Urban Governance 4 (2024) 255–273

Contents lists available at ScienceDirect

Urban Governance

journal homepage: www.elsevier.com/locate/ugj

Understanding big data and data protection measures in smart city strategies: An analysis of 28 cities

Martin Lnenicka[a,*], Petr Hervert[b], Oldrich Horak[b]

[a] *University of Hradec Kralove, Faculty of Informatics and Management, Hradec Kralove, Czech Republic*
[b] *University of Pardubice, Faculty of Economics and Administration, Pardubice, Czech Republic*

ARTICLE INFO

Keywords:
Big data
Data protection
Smart city
Strategy
Delphi method
Recommendations

ABSTRACT

The Smart City concept aims to improve urban governance and optimize public services, ultimately enhancing the quality of life for citizens. As data generation and processing grow rapidly in volume, velocity, and variety, Smart Cities must integrate secure big data considerations into their strategic frameworks and project implementations. This paper explores how big data and data protection measures are represented in the strategies of 28 cities worldwide. To achieve this, we employed a three-phase research methodology: 1) identifying resources, 2) conducting content analysis, and 3) using the Delphi method. Our findings indicate that only half of the cities explicitly address big data in their strategies, and most lack adequate data protection measures. Additionally, the paper presents a list of recommendations for big data management and data protection, derived from measures found in Smart City strategies and validated by domain experts through the Delphi method. These recommendations aim to enhance understanding of how to effectively incorporate big data and its protection into urban planning and Smart City projects. However, it is important to note that these insights primarily apply to larger urban areas with abundant resources.

1. Introduction

The city has always been associated with technological advancements, with cities providing the environment for developing modern technologies and being built using these technologies (Barlow & Levy-Bencheton, 2018; Bawany & Shamsi, 2015; Halegoua, 2020). It is still the case today; only technologies are becoming more advanced (Bibri et al., 2023; Rani et al., 2021). The Smart City concept should address questions about urban governance, such as what aspects of cities can be managed and how cities can be planned to create communities, manage growth, or allocate resources and services efficiently (Halegoua, 2020). At the same time, the specific definition and characteristics of Smart Cities exhibit variability across sources, a unifying theme persists: the core utilization of Information and Communication Technologies (ICT) to enhance the quality of life for citizens. It is achieved through the efficient and effective delivery of public services (Bawany & Shamsi, 2015; Czupich, 2019). Smart Cities can also be conceptualized as environments, or even ecosystems, where strategically implemented ICT elements function as data collection tools. In these data ecosystems, data are collected, processed, analyzed, published, and used to support the decision-making processes of city officials, better allocate resources, deliver public services, enhance transparency and openness of public actions, and sustainable development of the city (Lnenicka et al., 2024).

The significance of big data has increased in recent years. Smart City operations inherently generate voluminous datasets reflecting daily activities (Hashem et al., 2016; Osman, 2019). Trends like the interconnected networks and sensors comprising the Internet of Things (IoT), the techniques and algorithms of Artificial Intelligence (AI), and open data initiatives making large datasets publicly available and accessible without restrictions amplify data collection, encompassing both static and real-time elements (Bibri et al., 2023; Lnenicka & Komarkova, 2019). Considering the understanding of urban dynamics (Batty, 2013), they provide opportunities for improving decision-making and generating insights related to monitoring, analyzing, and planning cities to enhance sustainability, efficiency, resilience, and overall quality of life (Al Nuaimi et al., 2015; Hashem et al., 2016; Lim et al., 2018). Kitchin (2014)) posits that big data presents an opportunity for governments and officials to achieve enhanced efficiency and effectiveness in urban management and regulation. By leveraging these data, city officials can elucidate the impact of various change types and rates on urban communities and urban planning (Barlow & Levy-Bencheton, 2018; Batty, 2013). The increased volumes of data require focusing on challenges related to data protection and security (Hussein, 2020; Pelton &

* Corresponding author.
E-mail addresses: martin.lnenicka@uhk.cz (M. Lnenicka), petr.hervert@student.upce.cz (P. Hervert), oldrich.horak@upce.cz (O. Horak).

https://doi.org/10.1016/j.ugj.2024.12.008
Received 20 October 2024; Received in revised form 18 November 2024; Accepted 6 December 2024
Available online 10 December 2024

Singh, 2018). The interplay between these challenges and big data use embodies a balance between leveraging innovative technologies and ensuring robust protection mechanisms (Bokolo, 2023; Sun et al., 2021). As urban areas integrate complex IoT systems, cloud computing, and AI-driven analytics, their capacity to generate and utilize data improves, offering enhanced public services, traffic management, and sustainability solutions. However, this progress brings an inevitable rise in cybersecurity risks and privacy concerns, necessitating a nuanced examination of their dynamics (Braun et al., 2018; Ismagilova et al., 2022; Javed et al., 2022).

In the context of big data technology integration in Smart Cities, the existence of a Smart City strategy, together with the support of city officials, plays a key role (Hashem et al., 2016); Pivar, 2020; Wu & Chen, 2021). Best practices and guidelines included in these strategies and related projects can help understand respective challenges and how to overcome them. Without robust data protection, Smart Cities may face significant privacy violations, data breaches, and misuse of sensitive information, undermining public trust and putting citizens at risk. In addition, poorly managed data can lead to vulnerabilities in critical infrastructure, making cities susceptible to cyberattacks, which could disrupt essential services, from transportation to healthcare. The integration of big data and data protection into Smart City strategies is not just a technical requirement; it is essential for safeguarding citizens' rights, ensuring the resilience of urban infrastructure, and fostering public confidence in the Smart City vision.

Given the potential of big data to address these challenges faced by Smart Cities, this paper aims to provide an understanding of how big data and data protection measures are reflected in existing Smart Cities strategies in selected Smart Cities. A research methodology consisting of three phases, 1) resources identification, 2) content analysis, and 3) the Delphi method, is adopted to achieve the aim. This paper provides a list of big data and data protection recommendations, which were transformed from measures identified in Smart Cities strategies and then validated by domain experts using the Delphi method.

The remainder of the paper is structured as follows: Section 2 focuses on the research background, Section 3 provides a research methodology, Section 4 presents data analysis and results, Section 5 provides discussion and limitations, and Section 6 concludes the paper.

2. Research background

2.1. Basic concepts and their interpretation

Big data are defined through the model of 3V(s): volume, velocity, and variety. These attributes pose challenges for traditional data processing technologies due to limitations in handling massive data size, rapid data generation, and inherent data heterogeneity (Favaretto et al., 2020). This 3Vs model has served as a springboard for further refinement. Researchers like Holubová et al. (2015) have emphasized the significance of data veracity (accuracy) and value (importance to the owner), expanding the definition to the 5Vs. More recent considerations include validity (data usability lifespan) and volatility (data storage duration), highlighting the evolving nature of big data characteristics (Favaretto et al., 2020; Holubová et al., 2015). Several sources propose additional characteristics to capture the complexity of big data. Arockia et al. (2017) proposed an extensive 17Vs framework, while Hussein (2020)) suggested a comprehensive 56Vs characterization. These evolving definitions highlight the multifaceted nature of big data and how they affect not only the volume and variety of data generated but also the challenges associated with managing, securing, and analyzing them. Beyond the specific characteristics, big data require specialized platforms, tools, and services throughout their lifecycle management. It encompasses data acquisition, transmission, analysis, visualization, and dissemination of outputs (Favaretto et al., 2020; Holubová et al., 2015).

Data protection in Smart Cities is a critical concern as these urban environments increasingly rely on big data that are collected, processed, and shared at unprecedented scales (Hussein, 2020; Pelton & Singh, 2018). However, with this increased reliance on data comes heightened risks. Data breaches, cyberattacks, and unauthorized access to personal information have become more common, posing significant threats to individuals' privacy and security. Sensitive personal data can be misused for identity theft, fraud, and surveillance, leading to severe consequences for individuals, businesses, and governments (Braun et al., 2018; Hussein, 2020). The 3Vs of big data significantly amplify risks to citizens' privacy, underscoring the critical need for robust data protection in Smart Cities (Sun et al., 2021). The high velocity of data, generated and processed in real-time by IoT devices and sensors, often bypasses comprehensive security checks, creating vulnerabilities for potential breaches. Simultaneously, the immense volume of data produced by urban environments increases the attack surface, exposing systems to risks like unauthorized access and data misuse. These challenges demand advanced security measures, including real-time encryption and privacy-preserving technologies, to protect sensitive information while maintaining the operational efficiency of Smart City systems. Addressing these risks is essential to building citizen trust and ensuring the safe implementation of data-driven urban innovations (Bokolo, 2023; Ismagilova et al., 2022; Javed et al., 2022).

