried out. A sample of 237 papers were analysed in order to identify the evolution over time of the number of articles included on the list, the evolution of the number of citations generated by these articles, the number of articles published by author, the number of articles published by country, the number of articles published by institution, the content of the 10 most cited articles on the list, the number of articles published per journal, the indicators of relevance, impact and prestige of the 10 journals with the most articles published on the list, and the established and emerging research categories on the topic as well as their future research necessities.

The work presented in this paper contributes to the literature on IoT data visualization for BI processes, as it extends the existing bibliographical reviews: (1) it considers IoT, data visualization and Business Intelligence together, since to date none of the existing reviews had considered these three research areas together; (2) it extends the period of the systematic review to 2020; (3) it has a greater coverage of information sources since it uses both the Scopus and Web of Science databases jointly; (4) it identifies the main authors, countries and institutions that contribute in the field of IoT data visualization for BI processes, using statistical analysis and bibliometric analysis techniques to obtain and compare the most influential works (response to RQ1); (5) through a topic modelling using LDA, it identifies and proposes four research categories: Technology infrastructure, Case examples, Final-user experience, and Big Data tools (response to RQ2); and (6) it identifies future research needs in the field of IoT data visualization for BI processes (response to RQ3).

The bibliographical analysis has confirmed the initial hypothesis that an analysis of current research could facilitate the advancement of future research in this field. The main conclusion is that the area of study requires more research and a higher number of annual publications. It is also necessary to improve the relevance of the research carried out, something that could be achieved by accessing journals of greater impact. Finally, the number of papers published in each category is quite balanced. The category with the fewest papers is Case Examples. Therefore, research conducted in this category should be improved, since it is crucial to transfer the knowledge generated by academics to practitioners, thereby allowing the implementation of the advances in the other categories in real enterprises. In this regard, it is notable that not all corporate functions that can take advantage of IoT data visualization for BI have case examples, such as marketing, purchasing, sales, etc. The same happened with the business sectors. This is an important gap in the current research on this category.

Finally, it is important to highlight the limitations of the study. This research was limited mainly by (1) the biases introduced by studying only two bibliographical databases: the Web of Science and Scopus. There was also a language bias, due to the fact that these databases include mostly articles that were written in English, and the search was conducted only in English. Other academic databases specifically dedicated to computer science such as ACM Digital Library or IEEE Xplore Digital Library could be used to improve and compare the results; (2) choosing a series of specific keywords introduced another bias by default. Other keywords could have been used and might have yielded different results; (3) the bibliometric analysis based on LDA was used and the literature was classified in four research clusters. Other methods, such as network citation analysis, might be used for such analysis to compare results; and (4) a spreadsheet has been used to do the systematic literature research. The use of a specialized tool could help to find out new findings.

References

1. Al-Eisawi D, Serrano A, Koulouri T. The effect of organisational absorptive capacity on business intelligence systems efficiency and organisational efficiency. Industrial Management & Data Systems. 2020.

2. Liang T, Liu Y. Research landscape of business intelligence and big data analytics: a bibliometric study. Expert Systems with Applications. 2018;111:2-10.

3. Wang Y, Byrd T.A. Business analytics-enabled decision-making effectiveness through knowledge absorptive capacity in health care. Journal of Knowledge Management. 2017;21(3):517-39.

4. Laursen, G. and Thorlund, J. Business Analytics, Wiley, Hoboken, NJ. 2010

5. Loon M. Knowledge management practice system: theorising from an international meta-standard. Journal of Business Research. 2019;94:432-41.

6. Innovative Enterprise (2020). The Evolution of Business Intelligence. Available at: https://channels.theinnovationenterprise.com/articles/the-evolution-of-business-intelligence

7. Chalmeta R, Grangel R. Performance measurement systems for virtual enterprise integration. International Journal of Computer Integrated Manufacturing. 2005;18:73-84

8. Patel J.A, Sharma P. Online Analytical Processing for Business Intelligence in Big Data. Big data. 2020;8(6):501-518

9. Lavalle A, Maté A, Trujillo J, et al. Visualization Requirements for Business Intelligence Analytics: A Goal-Based, Iterative Framework. IEEE 27th International Requirements Engineering Conference (RE). 2019;109-119.

10. Aufaure, M. What's up in business intelligence? A contextual and knowledge-based perspective. Conceptual Modeling - 32th International Conference ER. 2013;9-18

11. Kumar N. Call for Special Issue Papers: Internet of Things Data Visualization for Business Intelligence. Big data. 2020;8(6):540-541.

12. Gray C.C, Teahan W.J, Perkins D. Understanding our analytics: A visualization survey. Journal of Learning Analytics. 2017

13. Levy Y, Ellis T. A systems approach to conduct an effective literature review in support of information systems research. Informing Science Journal. 2006;9:181-212.

14. Snyder H. Literature review as a research methodology: An overview and guidelines. Journal of Business Research. 2019;104:333-9.

15. Blei DM, Edu BB, Ng AY, Edu AS, Jordan MI, Edu JB 2003. Latent Dirichlet Allocation. Journal of Machine Learning Research. vol. 3, 30.

16. Henderson K, Eliassi-Rad T. Applying latent Dirichlet allocation to group discovery in large graphs. In: Proceedings of the 2009 ACM Symposium on Applied Computing. New York, NY, USA: ACM. 2009

17. van Altena AJ, Moerland PD, Zwinderman AH, Olabarriaga SD. Understanding big data themes from scientific biomedical literature through topic modeling. J Big Data. https://doi.org/10.1186/s40537-016-0057-0. 2016;(3)23

18. Shimada D, Kotani R, Iyatomi H. Document classification through image-based character embedding and wildcard training. Proceedings—2016 IEEE international conference on Big Data, Big Data 2016;3922–3927.