The effective implementation of data protection in Smart Cities is hindered by complex socio-technical factors. Smart Cities operate at the intersection of advanced technologies and governance frameworks, where differences in legislative structures can significantly influence the scope and effectiveness of data protection efforts. As Bellanova and De Goede (2022) emphasize, regulatory environments play a critical role in embedding security practices into socio-technical infrastructures, determining the degree to which data protection is enforced within urban spaces. Additionally, cities with fragmented infrastructures, as highlighted by Hacker and Binz (2021), often face difficulties in coordinating data management practices, leading to uneven application of privacy safeguards. The dynamic nature of Smart Cities requires continuous adaptation to evolving technologies and threats. Additionally, the practical implementation of data protection policies must account for the specific socio-cultural contexts in which they operate. To address these challenges, Williams et al. (2013) propose value-sensitive design principles that align technical solutions with societal values, ensuring that data protection measures are both effective and culturally appropriate.

Smart City strategies articulate each city's unique vision for leveraging ICT, data, and innovative solutions to achieve urban development goals. These strategies encompass comprehensive plans for integrating these elements to enhance resident quality of life, optimize service delivery and infrastructure, and promote sustainable practices. The implementation of Smart Cities can cover a broad range of issues tied to Sustainable Development Goals, which constituted an ambitious challenge to address the growing economic, social and environmental concerns development (Sharifi et al., 2024). Various approaches and methods have gradually introduced the framework for sustainable development in connection with the Smart City concept, as many countries lack a proper framework for the use of big data (Camerin et al., 2024). Ultimately, Smart City strategies aim to translate digital technologies and data-driven decision-making into concrete solutions for urban challenges, fostering economic growth and social well-being (Isarsoft GmbH, 2023; Pivar, 2020). Smart City strategies exhibit differentiation based on their geographical scope. National or country-wide strategies encompass broader goals, while local strategies (neighborhood, municipality, city, metropolitan area, or region) cater to specific contexts. A further distinction lies in their focus: infrastructure-oriented strategies prioritize physical elements like transportation, energy, or waste management. Soft infrastructure-oriented strategies emphasize human capital development, including knowledge, inclusivity, social innovation, and social justice (Angelidou, 2014).

In the Smart City context, a project's definition is divided into two directions. For example, Rosati and Conti (2016) discuss the project as an overall process of building a Smart City. Wu and Chen (2021), on the other hand, discuss projects in terms of project management characteristics and divide the Smart City concept into projects in different areas. Residents are key stakeholders in Smart Cities, contributing data and benefiting from data-driven services. While governments and technology companies often lead Smart City projects, public engagement is crucial for ensuring transparency, trust, and alignment with resident needs (Del-Real et al., 2023).

Multiple frameworks and indices exist for evaluating Smart Cities, each offering a unique perspective. These frameworks are typically published as reports, employing various categories and indicators to assess cities. Often, they include a methodology for applying the framework and ranking cities based on the results. The frequency of updates in these reports is usually annual or biennial. Researchers continually strive to refine Smart City assessment by comparing existing frameworks and proposing improvements (Escolar et al., 2019; Sharifi, 2019). It includes incorporating modern technologies and ensuring comprehensiveness. An example is the study of Lai and Cole (2023), which proposed nine core categories for a comprehensive index. These categories are built upon six widely recognized dimensions (economy, environment, governance, quality of life, mobility, and citizenry) and three additional elements: perception, privacy, and cybersecurity (Lai & Cole, 2023).

The concept of Smart Cities, though often hailed for its promise of increased efficiency and improved urban living, also poses a range of ethical and social challenges. Ziosi et al. (2024) point to concerns around privacy and data governance, as Smart Cities rely heavily on pervasive data collection, potentially leading to surveillance and loss of individual autonomy. This data-centric approach may prioritize economic efficiency over ethical considerations, risking public trust and complicating community relations. Additionally, Middha and McShane (2022) highlight a phenomenon they call e-gentrification, where digital infrastructure advancements and data-driven city planning can inadvertently displace marginalized groups and drive socioeconomic divides within urban environments. Such digital-driven urban development can privilege affluent, tech-savvy demographics while marginalizing others, undermining social inclusivity and equitable access to urban resources. Mykhnenko (2023) adds that Smart City initiatives may neglect urban areas on the inner peripheries, where declining populations and aging infrastructure are prevalent. This lack of attention further exacerbates regional inequalities, as resources and investments are funneled into thriving urban centers, leaving peripheral areas to stagnate. Collectively, these perspectives emphasize that the current trajectory of Smart City development, if unchecked, risks deepening socioeconomic divides, compromising digital rights, and potentially alienating less advantaged urban populations.

2.2. Applications of big data and data protection in smart cities

Most of the applications of big data and data protection in Smart Cities are presented in the context of big data analytics and the respective steps of working with big data (Chauhan et al., 2016; Hashem et al., 2016). Lim et al. (2018) reported that data-based innovations in Smart Cities should focus on developing innovative applications that generate value for stakeholders, particularly citizens. Smart City applications should be functional and affordable (Li et al., 2015). Al Nuaimi et al. (2015) proposed requirements for big data Smart City applications to ensure their effective and efficient design and implementation, highlighting the importance of security, privacy, as well as citizen awareness, and government role. Along similar lines, Chauhan et al. (2016) found that security and privacy are critical elements that affect big data Vs and must be considered in adopting the process of effective management of big data in Smart Cities. The Smart City sensor networks produce massive volumes of heterogeneous data at high velocities. A robust, scalable, and reliable big data management system capable of efficiently handling diverse data types should be thus implemented (Bawany and Shamsi, 2015).

According to Wu et al. (2018), developing a data-centric Smart City necessitates robust data mining capabilities and the seamless integration of diverse data formats. It entails using data mining techniques and practices to extract valuable insights from the vast amount of big urban data. Khan et al. (2013) proposed a cloud-based analysis service for big data processing and analysis in Smart Cities, aiming to generate information intelligence and enhance decision-making. The integration of IoT, cloud computing, and data mining approaches to deal with big data in Smart Cities was emphasized by Li et al. (2015). Smart Cities should ensure standardized data acquisition, access, and support for iterative and sequential data processing and scalability (Osman, 2019). Thus, many requirements can be identified. Table 1 summarizes the requirements for big data systems in the context of Smart Cities.

The complexity of Smart Cities requires a structured definition. Various sources and indexes use frameworks to break down the concept into distinct areas for thorough evaluation. This information provides insights into two key aspects: (1) it highlights current priorities in Smart City development, aiding in the selection of effective solutions and best practices, and (2) by comparing these dimensions over time, especially through regularly published indexes, we can learn from past successes and failures in strategies and projects. It is important to note that Table 2 only includes dimensions supported by multiple sources. The broader literature may cover additional areas, using different terms or merging existing categories. The table highlights the nine most frequently referenced dimensions, which are emphasized in academic literature and serve as criteria for Smart City assessment indexes.

The Smart People dimension focuses on building connections between individuals and their city, covering education, participation, and social interaction. While Smart Household overlaps with this dimension, it is more aligned with Smart Life, which addresses quality-of-life aspects such as housing, culture, and tourism. The Smart Industry dimension goes beyond traditional sectors, analyzing how businesses integrate with Smart City infrastructure and adopt new technologies. Smart Network refers to the interconnected infrastructure for resources like electricity, heating, and water distribution. Smart Healthcare encompasses the functionality of hospitals and private healthcare facilities, expanding to include Smart Health, which tracks factors like citizen satisfaction with healthcare services and ease of booking online appointments.

Smart Transport, or mobility, includes urban transportation elements such as public transit, smart parking, shared mobility services (like carpooling and bike rentals), congestion management, and traffic information systems. Smart Government involves technology-driven management strategies, including spatial planning, online civic engagement platforms (such as petitions and voting), and data analytics for citizen outreach, along with keeping residents informed about city operations. The Smart Environment dimension addresses the city's climate and ecological health, monitoring natural conditions and pollution through IoT sensors, while also managing natural resources. Finally, Smart Security focuses on cybersecurity and data privacy within the Smart City framework.

3. Research methodology

The research methodology, as illustrated in Fig. 1, can be conceptualized as a three-stage process encompassing a total of 15 steps. The first stage focuses on identifying relevant sources for the subsequent analysis, which constitutes the second stage. The third stage then involves validation utilizing the Delphi Method.

3.1. Resources identification

Initially, we select a representative sample of cities. This sample will serve as the foundation for our subsequent analysis. Given its prominence in the relevant literature and other sources discussed previously,

Table 1
Selected requirements for big data systems (source: own elaboration).