19. Wallach H.M. Topic modeling: beyond bag-of-words. Proceedings of the23rd international conference on Machine learning. 2006;977–984

20. Principe, V.A., de Souza Vale, R.G., de Castro, J.B.P., Carvano, L.M., Pereira Henriques, R.A., de Almeida e Sousa Lobo, V.J., de Alkmim Moreira Nunes, R. (2021) A computational literature review of football performance analysis through probabilistic topic modeling. Artificial Intelligence Review. https://doi.org/10.1007/s10462-021-09998-8

21. Tiba, S; van Rijnsoever, FJ; Hekkert, MP. Firms with benefits: A systematic review of responsible entrepreneurship and corporate social responsibility literature. Corporate Social Responsibility And Environmental Management 2019;26 (2) 265-284

22. Ammirato, S.; Felicetti, A.M.; Raso, C.; Pansera, B.A.; Violi, A. Agritourism and Sustainability: What We Can Learn from a Systematic Literature Review. Sustainability 2020;12, 9575

23. Koga, H. and Taniguchi, T. Developing a User Recommendation Engine on Twitter Using Estimated Latent Topics. Procs. of Human-Computer Interaction. Design and Development Approaches, 2011;461–470

24. Maier D, Waldherr A, Miltner P, et al. Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. Communication Methods and Measures. 2018;12(2-3):93-118.

25. Miwa M, Thomas J, O'Mara-Eves A, Ananiadou S. Reducing systematic review workload through certainty-based screening. J Biomed Inform. 2014;51:242–53.

26. Campbell, J.C., Stroulia, E. (2015). The Art and Science of Analyzing Software Data. Edited by: Christian Bird, Tim Menzies and Thomas Zimmermann. Elsevier https://doi.org/10.1016/C2012-0-07289-4

27. Gensim. Available at: https://radimrehurek.com/gensim/. 2021

28. Pirola, F, Boucher, X, Wiesner, S and Pezzotta, G (2020). Digital Technologies in Product-Service Systems: A Literature Review and a Research Agenda. Computers in Industry, (123).

29 Mongeon, P., Paul-Hus, A. The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics 2016;106, 213–228

30. Martínez-López F. J, Merigó J. M, Valenzuela-Fernández L, et al. Fifty years of the European Journal of Marketing: a bibliometric analysis. In European Journal of Marketing. 2018.

31. Zhu, J., Liu, W. A tale of two databases: the use of Web of Science and Scopus in academic papers. Scientometrics 2020;123, 321–335

32. Hoffman M, Bach F, Blei D. Online learning for latent dirichlet allocation. Advances in neural information processing systems. 2010;23:856-64.

33. Rehurek R, Sojka P. Software framework for topic modelling with large corpora. In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. 2010.

34. Li J, Tao F, Cheng Y, et al. Big data in product lifecycle management. The International Journal of Advanced Manufacturing Technology. 2015;81(1-4):667-84.

35. Esmaeilian B, Behdad S, Wang B. The evolution and future of manufacturing: A review. Journal of Manufacturing Systems. 2016;39:79-100.

36. Zhong RY, Newman ST, Huang GQ, et al. Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. Computers & Industrial Engineering. 2016;101:572-91.

37. Oussous A, Benjelloun FZ, Lahcen AA, et al. Big Data technologies: A survey. Journal of King Saud University-Computer and Information Sciences. 2018;30(4):431-48.

38. Zhong RY, Lan S, Xu C, et al. Visualization of RFID-enabled shopfloor logistics Big Data in Cloud Manufacturing. The International Journal of Advanced Manufacturing Technology. 2016;84(1-4):5-16.

39. Darwish A, Hassanien AE, Elhoseny M, et al. The impact of the hybrid platform of internet of things and cloud computing on healthcare systems: opportunities, challenges, and open problems. Journal of Ambient Intelligence and Humanized Computing. 2019;10(10):4151-66.

40. Li R, Lu B, McDonald-Maier KD. Cognitive assisted living ambient system: a survey. Digital Communications and Networks. 2015;1(4):229-52.

41. Appelbaum D, Kogan A, Vasarhelyi MA. Big Data and analytics in the modern audit engagement: Research needs. Auditing: A Journal of Practice & Theory. 2017;36(4):1-27.

42. Liu C, Jiang P. A cyber-physical system architecture in shop floor for intelligent manufacturing. Procedia Cirp. 2016;56:372-7.

43. Aceto G, Persico V, Pescapé A. The role of Information and Communication Technologies in healthcare: taxonomies, perspectives, and challenges. Journal of Network and Computer Applications. 2018;107:125-54.

44. Chang, J.; Gerrish, S.; Wang, C.; Boyd-Graber, J.; Blei, D. Reading tea leaves: How humans interpret topic models. Adv. Neural Inf. Process. Syst. 2009;22, 288–296

45. Blei, D.M. Probabilistic topic models. Commun. ACM 2012;55, 77–84.

46. Zach R, Hofstätter H, Mahdavi A. A distributed and scalable approach to building monitoring. eWork and eBusiness in Architecture, Engineering and Construction: ECPPM. 2014;231-6.

47. Donnal JS, Paris J, Leeb SB. Energy applications for an energy box. IEEE Internet of Things Journal. 2016;3(5):787-95.

48. Song W, Feng N, Tian Y, et al. A Deep Belief Network for Electricity Utilisation Feature Analysis of Air Conditioners Using a Smart IoT Platform. Journal of Information Processing Systems. 2018;14(1).

49. Zhang H, Yan X, Li H, et al. Real-time alarming, monitoring, and locating for non-hard-hat use in construction. Journal of construction engineering and management. 2019;145(3).

50. Assad F, Alkan B, Chinnathai MK, et al. A framework to predict energy related key performance indicators of manufacturing systems at early design phase. Procedia CIRP. 2019;81:145-50.

51. Jin R, Zhang H, Liu D, et al. IoT-based detecting, locating and alarming of unauthorized intrusion on construction sites. Automation in Construction. 2020;118:103278.

52. Gao G, Jia Y, Xiao K. An IOT-based Multi-sensor Ecological Shared Farmland Management System. International Journal of Online Engineering. 2018. 14(3).

53. Kurniawan F, Nurhayati H, Arif YM, et al. Smart Monitoring Agriculture Based on Internet of Things. 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT). 2018;363-366.

54. Pachayappan M, Ganeshkumar C, Sugundan N. Technological implication and its impact in agricultural sector: An IoT Based Collaboration framework. Procedia Computer Science. 2020;171:1166-73.

55. Alam MF, Katsikas S, Beltramello O, et al. Augmented and virtual reality based monitoring and safety system: A prototype IoT platform. Journal of Network and Computer Applications. 2017;89:109-19.