Requirement	Description	References
Scalable and resilient data infrastructure	The ever-increasing volume and fluctuating nature of data streams necessitate the development of robust networks within Smart City infrastructures. These networks must be equipped to handle the growing data demands while maintaining stability amidst data flow variations.	Al Nuaimi et al. (2015), Bawany and Shamsi (2015), Hashem et al. (2016), Khan et al. (2013)
Data management	Smart Cities inherently generate vast amounts of data from diverse sources. The subsequent challenge lies in effectively extracting, processing, and integrating this data from various sources.	Al Nuaimi et al. (2015), Bawany and Shamsi (2015), Khan et al. (2013), Li et al. (2015)
Data storage	Data storage presents a significant challenge within Smart City ecosystems, necessitating the exploration of various solutions. While data storage could be considered a subcomponent of data management, its critical importance warrants its independent consideration.	Bawany and Shamsi (2015), Hashem et al. (2016), Li et al. (2015)
Data processing	Big data analytics is the preeminent application domain for big data within Smart City contexts. This field encompasses real-time functionalities, demonstrably exemplified by its role in optimizing transportation networks.	Hashem et al. (2016), Li et al. (2015), Pelton and Singh (2018)
Data protection	The collected and processed data may encompass sensitive or private information, partly or entirely. Consequently, implementing robust security measures across all facets of Smart City systems becomes paramount.	Al Nuaimi et al. (2015), Braun et al. (2018), Lim et al. (2018)

Table 2
Summary of the Smart City concept areas (source: own elaboration).

Dimension	**Source**
Smart People	Giffinger et al. (2007), Hashem et al. (2016), Khan et al. (2013), Lai and Cole (2023), Sharifi (2019)
Smart Life	Giffinger et al. (2007), Lai and Cole (2023), Sharifi (2019)
Smart Industry	Giffinger et al. (2007), Hashem et al. (2016), Khan et al. (2013), Lai and Cole (2023), Sharifi (2019)
Smart Network	Hashem et al. (2016)
Smart Healthcare	Hashem et al. (2016), IMD Business School (2023)
Smart Transport	Berrone a Ricard (2022), Giffinger et al. (2007), Hashem et al. (2016), IMD Business School (2023), Khan et al. (2013), Sharifi (2019)
Smart Government	Berrone a Ricard (2022), Giffinger et al. (2007), IMD Business School (2023), Khan et al. (2013), Lai and Cole (2023), Sharifi (2019)
Smart Environment	Berrone and Ricard (2022), Giffinger et al. (2007), Khan et al. (2013), Lai a Cole (2023), Sharifi (2019)
Smart Security	Hashem et al. (2016), IMD Business School (2023), Lai and Cole (2023)

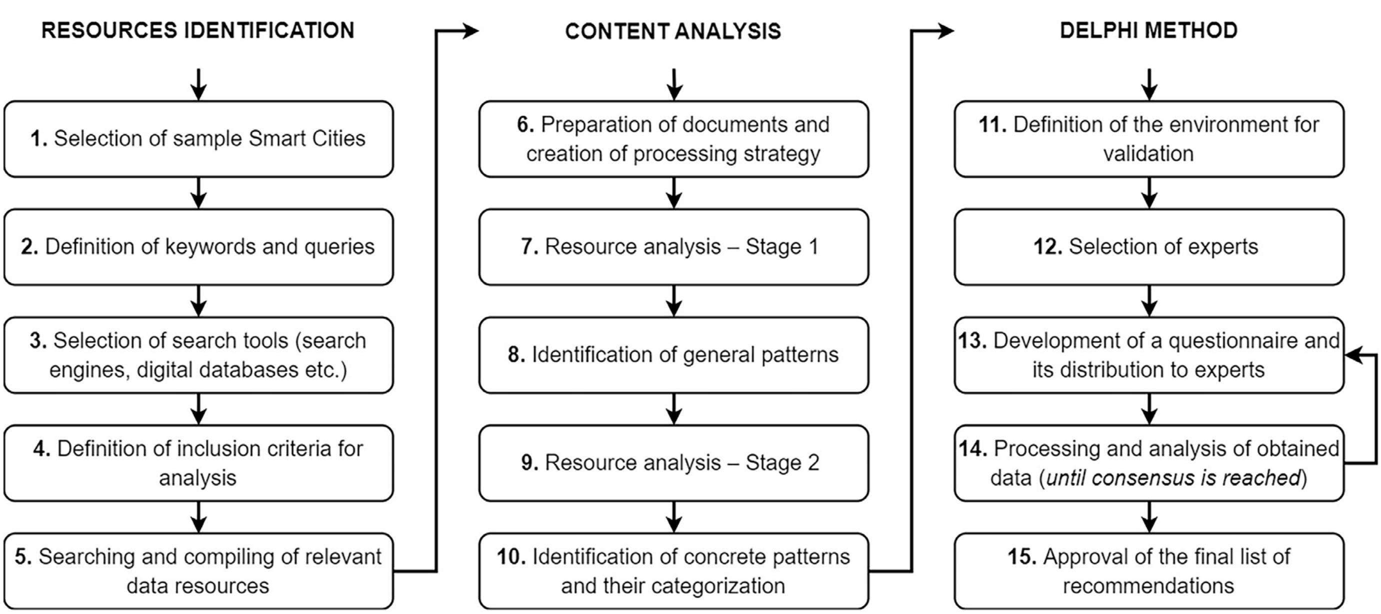

Fig. 1. Phases and steps of the research methodology (source: own elaboration).

the 2023 Smart City Index (SCI) report was chosen for this study. The SCI offers a comprehensive evaluation of 141 cities worldwide, providing a global perspective on Smart City trends and challenges. This index encompasses a diverse range of urban contexts and levels of Smart City development. Moreover, the 2023 update of the SCI methodology reflects the increasing importance of sustainability in contemporary urban planning. To identify big data and data protection best practices, we selected the top-performing Smart Cities listed in the SCI in which we expected to identify relevant measures.

Specific restrictions were applied to the selection process to ensure the representativeness of the sample and the generalizability of the findings, thereby mitigating the influence of localized factors. Four regions (Europe, North America, Asia, and Others) were identified as city selection sources. Considering the 2023 SCI rank, a maximum of ten cities

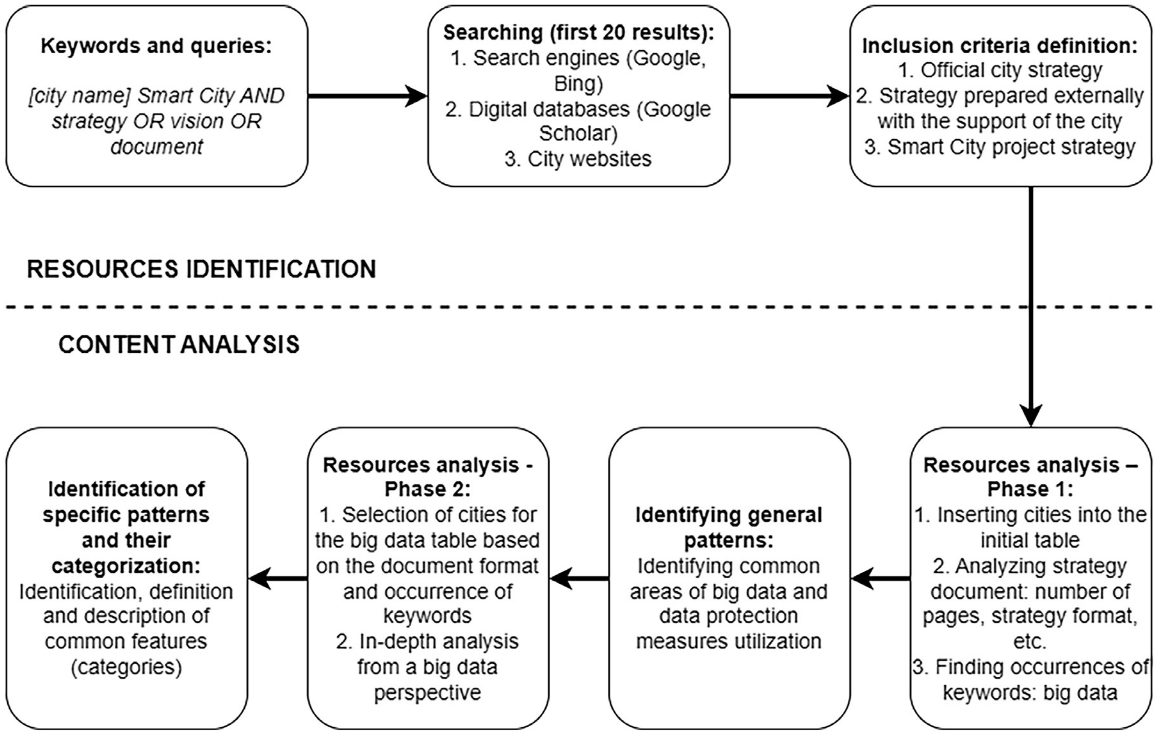

Fig. 2. The steps of searching for strategies and their analysis (source: own elaboration).

were selected from each region. Additionally, within Europe, Asia, and Others, only one city per country was included. This approach yielded a sample size of 40 cities, drawn from the 141 cities listed in the 2023 SCI report. Although including cities from underrepresented regions like South America and Africa would strengthen the global relevance of our study, the unavailability of reliable data for these regions is a significant challenge. Nevertheless, these regions can be included in future research where we can apply our framework to assess the implementation of the respective measures.

A systematic approach for identifying and analyzing Smart City strategies for the 40 selected cities was employed, see Fig. 2. The first step involved defining relevant keywords and queries for search engines, primarily Google and Bing. Additionally, the Google Scholar database and official websites of the selected cities were also utilized. Initially, we conducted searches in English. If no relevant documents or websites were found, we translated our search terms into the official language of the country where the city is located. Inclusion criteria were established to ensure the retrieved documents were relevant to the study. These included official city strategies, strategies developed by external entities with city support, and Smart City project strategies (specific projects focusing on Smart City development). We initially compiled a list of documents and then analyzed each one to determine its relevance to big data and data protection. Throughout the research process, we continuously updated this list. Documents were removed or added based on several factors: (1) low quality, relevance, or usefulness; (2) outdated information or approaches; (3) limited accessibility (restricted access, only example pages, or summary information); and (4) lack of machine processability (locked for changes – search, copy etc.). This dynamic approach ensured that our research remained current and responsive to the evolving landscape of Smart City big data and data protection. Relevant documents that have met these criteria were found for 30 Smart Cities. No strategy was found for eight of the remaining cities; only two cities had a national smart strategy. Also excluded from the content analysis were the strategies for Beijing (unavailable for translation) and Tel Aviv (inaccessible from Europe). The selected Smart Cities and their strategies are listed in Appendix 1.

3.2. Content analysis

The content analysis followed the steps displayed in Fig. 2. To conduct our content analysis, we systematically examined and explored the documents. This involved scrutinizing both the textual content and the structural elements, such as headings and subheadings. We focused on identifying key terms and phrases that were indicative of the relevant themes of big data and data protection. The first step of the content analysis involved the development of an initial table summarizing key characteristics of the 40 Smart Cities and their strategies included in the sample (see Appendix 1). A systematic approach to knowledge acquisition was employed, with detailed notes recorded in Microsoft Excel during the coding process. This table used several indicators derived from the SCI 2023 report: city index ranking, city Human Development Index (HDI), country HDI, and population. Additional columns within the table captured data derived from the analysis of the selected documents: (1) strategy format (text or presentation), (2) language (primary language of the strategy and/or other languages used), (3) strategy title (focusing on the main topic covered by the strategy), (4) year of publication, (5) validity of the strategy, (6) number of pages (length of the strategy), and (7) the last three columns captured big data utilization: presence/absence (mention), frequency of mention, and areas discussed.

These characteristics are needed for further analyses, such as cluster analysis, to identify big data utilization patterns, similarities between Smart Cities, and the characteristics that affect them most. The following phase 2 included an in-depth analysis from a big data perspective. Based on the initial table, areas where the Smart City strategies addressed big data were identified and incorporated into a second big data table. Only cities possessing a documented Smart City strategy, available in written text or presentation format, were included. Furthermore, the strategy had to mention the concept of big data within its content explicitly. Considering these criteria, the number of Smart Cities suitable for identifying and categorizing specific patterns based on the data in the second table was reduced to 14.

Table 3 presents the results of phase 2 involving Smart City strategies, which included the topic of big data. The following categories fo-

Table 3
Categorization of specific big data applications for selected Smart Cities (source: own elaboration).

SCI 2023 Rank	City	Country	Strategy format	Transportation	Big data analytics	City governance	Economy, business	Technological mentions	Other big data terms	Other specific areas
4	Copenhagen	Denmark	Presentation	1	0	1	0	0	0	Waste management
6	London	United Kingdom	Text	1	1	0	0	0	0	x
7	Singapore	Singapore	Text	0	1	0	0	0	0	x
10	Stockholm	Sweden	Text	1	1	0	0	1	1	x
11	Hamburg	Germany	Text	0	1	1	0	1	0	x
13	Abu Dhabi	United Arab Emirates	Presentation	0	1	0	0	0	1	x
14	Prague	Czech Republic	Presentation	1	1	0	1	0	0	Tourism, social media, energy
16	Seoul	Korea	Presentation	0	0	1	0	0	0	x
19	Hong Kong	Hong Kong	Text	0	1	0	0	1	0	x
29	Taipei	Taiwan	Text	0	1	1	0	1	0	Water supply management
40	Ottawa	Canada	Text	0	1	0	0	0	0	x
55	Seattle	USA	Text	1	1	0	0	0	1	x
72	Tokyo	Japan	Presentation	1	0	1	1	0	0	Research
102	Jakarta	Indonesia	Presentation	0	1	0	0	1	0	x

cused on specific big data applications were identified: (1) transportation, (2) big data analytics (including general data analytics), (3) city governance (city planning, local government projects, urban development, policymaking), (4) economy, business, (5) technological mentions (including IoT, real-time data, AI, machine learning), (6) other big data terms (including management, engines, platforms, etc.), and (7) other specific areas (big data utilization areas mentioned for only a single city).

3.3. Delphi method

The Delphi method is a structured technique for obtaining consensus among experts on a particular topic. It is a multi-round process that involves successive iterations of questionnaires, feedback, and revisions. The goal is to reach a convergence of expert opinions through a series of anonymous rounds, minimizing the influence of groupthink and social pressures (Crisp et al., 1997). Lilja et al. (2011) emphasize the flexibility of the Delphi method, allowing it to be adapted to different contexts and research needs. Its ability to facilitate consensus makes it a valuable tool in fields like technology management and policy development, where input from multiple experts is necessary to address future challenges.

The Delphi method was employed to validate the identified recommendations through a three-round iterative process. The initial phase of the Delphi method involved establishing the context for expert evaluation. Given the selection of an online format with an electronic questionnaire, a detailed description of a model Smart City scenario was provided to the experts. This scenario included potential limitations the Smart City might face. The sample of cities used for reference was primarily composed of successful cities with ample resources based on the SCI 2023 ranking. To reflect this context, the model Smart City was characterized by several key features: interest in utilizing big data for various purposes, a focus on ensuring big data security, and an assumption of sufficient technological, human, and financial resources, along with the necessary infrastructure to support Smart City strategies.

Following the definition of the validation environment, expert recruitment commenced. Selection criteria prioritized expertise in the key areas relevant to the study: Smart Cities, information systems and processes, data requirements, and data security and protection. Notably, individual experts did not require in-depth knowledge across all domains. Seven experts with complementary skillsets were identified and invited to participate, all accepting the invitation. Their qualifications are summarized in Table 4. The diversity of big data, data protection, and urban governance experiences within the group ensures a comprehensive range of perspectives, enhancing the depth and quality of the Delphi process. The number of experts is consistent with Mullen's research from 2003. She found that seven is a good minimum number of experts for a study like this, and that the accuracy of the results decreases significantly if there are fewer experts. While more experts can improve accuracy, the improvement gets smaller and smaller as the panel size grows beyond seven (Mullen, 2003).

4. Data analysis and results

Data analysis employed an approach utilizing both Excel and Python. The Excel environment facilitated the analysis of both tables (Appendix 1 and Table 3), including counts, averages, missing values, replacements, and other data preprocessing steps needed to prepare data suitable for the Python-based analysis. Data preparation of the initial table involved removing several columns deemed unnecessary for the Python-based analysis, i.e., *Country* and *Applications of big data*. The *Validity of the strategy* column was also removed due to numerous missing values. Decimal values were converted from comma to dot format. Fo*ur unspecified* values were replaced with empty cells for the Year of publication column. The missing values were then replaced using the average value of 2018. Fourteen missing values in the *No of mentions of big data*

Table 4
List of experts and their expertise (source: own elaboration).

ID	Country	Job position	Years of expertise	Professional expertise
1	Czech Republic	Assistant professor, data analyst, e-government consultant	15	Expert in Smart City data governance. Experienced in big data, open data, data security, and the impact of technology on public sector governance.
2	Germany	Civil servant	12	Information systems and public sector expert. Experienced in systems analysis, design, and implementation, with a focus on public sector applications.
3	Austria	Professor, project manager	25	Data management and project management expert. Experienced in data modeling, data management, data security, and delivering complex projects in the public sector.
4	Poland	Researcher	5	Smart City innovation and citizen engagement expert. Experienced in leveraging technology to enhance citizen participation and urban governance.
5	Slovakia	IT consultant and data specialist	20+	Information systems and cybersecurity expert. Experienced in data management, data exchange, and security assessments in both public and private sectors.
6	Switzerland	PhD student	3	E-government and Smart City expert. Experienced in applying privacy-enhancing technologies to protect personal data in urban settings.
7	Czech Republic	Assistant professor	25	Cybersecurity and citizen participation expert. Experienced in ICT systems, computer networks, and cybersecurity, with a focus on citizen engagement in digital environments.