56. Tan JS, Ang AK, Lu L, et al. Quality analytics in a big data supply chain: commodity data analytics for quality engineering. IEEE Region 10 Conference (TENCON). 2016;3455-3463.

57. Benedetto F, Brunetti D, Gena C, et al. Intelligent monitoring applications for Industry 4.0. Proceedings of the 25th International Conference on Intelligent User Interfaces Companion. 2020;67-68.

58. Jinushi T, Chisaki K, Kawakami Y. Optimization of Factory Production Activities by Utilizing IoT. Fujitsu Science and Technology Journal. 2016;52(4):77-83.

59. Goti-Elordi A, de-la-Calle-Vicente A, Gil-Larrea MJ, et al. Aplicación de un sistema business intelligence en un contexto big data de una empresa industrial alimentaria. DYNA-Ingeniería e Industria. 2017;92(3).

60. Rane SB, Mishra N. Roadmap for business analytics implementation using DIPPS model for sustainable business excellence: case studies from the multiple fields. International Journal of Business Excellence. 2018;15(3):308-34.

61. Yu Z, Peng H, Zeng X, et al. Smarter construction site management using the latest information technology. Proceedings of the Institution of Civil Engineers-Civil Engineering. 2018;172(2)89-95.

62. Beneicke J, Juan AA, Xhafa F, et al. Empowering Citizens’ Cognition and Decision Making in Smart Sustainable Cities. IEEE Consumer Electronics Magazine. 2019;9(1):102-8.

63. Fagiani M, Severini M, Valenti M, et al. rEMpy: a comprehensive software framework for residential energy management. Energy and Buildings. 2018;171:131-43.

64. Lee YJ, Kim EK. Smart device based power generation facility management system in smart grid. 17th International Conference on Advanced Communication Technology (ICACT). 2015;526-532.

65. Lupión M, Redondo JL, Sanjuan JF, et al. Deployment of an IoT Platform for Activity Recognition at the UAL's Smart Home. International Symposium on Ambient Intelligence. 2020;82-92.

66. Mousannif H, Sabah H, Douiji Y, et al. From big data to big projects: a step-by-step roadmap. International Conference on Future Internet of Things and Cloud. 2014;373-378.

67. Fedushko S, Ustyianovych T. Operational Intelligence Software Concepts for Continuous Healthcare Monitoring and Consolidated Data Storage Ecosystem. International Conference on Computer Science, Engineering and Education Applications. 2020;545-557.

68. Ristevski B, Savoska S, Blazheska-Tabakovska N. Opportunities for Big Data Analytics in Healthcare Information Systems Development for Decision Support. ISGT 2020 Information Systems and Grid Technologies. 2020;38-50.

69. Xu Q, Chong HY, Liao PC. Collaborative information integration for construction safety monitoring. Automation in Construction. 2019;102:120-34.

70. Pillsbury W, Clevenger CM, Abdallah M, et al. Capabilities of an Assessment System for Construction Worker Physiology. Journal of Performance of Constructed Facilities. 2020;34(2):04019120.

71. Edirisinghe R, Woo J. BIM-based performance monitoring for smart building management. Facilities. 2020.

72. Taşabat S.E, Özçay T, Sertbaş S, et al. Industry 4.0 Application on Diagnosis Prediction of Construction Machinery: A New Model Approach. Civil Engineering and Architecture, 2020;8(4):404-416.

73. Berglund EZ, Monroe JG, Ahmed I, et al. Smart Infrastructure: A Vision for the Role of the Civil Engineering Profession in Smart Cities. Journal of Infrastructure Systems. 2020.

74. Chen KW, Tsai HM, Hsieh CH, et al. Connected vehicle safety science, system, and framework. IEEE World Forum on Internet of Things (WF-IoT). 2014. 235-240.

75. Avazpour I, Grundy J, Zhu L. V for variety: Lessons learned from complex smart cities data harmonization and integration. IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops). 2016;1-6.

76. Li Z, Zhu S, Hong H, et al. City digital pulse: a cloud based heterogeneous data analysis platform. Multimedia Tools and Applications. 2017;76(8):10893-916.

77. Komamizu T, Amagasa T, Shaikh SA, et al. Towards real-time analysis of smart city data: A case study on city facility utilizations. IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS). 2016;1357-1364.

78. Iliashenko V, Filippova K, Overes E, et al. The modelling of intelligent transport system architecture based on big data technologies. Proceedings of the 33rd International Business Information Management Association Conference, IBIMA 2019: Education Excellence and Innovation Management through Vision 2020. 2019;8553-8561.

79. Rojas E, Bastidas V, Cabrera C. Cities-Board: A Framework to Automate the Development of Smart Cities Dashboards. IEEE Internet of Things Journal. 2020;7(10):10128-36.

80. Gupta R, Budhiraja N, Mago S, et al. An IoT-Based Smart Parking Framework for Smart Cities. Data Management, Analytics and Innovation. 2020;19-32.

81. Gu Z, Gao T. A Visual Particle System Based on Mechanism Model Data in Digital Twin. International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. 2020;183-193.

82. Yu W, Dillon T, Mostafa F, et al. A global manufacturing big data ecosystem for fault detection in predictive maintenance. IEEE Transactions on Industrial Informatics. 2020;16(1):183-92.

83. Redondo R, Herrero Á, Corchado E, et al. A Decision-Making Tool Based on Exploratory Visualization for the Automotive Industry. Applied Sciences. 2020;10(12):4355.

84. Park B, Jeong J. A Novel Rule-based MPS model for SME Manufacturing Factory in Korea. Procedia Computer Science. 2019;155:716-21.

85. Zimmer B, Zenisek J, Jetter HC. Towards Uncertainty Visualization in Smart Production Environments. Proceedings of the 11th International Symposium on Visual Information Communication and Interaction. 2018;116-117.

86. Shafiq SI, Velez G, Toro C, et al. Designing intelligent factory: conceptual framework and empirical validation. Procedia Computer Science. 2016;;96:1801-8.

87. Arora S, Rani R. A Streamlined Approach for Real-Time Data Analytics. Second International Conference on Inventive Communication and Computational Technologies (ICICCT). 2018;732-736

88. Adwernat S, Wolf M, Gerhard D. Optimizing the design review process for cyber-physical systems using virtual reality. Procedia CIRP. 2020;91:710-5.