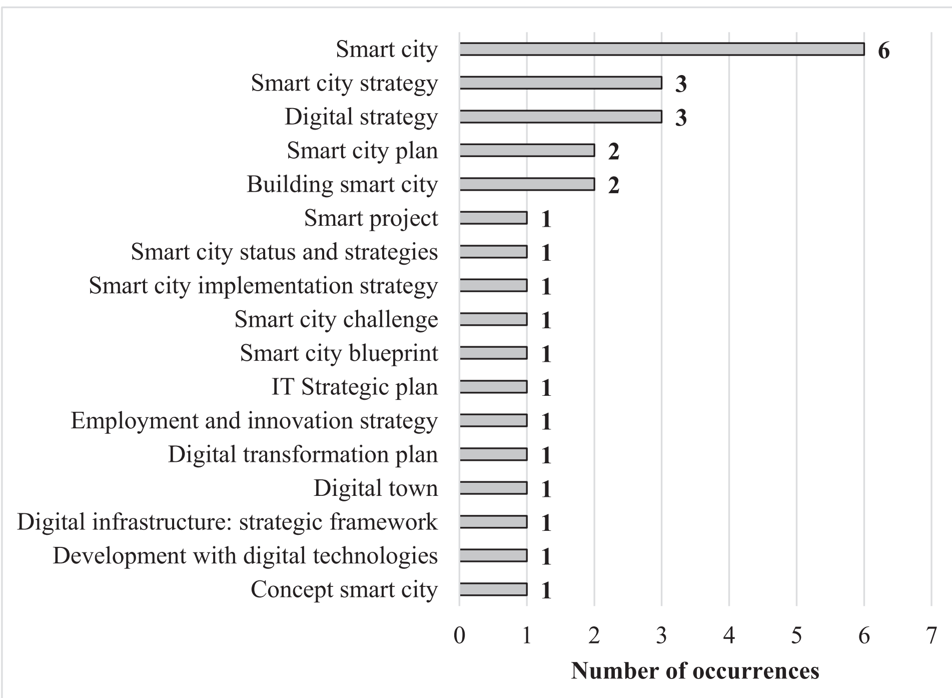

Fig. 3. Analysis of strategies' titles (source: own elaboration).

columns were replaced with 0. The file format was also converted from XLSX to CSV.

The data underwent a text-to-numeric conversion process for several columns, i.e., *Strategy format, Does it include big data?*, and *Region.* This process involved assigning binary codes for the first two columns: *Text* was converted to 0, while *Presentation* was converted to 1. Similarly, for *Does it include big data?, No* was coded as 0, and *Yes* was coded as 1. Furthermore, the *Region* column underwent categorical encoding. *Europe* was assigned a value of 0, *Asia* a value of 1, *Others* a value of 2, and *North America* a value of 3. Following this initial conversion, data normalization and/or standardization techniques were employed to ensure consistent scaling across all attributes, facilitating subsequent analysis.

The first phase of the Delphi method involved the development of a structured questionnaire for the first round. This instrument served two key functions: 1) defining the intended outcomes of the expert evaluation process (validation objectives) – the questionnaire explicitly outlined the desired results of the expert evaluation, which ensured that expert responses were focused and directly addressed the research objectives, and 2) providing instructions for the participating experts to facilitate expert understanding of the evaluation process, which included the rating scale (0-None, 1-Low, 2-Moderate, 3-High) used to assess their level of agreement with each recommendation presented.

4.1. Big data context in smart cities strategies

It was found that only 75 %, i.e., 30, of Smart Cities have a strategy document. However, the two strategies cannot be further analyzed since the strategy of Beijing cannot be translated into English, and the strategy of Tel Aviv is inaccessible from Europe. Thus, subsequent analyses focus on 28 cities with documented strategies.

Strategy title was one of the attributes examined to identify any recurring patterns or define keywords among the titles. An overview of these findings is presented in Fig. 3. The most common title pattern included the term Smart City (encompassing variations such as Smart Seattle and Medellín: Smart City), appearing six times. Another pattern, observed three times each, was the use of the titles Smart City Strategy and Digital Strategy. While not explicitly incorporating Smart City in its title, the latter category exhibited characteristics typically associated with Smart City strategies. Titles appearing twice included Smart City Plan and Building Smart City. The remaining titles were unique, appearing only once.

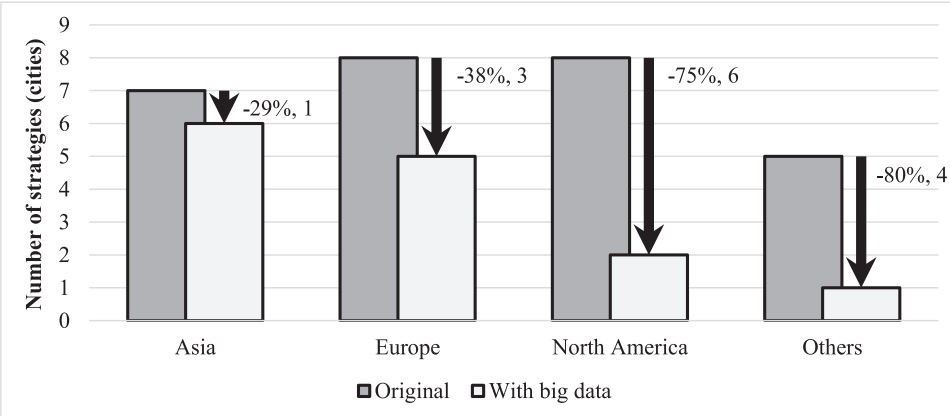

Fig. 4. Comparison of strategies that include the term big data according to regions (source: own elaboration).

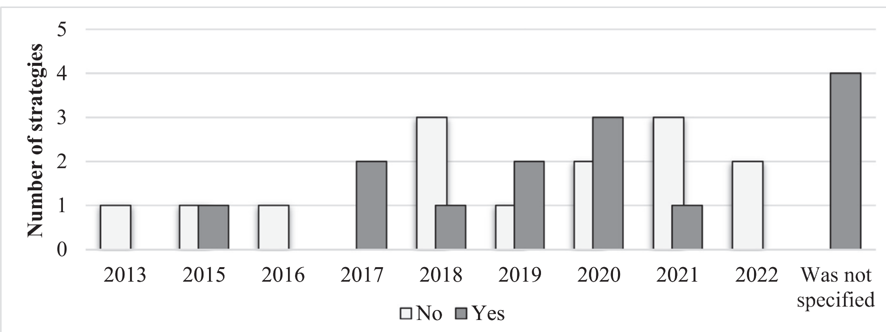

Fig. 5. Overview of publication years of strategies (source: own elaboration).

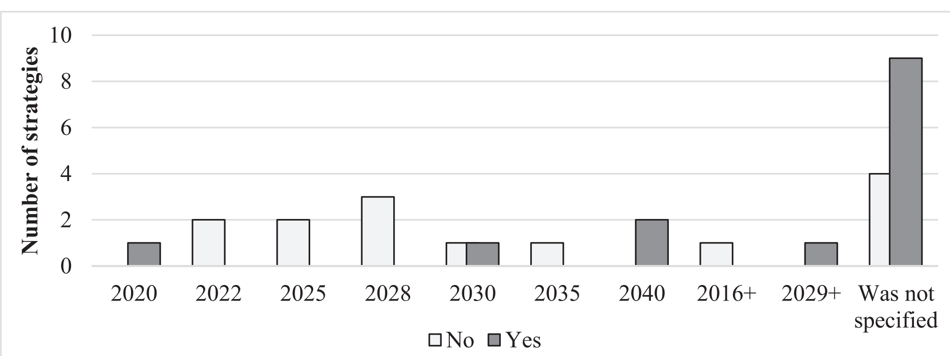

Fig. 6. Overview of years of strategies' validity (source: own elaboration).

We found that only 19 out of 28 strategies incorporate the term Smart in their titles. The remaining documents utilize diverse titles, predominantly featuring the concept of digitalization, often in the form of the adjective digital. Despite these variations in title, the strategies exhibited, at least partially, the characteristics of Smart City strategies. Then, the keyword *big data* occurrences within the analyzed documents were explored. Exactly 50 %, i.e., 14 cities mention *big data* in their strategic documents.

While the study initially encompassed 40 cities, with a quota of 10 cities per region, the selection for analysis was narrowed down due to two factors: the absence of a city-specific strategy and the occurrence of the keyword big data. Fig. 4 presents the changes in regional proportions for the number of strategies that include the term big data. The original number of 28 cities was reduced to 14 with big data. Europe and North America each have eight cities in the narrowed-down selection, Asia has one fewer, and the Others has only five, half its original number. As previously mentioned, the two cities address their strategies solely within the context of their respective national strategies. The right column of Fig. 4 displays only cities within each region that mention big data. The figure further reveals that the concept of big data is most frequently addressed by cities in Asia, followed by Europe. The arrows in the graph represent the changes in city counts upon focusing solely on strategies that mention big data. Furthermore, it can be inferred that big data is the least addressed in the Others and North America, while Asia has the highest proportion of cities that mention big data.

Fig. 5 illustrates the distribution of strategy publication years. The figure reveals that the highest number of strategies were published in 2020 (5 strategies), followed by 2018 (4 strategies), 2021 (4 strategies), and 2019 (3 strategies). The number of strategies published in subsequent years decreases, with two or fewer strategies published yearly. The publication year of their strategies could not be determined for four cities. The right columns titled Yes focus on the publication years of strategies mentioning big data. The earliest publication year is 2015, which is represented by Copenhagen's strategy. Most documents (3 strategies) were published in 2020, with one additional strategy's publication year remaining uncertain. The most recent document, belonging to Stockholm, was published in 2021. For comparison, the left columns present the distribution of city publication years and their strategies

Table 5
Document format (source: own elaboration).

Document format	All cities with strategy	Cities with big data	Cities without big data
Presentation	12; 43 %	6; 43 %	6; 43 %
Text	16; 57 %	8; 57 %	8; 57 %

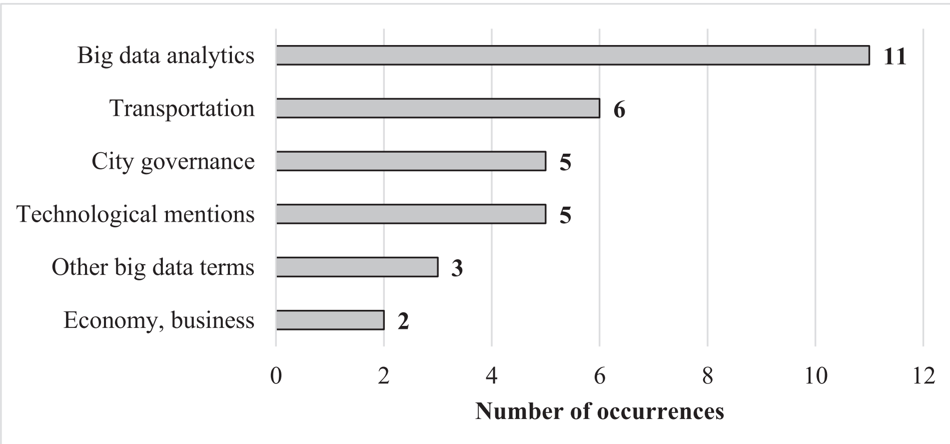

Fig. 7. Counts of specific big data applications (source: own elaboration).

that do not mention big data. The highest number of strategies were published in 2018 and 2021. The most recent documents, published in 2022, belong to Toronto and Reykjavík. The oldest document that does not mention big data, dating back to 2013, belongs to Vancouver. Comparing these findings, we can conclude that big data are not exclusively associated with newer strategies. In contrast to strategies mentioning big data, the publication year of all strategies not mentioning big data could be determined.

Building upon the previous analysis of publication years, the next step of the analysis delves into a comparison of the validity periods (range of years) for which the strategic documents were created. Fig. 6 provides an overview of the validity of strategies over the years. Notably, 13 documents do not have any specified validity period. Based on the derived data, three cities have already exceeded the validity period of their documents. One document, belonging to London, had a validity period extending until 2020. Two additional documents about Canberra and Bangkok have validity periods set for 2022. Furthermore, two documents have a specific year and beyond as their validity period, represented on the figure as *year+*. These are Vancouver (2016 and beyond) and Abu Dhabi (2029 and beyond). The document with the longest prolonged validity period until 2040 belongs to Tokyo.

The right column of Fig. 6 focuses exclusively on cities that incorporate the term big data in their strategies. Like the overall distribution, most cities do not have a specified document validity period. One document belonging to London has already exceeded its validity period. For comparison, the left column of the figure presents the validity periods for cities without big data mentions. The upper validity limit in this category is 2035, where the Smart City strategy is integrated into another strategic document with a validity period extending until that year. This is the case for the Swiss city of Zurich.

Table 5 examines the format in which city strategies are presented, distinguishing between two primary approaches: presentation style and written-text style. The categorization was based on the overall layout of the document and the proportion of text compared to visual elements such as images or diagrams. Table 5 reveals that most strategies (57 %, or 16 in absolute terms) are presented in a written-text format, while 12 strategies adopt a more presentation-style approach. Notably, this distribution remains consistent when comparing cities with and without big data.

Fig. 7 visually represents the thematic distribution of big data applications across the analyzed cities. Big data analytics emerged as the most prevalent theme, with 11 cities mentioning its utilization. Transportation ranked as the second most frequently discussed domain, with six cities incorporating big data in this area. Five cities mentioned city governance, encompassing urban planning, local government projects, urban development, and policymaking. Technological mentions, including concepts like IoT, real-time data processing, AI, and machine learning, were also highlighted by five cities. Other big data terms, such as big data management, big data engines, big data platforms, or big data systems, were mentioned by three cities. The thematic overview concludes with the area of economics and business, which was mentioned in the strategies of the two cities.

4.2. Descriptive and cluster analyses

Descriptive statistics for the initial table are presented first in Table 6. It is important to note that the *Number of big data mentions* attribute is non-representative due to records with zero values. However, this attribute also appears in the descriptive statistics for the big data table, which is provided later. The most interesting results probably concern the attribute of the year of publication or the number of pages, which is, however, influenced by the fact that the strategies are divided into presentations or written texts.

Table 7 presents the descriptive statistics for the big data table. The attribute *Number of big data mentions* indicates an average of 5.5 mentions per strategy, with a median value of 4.5. The highest occurrence is 20, corresponding to the city of Prague. Comparing the index rankings with the previous table reveals lower values for cities addressing big data in their Smart City strategies. It suggests that cities with a higher ranking in the index are more likely to deal with big data challenges.

Then, the correlation analysis was performed for the initial table. However, there is no significant relation between attributes. Table 8 presents correlation coefficients for the big data table. A moderately positive correlation is observed between the *Number of big data mentions* and *Transportation* and *Economy, business* attributes. This finding suggests that if big data are mentioned in the strategy, then these are related to the transportation economy and business areas. This table also reveals several negative correlations between attributes. However,

Table 6
Outputs of descriptive statistics for 28 cities in the initial table (source: own elaboration).

Stat.	SCI Ranking	City HDI	Population (2023)	Country HDI	Year of publication of strategy	Number of pages	Number of big data mentions
COUNT	28	28	28	28	28	28	28
MEAN	40.36	0.93	5,002,037.61	0.91	2018.64	43.61	2.75
STD	35.74	0.06	7,276,888.12	0.06	2.18	33.17	4.46
MIN	1	0.76	137,618	0.71	2013	12	0
25 %	12.50	0.93	1,133,116.5	0.91	2018	24	0
50 %	27.50	0.94	2,197,972.5	0.93	2018.50	32.50	0.50
75 %	56.50	0.96	7,735,631	0.94	2020	51	4.25
MAX	118	0.99	37,194,105	0.96	2022	167	20

Table 7
Outputs of descriptive statistics for 14 cities in the big data table (source: own elaboration).

Statistics	SCI Ranking	Number of big data mentions	Transportation	Big data analytics	City governance	Economy, business	Technological mentions	Other specific areas
COUNT	14	14	14	14	14	14	14	14
MEAN	28,43	5,50	0,43	0,79	0,36	0,14	0,36	0,21
STD	29,15	5,00	0,51	0,43	0,50	0,36	0,50	0,43
MIN	4	1	0	0	0	0	0	0
25 %	10,25	3	0	1	0	0	0	0
50 %	15	4,50	0	1	0	0	0	0
75 %	37,25	5	1	1	1	0	1	0
MAX	102	20	1	1	1	1	1	1

Table 8
Correlation outputs for 14 cities in the big data table (source: own elaboration).

	SCI Ranking	Region	Document format	Number of big data mentions	Transportation	Big data analytics	City governance	Economy, business	Technological mentions	Other specific areas
SCI Ranking	1	0.427	0.301	-0.167	-0.028	-0.053	-0.014	0.273	0.150	0.022
Region		1	-0.017	-0.410	-0.318	0.145	-0.294	-0.117	-0.290	0.310
Document format			1	0.311	0.137	-0.658	0.244	0.444	-0.323	0.134
Number of big data mentions				1	0.619	-0.098	0.055	0.552	-0.018	0.212
Transportation					1	-0.252	-0.083	0.406	-0.336	0.282
Big data analytics						1	-0.592	-0.321	0.347	0.166
City governance							1	0.129	0.211	-0.309
Economy, business								1	-0.225	-0.085
Technological mentions									1	-0.085
Other specific areas										1

given the small sample size of 14 cities, some of these relationships may require further investigation in future research.

Fig. 8 presents the clustering results for the initial table. Three clusters emerged as the optimal choice for effective analysis. With four clusters, the clusters overlapped significantly. The figure shows that only the blue cluster (Cluster 0) represents a distinct grouping. The remaining clusters contain cities scattered away from the cluster centers.