89. de Regt A, Barnes SJ, Plangger K. The virtual reality value chain. Business Horizons. 2020;63(6):737-48.

90. Dimitris M, Ekaterini V, Zogopoulos V. An iot-based platform for automated customized shopping in distributed environments. Procedia CIRP. 2018;72:892-7.

91. Wu Q, Deng C, Sun R, et al. Multi-dimensional information visualization analysis of business circles and products based on density clustering. IFAC-PapersOnLine. 2018;51(17):811-6.

92. Pfeffer J, Graube M, Reipschlaeger P, et al. Towards collaborative plant control using a distributed information and interaction space. IEEE 20th Conference on Emerging Technologies & Factory Automation (ETFA). 2015;1-4.

93. Bojan VC, Raducu IG, Pop F, et al. Cloud-based service for time series analysis and visualisation in Farm Management System. IEEE International Conference on Intelligent Computer Communication and Processing (ICCP). 2015;425-432.

94. Villeneuve E, Abi Akle A, Merlo C, et al. Decision support in precision sheep farming. IFAC-PapersOnLine. 2019;51(34):236-41.

95. Li Y, Gong J, Hu B. Multisource Remotely Sensed Wetland Information Exploration Using Interactive Visualization Methods. Proceedings of the 2018 2nd International Conference on Big Data and Internet of Things. 2018;95-99.

96. Guillermo JC, García-Cedeño A, Rivas-Lalaleo D, et al. Iot architecture based on wireless sensor network applied to agricultural monitoring: A case of study of cacao crops in ecuador. International Conference of ICT for Adapting Agriculture to Climate Change. 2018;42-57.

97. Botes LA, Hamer W, van Jaarsveld S, et al. Finding the four qualities of intelligent industrial reporting. South African Journal of Industrial Engineering. 2019;30(3):262-76.

98. Rauch E, Dallasega P, Matt DT. The way from lean product development (LPD) to smart product development (SPD). Procedia CIRP. 2016;50:26-31.

99. Richter TO, Felber A, Troester PM, et al. Visualization of requirements engineering data to analyse the current product maturity in the early phase of product development. Procedia CIRP. 2020;91:271-7.

100. Ferreira L, Putnik GD, Lopes N, et al. Disruptive data visualization towards zero-defects diagnostics. Procedia CIRP. 2018;67:374-9.

101. Rubart J, Lietzau B, Söehlke P, et al. Semantic Navigation and Discussion in a Digital Boardroom. IEEE 11th International Conference on Semantic Computing (ICSC) 2017;290-296.

102. Shao Y, Lessio N, Morris A. IoT avatars: mixed reality hybrid objects for core ambient intelligent environments. Procedia Computer Science. 2019;155:433-40.

103. Yun Y, Ma D, Yang M. Human–computer interaction-based Decision Support System with Applications in Data Mining. Future Generation Computer Systems. 2020;114:285-9.

104. Abuzaid F, Bailis P, Ding J, et al. MacroBase: Prioritizing Attention in Fast Data. ACM Transactions on Database Systems (TODS). 2018;43(4):1-45.

105. García-Esteban JA, Curto B, Moreno V, et al. Quality Control Monitoring by using Artificial Neural Networks for an Iberian Ham Industry. ICINCO (2). 2018;638-645.

106. Olivotti D, Eilers D. Understanding Anomalies: Visualizing Sensor Data for Condition Monitoring of Manufacturing Machines. Multikonferenz Wirtschaftsinformatik. 2018.

107. Iftikhar N, Lachowicz BP, Madarasz A, et al. Real-time Visualization of Sensor Data in Smart Manufacturing using Lambda Architecture. Proceedings of the 9th International Conference on Data Science, Technology and Applications. 2020;1:215-222.

108. Fraternali P, Cellina F, Herrera S, et al. A Socio-Technical System Based on Gamification Towards Energy Savings. IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). 2018;59-64.

109. Carneiro JP, Varnosfaderani MP, Balali V, et al. Comprehensible and Interactive Visualizations of Spatial Building Data in Augmented Reality. Computing in Civil Engineering: Visualization, Information Modeling, and Simulation 2019;79-86.

110. Kazado D, Kavgic M, Eskicioglu R. Integrating Building Information Modeling (BIM) and sensor technology for Facility Management. Journal of Information Technology in Construction (ITcon). 2019;24(23):440-58.

111. Ceccarini C, Mirri S, Prandi C, et al. A data visualization exploration to facilitate a sustainable usage of premises in a Smart Campus context. Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good. 2020;24-29.

112. Yang H, Rothrock L, Kumara S, et al. VisOSA: Visualizing obstructive sleep apnea symptoms and comorbidities. IIE Annual Conference. 2013.

113. Marques G, Pitarma R. Noise Exposure in Residential Buildings: An Internet of Things Approach for Enhanced Acoustic Comfort and Occupational Health. 15th Iberian Conference on Information Systems and Technologies (CISTI). 2020;1-6.

114. Elouni J, Ellouzi H, Ltifi H, et al. Intelligent health monitoring system modeling based on machine learning and agent technology. Multiagent and Grid Systems. 2020;16(2):207-26.

115. Teong KS, En Tan Y, Shabbir M, et al. Finding Insights through Interactive Visualization. InProceedings of the 6th International Conference on Information Technology: IoT and Smart City. 2018;49-53.

116. Hingant J, Zambrano M, Pérez FJ, et al. Hybint: a hybrid intelligence system for critical infrastructures protection. Security and Communication Networks. 2018.

117. Ltifi H, Benmohamed E, Kolski C, et al. Adapted Visual Analytics Process for Intelligent Decision-Making: Application in a Medical Context. International journal of information technology & decision making. 2020;19(01):241-82.

118. Alves CM, Cota MP, Castro MR. MLV-Viewer: Data Mining Visualizations in Decision Support System. 15th Iberian Conference on Information Systems and Technologies (CISTI). 2020;1-5.

119. Sun D, Huang R, Chen Y, et al. PlanningVis: A visual analytics approach to production planning in smart factories. IEEE transactions on visualization and computer graphics. 2020 ;26(1):579-89.