Table 9 presents the distribution of cities into the identified clusters. Cluster 0 is the most populous, while Cluster 2 is the least populous. The columns *Does it include big data?* and *Number of big data mentions* did not significantly influence the cluster distribution. Cluster 0 comprises all European cities and all but three Asian cities. Cluster 1 includes all North American cities and two additional cities from the Others: Canberra and Abu Dhabi. Notably, while language was not explicitly incorporated into the clustering process, all cities in Cluster 2 share English as their official language, distinguishing them from the two non-English speaking cities in Cluster 2. Apart from language, the index ranking and City HDI also played a role in cluster formation, with a few exceptions. Cluster 0 had an average index ranking of 13.17 (median 12.5), Cluster 1 had an average index ranking of 38.6 (median 45), and Cluster 2 had an average index ranking of 97.67 (median 95.5). In terms of City HDI, Cluster 0 had an average of 0.959 (median 0.961), Cluster 1 had an average of 0.940 (median 0.941), and Cluster 2 had an average of 0.843 (median 0.849). However, Clusters 1 and 2 exhibited outliers that

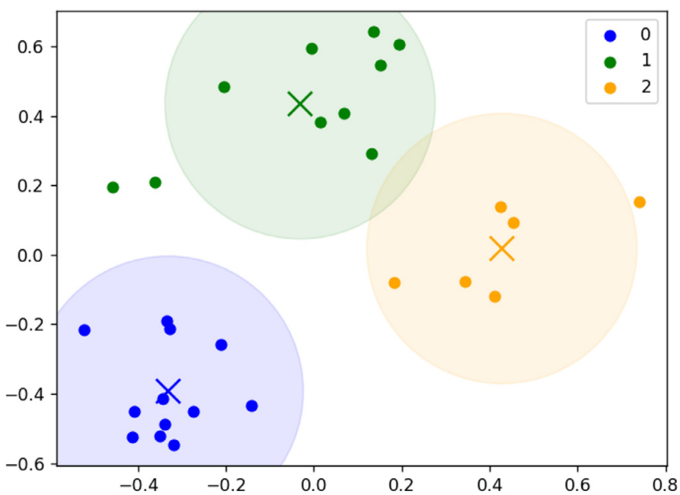

Fig. 8. Clustering of 28 cities in the initial table (source: own elaboration).

Table 9
Classification into clusters (for the initial table) (source: own elaboration).

Cluster – 0	Cluster – 1	Cluster – 2
Zurich	Canberra	Tokyo
Copenhagen	Abu Dhabi	Bangkok
London	New York	Kuala Lumpur
Singapore	Ottawa	Jakarta
Stockholm	Vancouver	Nicosia
Hamburg	Toronto	Medellin
Prague	Los Angeles	
Amsterdam	Denver	
Seoul	Seattle	
Hong Kong	Chicago	
Reykjavik		
Taipei City		

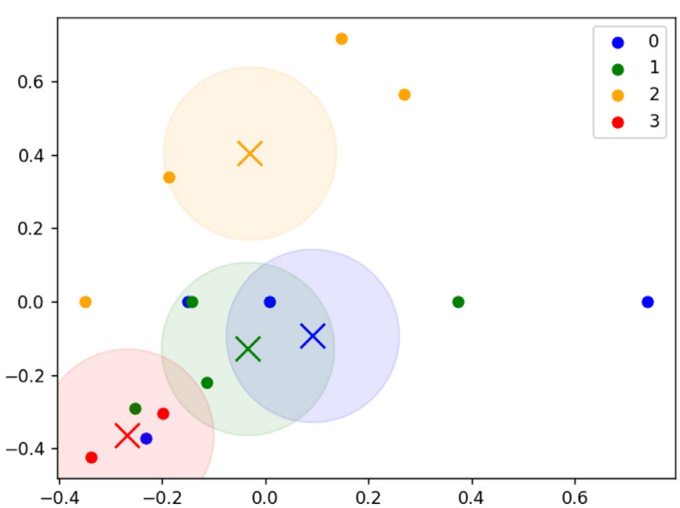

Fig. 9. Clustering of 14 cities in the big data table (source: own elaboration).

deviated substantially from these averages. Canberra was an outlier in Cluster 1, while Tokyo was an outlier in Cluster 2.

Fig. 9 presents the clustering results for the big data table. The clustering outcomes for this table are suboptimal, as clusters overlapped significantly across different numbers of clusters, and some cases were outliers. Four clusters emerged as the optimal solution, as five or more clusters resulted in a single city being assigned to a cluster. As illustrated in the figure, the cities in the red cluster (Cluster 3) exhibit the closest proximity. A pattern can also be observed in Cluster 2, which primarily consists of cities in the upper left corner. Clusters 0 and 1 overlap with one city from each cluster, also located in the red cluster (Cluster 3).

Table 10
Classification into clusters (for the big data table) (source: own elaboration).

Cluster – 0	Cluster – 1	Cluster – 2	Cluster – 3
Hamburg	Copenhagen	Singapore	London
Hong Kong	Prague	Abu Dhabi	Stockholm
Taipei City	Seoul	Ottawa	
Jakarta	Tokyo	Seattle	

Table 10 presents the cluster assignments. As observed in the figure above, the cities in Cluster 3 exhibit the closest proximity, namely London and Stockholm, which were also grouped in the clustering of the initial table. Cluster 2 comprises cities within a specific region, free from overlapping with other clusters. This cluster includes Singapore, Abu Dhabi, Ottawa, and Seattle, which, except for Singapore, were also grouped together in the initial clustering. Unlike the initial clustering, the index ranking did not play a significant role in cluster formation for this table.

For Cluster 0, the determining attribute is the *Technological mentions* attribute, which consistently holds a value of 0 for all cities in this cluster and the adoption of *Big data analytics*. Notably, cities in this cluster do not incorporate big data applications in *Transportation* and *Economy, business*. Cluster 1 is characterized by the application of big data in *Transportation*. Notably, with one exception, cities in this cluster do not mention *Big data analytics*. Additionally, all cities in this cluster have their strategy presented as a presentation. The primary distinguishing attribute of Cluster 2 is the adoption of *Big data analytics*, while the remaining areas are generally not utilized, except Seattle. Cluster 3 comprises two European cities with written strategic documents and similar index rankings. Both cities utilize big data in the area of *Transportation* and mention *Big data analytics*. However, they do not employ big data in *City governance* and *Economy, business*.

4.3. Data protection in strategies

Data protection receives limited attention within the analyzed Smart City strategies. This subsection explores this issue by examining data protection mentions in the strategies of cities utilizing big data. Copenhagen's strategy makes no mention of data protection. London references data protection in the context of the Oyster card, stating that data protection standards are adhered to without specifying the specific standards employed. Singapore, in contrast, dedicates more substantial attention to data protection. Stockholm also devotes a significant portion of its strategy to this topic. Hamburg addresses data protection in two paragraphs. The first instance employs the term in a general sense, emphasizing the need for high data protection. The second paragraph focuses on cloud technology utilization and encourages businesses and organizations to engage in data protection on their platforms actively. The remaining cities dealing with big data, namely Abu Dhabi, Prague, Seoul, Hong Kong, Taipei, Ottawa, Seattle, Tokyo, and Jakarta, do not address data protection in their strategies.

The strategic document for Stockholm dedicates a chapter to this topic. The first paragraph acknowledges the fact that a vast amount of information is generated within the Smart City framework. It highlights various systems and expertise for information management and security. However, it emphasizes ensuring all new solutions adhere to existing information security guidelines, including current laws. The subsequent paragraph focuses on reviewing information security and classification guidelines, with particular emphasis, according to the strategy, on handling data that integrates information from multiple sources. It further emphasizes the need to develop standard working methods and procedures, along with a shared information structure, to ensure consistency in the collection of information. The third paragraph addresses

Table 11
The final list of recommendations for big data and data protection measures (source: own elaboration).