120. Campos J, Sharma P, Gabiria UG, et al. A big data analytical architecture for the Asset Management. Procedia CIRP. 2017;64:369-74.

121. Jung JJ, Kim K, Park J. Framework of Big data Analysis about IoT-Home-device for supporting a decision making an effective strategy about new product design. International Conference on Artificial Intelligence in Information and Communication (ICAIIC). 2019;582-584.

122. Godinho EM, Piedade MB. ACCEPT–Web Applications for visualization and analytics of shop floor data. 15th Iberian Conference on Information Systems and Technologies (CISTI) 2020;1-6.

123. Xiong G, Zhu F, Dong X, et al. A kind of novel ITS based on space-air-ground big-data. IEEE intelligent transportation systems magazine. 2016;8(1):10-22.

124. Soares A, Dividino R, Abreu F, et al. CRISIS: integrating AIS and ocean data streams using semantic web standards for event detection. International Conference on Military Communications and Information Systems (ICMCIS). 2019;1-7.

125. Luo S, Liu H, Qi E. Big data analytics–enabled cyber-physical system: model and applications. Industrial Management & Data Systems. 2019.

126. Khan NA, Nebel JC, Khaddaj S, et al. Scalable system for smart urban transport management. Journal of Advanced Transportation. 2020.

127. Xu P, Mei H, Ren L, et al. ViDX: Visual diagnostics of assembly line performance in smart factories. IEEE transactions on visualization and computer graphics. 2017;23(1):291-300.

128. Wu W, Zheng Y, Chen K, et al. A visual analytics approach for equipment condition monitoring in smart factories of process industry. IEEE Pacific Visualization Symposium (PacificVis). 2018;140-149

129. Qian X, Tu J, Lou P. A general architecture of a 3D visualization system for shop floor management. Journal of Intelligent Manufacturing. 2019;30(4):1531-45.

130. Priyadarshini R, Kayalvizhi S, Geetha A, et al. Data driven scene classification to analyze ADAS. International Journal of Recent Technology and Engineering (IJRTE). 2019;8.

131. Bornschlegl, MX, Reis, T, Hemmje, ML. A prototypical reference application of an ivis4bigdata infrastructure supporting anomaly detection on car-to-cloud data. 27th International Conference on Software Engineering and Data EngineeringAt: New Orleans, LA, USA. 2018

132. Kang X, Li J, Fan X. Spatial-Temporal Visualization and Analysis of Earth Data Under Cesium Digital Earth Engine. Proceedings of the 2018 2nd International Conference on Big Data and Internet of Things. 2018;29-32.

133. Bouloukakis M, Partarakis N, Drossis I, et al. Virtual Reality for Smart City Visualization and Monitoring. Mediterranean Cities and Island Communities 2019;1-18.

134. Lock O, Bednarz T, Pettit C. HoloCity–exploring the use of augmented reality cityscapes for collaborative understanding of high-volume urban sensor data. The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry. 2019;1-2.

135. Singh A, Payal A, Bharti S. A walkthrough of the emerging IoT paradigm: Visualizing inside functionalities, key features, and open issues. Journal of Network and Computer Applications. 2019;143:111-51.

136. Chen Y. A survey on industrial information integration 2016–2019. Journal of Industrial Integration and Management. 2020;5(01):33-163.

137. Munawar HS, Qayyum S, Ullah F, et al. Big Data and Its Applications in Smart Real Estate and the Disaster Management Life Cycle: A Systematic Analysis. Big Data and Cognitive Computing. 2020;4(2):4.

138. Fei X, Shah N, Verba N, et al. CPS data streams analytics based on machine learning for Cloud and Fog Computing: A survey. Future Generation Computer Systems. 2019;90:435-50.

139. Yin X, Liu H, Chen Y, et al. Building information modelling for off-site construction: review and future directions. Automation in Construction. 2019;101:72-91.

140. Himeur Y, Alsalemi A, Bensaali F, et al. Building power consumption datasets: Survey, taxonomy and future directions. Energy and Buildings. 2020.

141. Zappatore M, Refolo S, Longo A. Challenges for Air Pollution Monitoring: A Cyber-Physical Social Systems Approach. CPSS@ IOT 2019;38-49.

142. Eldin DM, Hassanien AE, Hassanien EE. Challenges of Big Data Visualization in Internet-of-Things Environments. International Conference on Innovative Computing and Communications. 2020;873-885.

143. Khare S, Totaro M. Big Data in IoT. 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). 2019;1-7.

144. Roman R, Zhou J, Lopez J. On the features and challenges of security and privacy in distributed internet of things. Computer Networks. 2013;57(10):2266-79.

145. Karim A, Siddiqa A, Safdar Z, et al. Big data management in participatory sensing: Issues, trends and future directions. Future Generation Computer Systems. 2020;107:942-55.

146. Mrsic L, Surla G, Balkovic M. Technology-Driven Smart Support System for Tourist Destination Management Organizations. International Conference on Innovative Computing and Communications. 2020;65-76.

147. De Mauro A, Greco M, Grimaldi M. Understanding Big Data through a systematic literature review: The ITMI model. International Journal of Information Technology & Decision Making. 2019;18(04):1433-61.

148. Zhou Y, Lau BP, Koh Z, et al. Understanding Crowd Behaviors in a Social Event by Passive WiFi Sensing and Data Mining. IEEE Internet of Things Journal. 2020;7(5):4442-54.

149. Himeur Y, Alsalemi A, Al-Kababji A, et al. Data fusion strategies for energy efficiency in buildings: Overview, challenges and novel orientations. Information Fusion. 2020;64:99-120.

150. Alavi AH, Buttlar WG. An overview of smartphone technology for citizen-centered, real-time and scalable civil infrastructure monitoring. Future Generation Computer Systems. 2019;93:651-72

Appendix A

A1 Author influence

Regarding the authors, none stood out significantly in terms of the number of publications. This may be a reflection of the small number of experts in the field. There were 21 authors with two papers (see Table A1).

A2 Affiliation statistics

The distribution by countries revealed the leadership of China, followed by the United States, India and Germany. These four countries plus the United Kingdom accounted for approximately half of the sample. On the other hand, from a continental point of view, more research can be highlighted in Asia, followed by Europe (see Table A2).

A3 Analysis by institution

In the analysis of the institutions, none are detected that stand out significantly in terms of the number of publications, so there is no institution with a high degree of specialization in this field. There are twenty institutions with two published documents and only two institutions with more than two publications on the subject (Table A3).

A4 Citation analysis

Table A4 shows the articles with the highest number of citations. The three most cited articles had a total of 708 citations, a figure that is quite high and is the result of the popularity of the research topic.

A5 Sources analysis

In the analysis of the sources, the weight of the three sources with the most publications must be highlighted: ACM International Conference Proceeding Series with 14 publications, followed by Advances in Intelligent Systems and Computing publications, and Procedia CIRP both with 11 (Table A5).
Three impact indicators have been used to assess the relevance of the sources in question: CiteScore, Source Normalized Impact per Paper (SNIP), and SCImago Journal Rank (SJR). CiteScore measures the average number of citations received per document published in the journal. Values are calculated by counting citations over a year for documents published in the three years prior to the calculation and dividing by the number of documents published in those three years. The SNIP measures the impact of citations in a given context and is based on total citations per field of study. The impact of a citation has a greater value in fields where citations are less likely to occur. SJR takes into consideration the prestige of the source in which the article is published. It uses an algorithm similar to Google to establish rankings between websites. It also takes into account the citations of the article. The indicators reflected in Table A5 express the degree of impact, relevance and importance of the source, according to these indicators.

Appendix B

Table B1 synthesizes the Topic 1 proposals by business sector and corporate function.

Figure legends

Figure 1. Research methodology

Figure 2. Trend in the generation of articles

Figure 3. Number of papers per topic and year

Figure 4. Main research lines identified in each topic

Table A1. Authors with the most articles

Author	No. docs
Ziefle M.; Zhong R.Y.; Lan S.; Edirisinghe R.; Huang G.Q.; Shafiq S.I.; Sanin C.; Szczerbicki E.; Xhafa F.; Wolf M.; Lau B.P.L.; Zhou Y.; Yuen C.; Zhang H.; Yan X.; Jin R.; Ltifi H.; Himeur Y.; Alsalemi A.; Bensaali F.; Amira A.	2

Table A2. Countries with the most articles

Country	No. docs
China	38
United States	26
India	19
Germany	18
United Kingdom	15
Italy	12
Spain	11
Australia	10
Portugal	8
Singapore	7

Table A3. Institutions with the most articles

Institution	No. docs
Karlsruhe Institute of Technology, University College London,	**3**
Zhejiang University, Aarhus University, Tsinghua University, Qatar University, University of Alicante, RMIT University, Purdue University, VSB-Technical University of Ostrava, Singapore University of Technology and Design, Tianjin University, University of Sfax, Ruhr-University Bochum, Massachusetts Institute of Technology, Beijing University of Posts and Telecommunications, University of Salamanca, RWTH Aachen University, University of Minho, Politehnica University of Bucharest, University of Colorado, Chinese Academy of Sciences.	**2**

Table A4. Articles with the most citations

Article	Abstract	Cate-gory	Total Cita-tions
Big Data in product lifecycle management [34]	The document reviews the different aspects of Big Data as well as product lifecycle management and answers questions about the feasibility and benefits of the application of Big Data techniques in manufacturing.	1	**257**
The evolution and future of manufacturing: A review [35]	A survey of the elements in manufacturing systems and the state-of-the-art, as well as the trends in manufacturing, is presented in this document.	3	**238**
Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities and future perspectives [36]	The document explores the application of Big Data in various sectors like manufacturing and healthcare, with special attention to Supply Chain Management. Additionally, it reviews the challenges and opportunities, as well as current technologies, in each step of Big Data application, namely data storage, data analysis and processing, and data visualization.	1	**213**
Big Data technologies: A survey [37]	The document discusses Big Data technologies in detail to help organizations choose the most suitable combination of technologies according to their requirements and specifications.	3	**183**
Visualization of RFID-enabled shop floor logistics Big Data in Cloud Manufacturing [38]	Using cloud manufacturing, the document presents a visualization approach to deal with an enormous amount of data that is collected from the RFID sensors on the shop floor.	1	**113**
The impact of the hybrid platform of internet of things and cloud computing on healthcare systems: opportunities, challenges, and open problems [39]	The document mentions that the combination of cloud computing and Internet of Things can bring about significant advantages in various areas within the healthcare system, such as smart hospitals, medicine control and remote medical services, and it reviews the current literature on this combination.	0	**75**
Cognitive assisted living ambient system: a survey [40]	The document presents a review of the information and communication technologies that are used in the field of Ambient Assisted Living in assisting elderly people with their personal and social life.	2	**55**

Article	Abstract	Category	Total Citations
Big Data and analytics in the modern audit engagement: Research needs [41]	The document argues that external auditors should move towards the integration of Big Data analytics in their profession, since client systems have been moving towards the technologies that lead to the production of Big Data, and then it reviews the concerns and opportunities in this regard.	3	**47**
A Cyber-physical System Architecture in Shop Floor for Intelligent Manufacturing [42]	The document introduces a comprehensive cyber-physical system architecture for the shop floor which would assist in intelligent manufacturing. It covers configurational and operational delicacies from data collection and interconnection between the entities to industrial Big Data analysis through their proposed framework, leading to the acquisition of knowledge that would help intelligent decision-making.	3	**41**
The role of Information and Communication Technologies in healthcare: taxonomies, perspectives, and challenges [43]	The document studies the relation between ICT and healthcare and presents a holistic view of the application of ICT technologies in healthcare.	0	**40**

Table A5. Ten sources with the most published articles and their impact indicators.

Source	n	CiteScore	SNIP	SJR
ACM International Conference Proceeding Series	14	0.8	0.333	0.200
Advances in Intelligent Systems and Computing	11	0.9	0.429	0.184
Procedia CIRP	11	3.6	1.144	0.728
CEUR Workshop Proceedings	5	0.6	0.293	0.177
Future Generation Computer Systems	5	10.2	2.687	1.216
Procedia Computer Science	4	2.5	1.000	0.342
Journal of Network and Computer Applications	4	13.8	3.154	1.389
IEEE Internet of Things Journal	4	12.6	4.110	2.607
Journal of Information Technology in Construction	4	3.6	1.382	0.630
Automation in Construction	4	9.5	2.681	1.690

Table B1. Topic 1 proposals by business sector and corporate function

BUSINESS SECTOR	CORPORATE FUNCTIONS			
	Strategic Planning	**Manufacturing**	**Facility Management**	**All Corporate functions**
Energy	1- Fagiani et al. [63] proposed a framework in Python to manage the energy consumption in residential sections. They made use of task scheduling logic and a configuration structure which represents different subsystems, as well as Fault Diagnosis and Overload Manager modules.	1- (T0) Assad et al. [50] introduced a framework, based on the use and incorporation of virtual models, to predict the key performance indicators in energy consumption in manufacturing systems before being set up, implementing the VueOne virtual engineering tool and aiming at achieving the best energy-efficient production systems.	1- A management system for smart grid power generation facilities was proposed by Lee & Kim [64], which made use of image sensors and GPS sensors to identify smart grid power generation facilities and gave the site manager some details such as the management history of the facility and the hardware diagram, so as to facilitate regular check-ups and obtain efficient maintenance.	1- (T0) Song et al. [48] developed an IoT platform capable of monitoring and refining the air conditioning operation habits of a house. In their development, they employed a deep belief network algorithm to analyse the data and recognize consumption patterns, a cloud server to collect the data, and visualization techniques to display power consumption, which consequently led to a better user experience and a reduction in energy consumption.
Health	**Human Resources**	**All Corporate functions**		
	1- (T2) In order to monitor the health condition of elderly or disabled people in smart homes, Lupión et al. [65] developed an IoT platform with the help of sensors that send the data to the cloud. The data would then be analysed by means of activity recognition algorithms and finally the results would be visualized through a web-based system and necessary alerts or notifications would be issued for doctors or the user's family.	1- Mousannif et al. [66] introduced the steps to build Big Data projects in any organization and reviewed the possible platforms and tools to be used in each step. 2- Fedushko & Ustyianovych [67] developed a software package for automating healthcare functions through the implementation of a complex monitoring system and a data collection mechanism. Their application contained consolidated information and interactive dashboards to see all the information that would help in decision-making. 3- Ristevski et al. [68] indicated that software platforms should be developed to deal with healthcare Big Data in which special attention must be paid to security and privacy for all the parties, especially the patients. Additionally, this software should be able to classify the analysed data by patients, population, epidemic, clinical symptoms and country to make decision-making more effective.		
	Human Resources	**Risk Management**	**Facility Management**	**All Corporate functions**

Building / construction	1- Aimed at safety monitoring in building projects, Xu et al. [69] developed a collaborative framework to integrate information gathered from five layers, namely: the preparation layer, monitoring layer, integration layer, reporting layer and intervention layer by using various smart sensors. Through this framework, information can be collected and analysed and distributed among the stakeholders. The authors managed to evaluate their proposal by developing a prototype in China as a case study to prove its ability to improve site safety.	1- Pillsbury et al. [70] used physiological status monitors to collect and analyse physical indicators related to vital signs such as the respiration rate of the workers on a construction site. 2- (T0) A monitoring system is proposed by Zhang et al. [49] that is capable of recognizing non-hardhat use, which is considered a crucial safety measure on construction sites. In their IoT-based proposal, they used sensors, radio-frequency identification (RFID) triggers, smartphones, a web-based application for data visualization, and a cloud server which would be used for storing and retrieving the collected data.	1- Edirisinghe & Woo [71] proposed a solution for visualizing building performance data, both physical sensor measurements and occupancy evaluation, in real time in order to assist facility management decisions and to provide designers with feedback on workplace preferences and space needs.	1- Taşabat et al. [72] developed a model to predict possible faults and vehicle maintenance needs in construction vehicles. In their model they made use of data mining, machine learning, the internet of things and cloud technology. 2- (T0) Yu et al. [61] exploited information technology to propose a BIM-based smart management model for a construction site in order to increase productivity and efficiency. They deployed IoT, cloud computing and Big Data analytics to address the issues related to data analysis and storage, augmented and virtual reality, digital processing to pre-process the components before transporting them to the construction site, and three-dimensional scanning to detect probable errors in measurements.
	Human Resources	**All Corporate functions**		
Smart Cities & transportation	1- Berglund et al. [73] identified various new roles in civil engineering which would appear as a result of applying smart technologies such as IoT to civil engineering professions. 2- Chen et al. [74] introduced a framework to develop a Machine-to-Machine-based proactive driver assistance system and used it in an intelligent transportation system to help drivers to avoid potential traffic accidents. 3- (T0) In an attempt to achieve optimization in various areas of smart sustainable cities such as freight logistics and citizens' transportation, Beneicke et al. [62] suggested using analytical tools, namely hybrid simulation-optimization and machine learning algorithms to analyse	1- Avazpour et al. [75] managed to deal with the huge amount of data obtained from a transportation system in a smart city and pointed out the challenges they face during their work, including noisy data, diverse data formats, data modelling and increasing demand for sophisticated visualization support. 2- Li et al. [76] introduced a cloud-based platform called City Digital Pulse (CDP), which is an end-to-end architecture covering all parts of Big Data analysis that consists of five major components: collection of data through soft sensors, storing the data in a cache database and a main database, analysing the data by algorithms, visualization of the data via a web service, and finally integration of all the components into the cloud. 3- Komamizu et al. [77] presented a real-time analytical system based on an OLAP system for streaming data in order to analyse the smart city real-time data. 4- Iliashenko et al. [78] proposed an Intelligent Transport System that allowed for Big Data analysis and on-line decision-making in a transportation system. 5- Rojas et al. [79] presented a framework called Cities-Board based on model-driven engineering to automate the development of smart cities dashboards by means of a graphic domain-specific language (DSL). 6- Gupta et al. [80] proposed a smart parking system using IoT, cloud-based services and smart phones that enables payment systems along with real-time data analytics.		

	masses of data in order to enhance citizens' insights and cognition.			
	Product Design and Development	**Strategic Planning**	**Manufacturing**	**Human Resources**

Smart factories	1- (T2) Gu & Gao [81] proposed a visual particle system in digital twin that is able to predict and monitor industrial production processes and would reduce production costs. Their system is helpful in simulating limitations of current digital twin systems and is capable of simulating more complex objects such as gas and fluid.	1- (T3) Yu et al. [82] proposed a comprehensive architecture to be used in IoT-based smart factories to assist with fault detection and predictive maintenance. Several technologies such as Apache Spark, OPC Collector, transformation protocols and encryption methods have been used to produce this manufacturing Big Data ecosystem. Unlike many other papers addressing the same subject, they paid attention to data security issues and guaranteed data security. 2- (T3) Redondo et al. [83] attempted to handle tremendous amounts of data in industrial companies, so as to support the continuous monitoring of machines and extract the knowledge and the patterns through a visualization technique called Hybrid Unsupervised Exploratory Plots (HUEPs) that combines Exploratory Projection Pursuit (EPP) and Clustering methods. They applied their proposal to the practical case of an automotive industry sector to test its ability to predict failures.	1- Park & Jeong [84] solved manufacturing and production defects by applying a smart rule-based MPS model to small and medium enterprise manufacturing. 2- Zimmer et al. [85] proposed data visualization as a means to analyse the potentials and limits of prediction models in smart manufacturing in order to manage the uncertainties appearing in predictive models. 3- (T0) To achieve optimization of production and productivity in factories' manufacturing lines, Jinushi et al. [58] suggested a comprehensive system based on IoT. This system covers the collection, aggregation and visualization of data, in an attempt to lessen the workers' responsibility and support decision-making.
	Risk Management	**All Corporate functions**	
Smart factories	1- Alam et al. [55] developed an Augmented and Virtual Reality (AR/VR)-based IoT prototype system, aimed at combining and integrating several technologies to enhance safety and availability, eliminating errors, and decreasing the time interventions in complex environments like factories.	1- (T3) In order to produce an intelligent factory through the implementation of Industry 4.0, Shafiq et al. [86] presented a conceptual all-inclusive framework that comprises four stages, namely, real-time data capture from sensors, PLCs, etc., data standardization and formalization, semantic analysis, and real-time visualization of key performance indicators (KPI) through a GUI-dashboard.	
	Risk Management		

Telecommunications	Arora & Rani [87] proposed a real-time data streaming method of analysis for fraud detection in the data gathered from telecommunication servers, using the Azure framework.			
	Product Design and development	**All Corporate functions**		
B2C companies/ Business/ Shopping/	1- Adwernat et al. [88] aimed to bridge the gap between the physical and digital world in the IoT and cyber-physical system by taking advantage of virtual reality, in order to optimize the design review process of products. 2- Dimitris et al. [90] used an innovative IoT platform that recognizes the customers' needs in order to carry out the adaptive and effective production of customized products.	1- De Regt et al. [89] developed a Virtual Reality value chain that makes it clear whether the use of Virtual Reality technologies adds value for key stakeholders.		
	Strategic Planning	**Human Resources**	**Facility Management**	**All Corporate functions**
Geography/Environment & Agriculture	1- (T3) Wu et al. [91] proposed a multi-dimensional information visual analysis approach for market sales Big Data by means of a density-based clustering algorithm that would expel excessive data and keep only the effective information, with the aim of providing the agricultural industries with a clear vision about market status and trends, and assisting with smart decision-making.	1- (T2) Pfeffer et al. [92] believed in using interactive surfaces as a means to facilitate collaborative work, as well as virtual and augmented reality to help in problem-solving through the virtual representation of a product. They illustrated a comparison between possible future technologies deployable in manufacturing plants and the current methods in order to demonstrate how plant control is going to be changed. They believed that the integration of the data collected from various steps in a product life cycle may help the internal and external stakeholders share the information in a collaborative manner. 2- (T2) Bojan et al. [93] presented a generic architecture to deal with the visualization of large-scale time-series data that can be used in various systems, if they work with structured data. By making use of cloud computing in data collection, sensors, and actuators, they managed to implement their proposal on a farm and	1- (T0) Gao et al. [52] proposed an IoT-based management system for farmlands that uses multi sensors in order to improve the economic return and achieve a better control over the state of the planted crops in real time.	1- Villeneuve et al. [94] fitted the sheep on a farm with sensors and made use of a decision support system to deal with huge amounts of real-time data and expert knowledge. 2- Li et al. [95] managed to classify the wetlands on Earth by visualizing the data gathered by Earth observation technologies. 3- Guillermo et al. [96] designed a Wireless Sensor Network based IoT architecture, called MOCCA, to monitor agricultural plantation and improve the cultivation process. Their system collects, stores, manages and visualizes the data and sends alerts to the farmer about specific events.

		decrease the number of workers.		
	All			
Mining Industry	1- Botes et al. [97] identified four qualities, namely, establishing a focus area, data availability, analytics and visualization, that contributed to intelligent reporting, and applied them to a case study to evaluate the current intelligence of the reporting of the case study and areas for improvement based on data-driven decision-making.			
	Product Design and development	**Manufacturing**	**Human Resources**	**Risk Management**

All Business Sectors	1. Rauch et al. [98] illustrated how to design Lean Product Development Processes based on the Axiomatic Design methodology. Their approach was then linked to Industry 4.0 while showing how to accomplish a smart and lean product development process by means of advanced technologies. 2. Li et al. [34] divided the life cycle of a product into three parts and investigated the application of Big Data in each part of the lifecycle of a product. 3. Richter et al. [99] proposed an approach to analyse product maturity in the early phase of production by visualization of the existing model of objectives.	1. Zhong et al. [38] proposed an approach to visualize the Big Data generated by the RFID-enabled shop floor logistics in Cloud Manufacturing, in order to enhance decision-making. 2. Ferreira et al. [100] presented a four-part proposal in order to achieve Zero-Defects Diagnostics: a) a set of visualization tools based on virtual and augmented reality to obtain a global perspective of production; b) a data analytics tool to help with decision-making; c) a resource federated network that enables communication between all existing resources; and d) a dynamic context-aware dashboard to contribute to intelligent manufacturing systems.	1- Lavalle et al. [9] proposed a comprehensive iterative goal-based modelling approach aimed at helping non-technical users to communicate their visual analytic needs and choose a suitable visualization technique, regardless of their lack of knowledge in this field.	1- Alam et al. [55] presented an integration of the IoT with augmented and virtual reality technologies in order to monitor and maintain the system and ensure the safety of the personnel in an extreme environment.