Label	Recommendation	Average	Standard deviation
D 1	BIG DATA	2.286	0.452
D 1.1	Big data should be used in these areas of Smart Cities…	2.286	0.452
D 1.1.1	Smart people – areas such as social interactions, education, public life, etc.	2.714	0.452
D 1.1.2	Smart life – quality of life areas like culture, housing, tourism, etc.	2.714	0.452
D 1.1.3	Smart industry/economy – supporting businesses and preparing conditions for growth (such as making a platform/sharing some data with them or creating/supporting infrastructure – optical cables, 5 G, etc.)	2.857	0.350
D 1.1.4	Smart grid – connection areas, such as electricity, heat, water, etc.	2.429	0.495
D 1.1.5	Smart healthcare – supporting the functioning of hospitals as well as connecting/supporting private sector health services	2.429	0.495
D 1.1.6	Smart mobility/transport – includes public transport and shared transportation services (car, bikes, etc.)	2.429	0.495
D 1.1.7	Smart traffic – traffic control (traffic lights on junctions, pedestrian crosswalks, etc.), smart parking and providing information about traffic	3.000	0.000
D 1.1.8	Smart governance – areas related to city management, such as urban planning, dealing with civil matters, voting, providing information, etc.	2.143	0.350
D 1.1.9	Smart environment – natural conditions (climate, greenery), pollution, resource management, etc.	2.286	0.452
D 1.1.10	Smart security – areas of creating a city more secure, such as the help of CCTV cameras, smart police, crime maps, etc.	2.571	0.495
D 1.2	The main requirements for big data systems in Smart Cities should include…	2.714	0.452
D 1.2.1	Scalable data infrastructure – must accommodate growing volumes of data	2.714	0.452
D 1.2.2	Resilient data infrastructure – must withstand various fluctuations in the volume of data flow.	2.571	0.495
D 1.2.3	Data management – must support acquiring, processing, and combining data from types of sources	2.571	0.495
D 1.2.4	Data storage for sensitive data – must have a particular protected place for sensitive data	2.857	0.350
D 1.2.5	Data storage for other big data – creating a platform for other kinds of big data	2.143	0.350
D 1.2.6	Computing infrastructure – as one of the most used areas of big data application is big data analytics, and in some cases done in real-time, it must be considered in building computing infrastructure	2.714	0.452
D 1.3	Smart Cities should include these NIST big data general requirements…	2.143	0.350
D 1.3.1	Transformation provider requirements	2.143	0.350
D 1.3.1.1	Needs to support diversified compute-intensive, statistical, and graph analytic processing and machine learning techniques	2.143	0.350
D 1.3.1.2	Needs to support batch and real-time analytic processing	2.286	0.452
D 1.3.1.3	Needs to support processing large, diversified data content and modeling	2.143	0.350
D 1.3.1.4	It must support processing data in motion (streaming, fetching new content, tracking, etc.).	2.143	0.350
D 1.3.2	Capability provider requirements	2.143	0.350
D 1.3.2.1	Needs to support legacy and advanced computing platforms	2.000	0.000
D 1.3.2.2	Needs to support legacy and advanced distributed computing clusters, co-processors, input/output processing (infrastructure)	2.000	0.535
D 1.3.2.3	Needs to support elastic data transmission (networking)	2.143	0.350
D 1.3.2.4	Needs to support legacy, large, and advanced distributed data storage (storage)	2.286	0.452
D 1.3.2.5	Needs to support legacy and advanced executable programming: applications, tools, utilities, and libraries (software)	2.000	0.535
D 1.3.3	Security and privacy requirements	2.429	0.495
D 1.3.3.1	Needs to protect and preserve the security and privacy of sensitive data	2.571	0.495
D 1.3.3.2	Needs to support sandbox, access control, and multilevel, policy-driven authentication on protected data	2.286	0.700
D 1.3.4	Life cycle management requirements	2.286	0.452
D 1.3.4.1	Needs to support data quality curation, including preprocessing, data clustering, classification, reduction, and format transformation	2.143	0.350
D 1.3.4.2	Needs to support dynamic updates on data, user profiles, and links	2.143	0.639
D 1.3.4.3	Needs to support data life cycle and long-term preservation policy, including data provenance	2.286	0.452
D 1.3.4.4	Needs to support data validation	2.286	0.452
D 1.3.4.5	Needs to support human annotation for data validation	2.000	0.000
D 1.3.4.6	Needs to support the prevention of data loss or corruption	2.429	0.495
D 1.3.4.7	Needs to support persistent identifier and data traceability	2.286	0.452
D 1.3.4.8	Needs to support standardizing, aggregating, and normalizing data from disparate sources	2.143	0.350
D 1.4	Other technological aspects related to big data in Smart Cities should include…	2.286	0.452
D 1.4.1	IoT	2.571	0.495
D 1.4.2	Traditional AI	2.429	0.495
D 1.4.3	Generative AI	2.286	0.452
D 1.4.4	Machine Learning	2.286	0.452
D 1.4.5	Real-time (transfer, analysis, etc.)	2.143	0.350
D 1.4.6	Cloud storage – only those forms of cloud deployment that meet the security requirements of the public sector	2.000	0.000
D 1.4.7	Cloud computing – only those forms of cloud deployment that meet the security requirements of the public sector	2.143	0.639
D 1.4.8	Big data analytics	2.286	0.452
D 1.4.9	Data lakes	2.000	0.535
D 1.4.10	NoSQL databases	2.143	0.639
D 2	DATA PROTECTION (all data in Smart Cities)	2.857	0.350
D 2.1	Smart City strategies should…	3.000	0.000
D 2.1.1	Be transparent about data protection – show detailed solutions of how data are protected	2.857	0.350
D 2.1.2	Refer to other standards and laws connected to data protection, such as following GDPR, etc.	2.714	0.452
D 2.1.3	Refer to other documents that deal with data protection in the city/country	2.286	0.452
D 2.1.4	Have separate document/strategy for data protection	2.857	0.350
D 2.2	Data protection in Smart City should include…	2.571	0.495
D 2.2.1	Expert or team of experts in data protection that would supervise smart city projects	3.000	0.000
D 2.2.2	Have (or work with) committee focused on data protection/cybersecurity	2.286	0.452
D 2.2.3	Using any existing standard of framework (e.g. NIST big data)	2.429	0.495
D 2.2.4	Educating (like training course or manual) employees of Smart City in questions of data protection	2.857	0.350
D 2.2.5	Educating (like training course or manual) citizens in Smart City on questions of data protection	2.286	0.452
D 2.2.6	Releasing regular control reports (annual) in the question of data protection	2.000	0.535

the allocation of responsibility for the ownership of information collected and utilized. It also addresses security concerns related to communication, which will primarily occur wirelessly and across various locations within the city. The final paragraph delves into the issue of new developments, where information processing must balance user benefits with personal integrity. Privacy considerations should be integrated into all new solutions developed within the Smart City framework. To address this issue, the city of Stockholm intends to establish a council for privacy and ethical issues related to new digital solutions, which would guide this area. The structure of this council, including its members and mandate, should be defined as part of the priority projects (Stockholms stad, 2021).

Data protection is also addressed within Singapore's Smart Nation strategy. It references the Singapore Cybersecurity Strategy of 2016, which establishes priorities for building resilient infrastructure and creating a more secure cyberspace. Cybersecurity plans have also been developed for 11 critical information infrastructure sectors. It highlights the collaboration between Smart Nation projects and the Smart Nation Cybersecurity Committee to ensure security from the outset, incorporating cybersecurity considerations into design, implementation, and security policy guidance. The strategy also addresses the prevention of data fusion by specific data protection centers. The final mention of data protection relates to international connectivity (Smart Nation & Digital Government Office, 2018).

4.4. A list of recommendations

This section concentrates on validating the findings from the preceding sections and transforming them into recommendations related to big data and data protection. These recommendations serve as guidelines for Smart City projects, providing a foundation for their endeavors in developing and implementing measures related to big data and data protection. The study employed three rounds of the Delphi method to reach a consensus among experts.

For the first round of the Delphi method, for which 2 wk were allocated, the experts evaluated 29 recommendations categorized into two sections. The first section focused on big data utilization within Smart City contexts. It delved into potential application areas for big data alongside essential system requirements and other relevant technological aspects. The second section explored data protection strategies. Here, the questionnaire investigated recommended approaches for cities to integrate data protection within their strategies while simultaneously examining potential data protection mechanisms that Smart Cities could adopt. The returned questionnaires from the first round were processed and analyzed.

Based on calculations of means and standard deviations, the following recommendations were removed for the second round: *Avoid talking about data protection, Mention data protection, but in a universal way – without details*, and *Create or support trademarks that would make companies in the city trustworthy*. These recommendations had low average scores among experts, at 1 – low. Based on expert feedback, new recommendations were also added. AI was divided into two categories: traditional AI and generative AI. Cloud was divided into cloud computing and cloud storage. The concepts of data lakes and NoSQL databases were added to the same category. Similarly, scalable and resilient infrastructure for big data systems was divided into two recommendations. Data storage was also divided. A new category, including National Institute of Standards and Technology (NIST) recommendations, was added to big data system requirements. However, this also meant removing data protection from the main requirements, as a new category was created. Finally, new recommendations were added in the last section: the use of a standard or framework, training (employees, companies, and citizens), and the release of regular control reports in the context of data protection. A paragraph was added to the introduction to clarify the problem and topic. Finally, some recommendations were further elaborated and specified.

With these modifications, the second round of the Delphi method was conducted, with experts having 1 wk to complete the questionnaires. The returned questionnaires were again processed, and the facilitator prepared a modified questionnaire for the next round, as there was still no consensus among experts after the second round. No new categories were added for the final third round of the questionnaire. However, the least significant recommendations were removed, specifically those with an average below 2. In the case of NIST recommendations in the *Capability provider requirements* category, the recommendation *Needs to support legacy and advanced software packages* was removed. The recommendation *Needs to support multisite archives* was removed in the Life cycle management requirements section. In the case of big data protection, the recommendation *of Educating (like training courses or manual) companies in Smart Cities on data protection* questions was removed.

Table 11 presents the final list of 55 recommendations on which experts reached a consensus in the third round of the Delphi method. Regarding expert agreement with the list, it was rated by four experts as Agree (5th level) and by three experts as Agree Strongly (6th level) on a six-point scale. It results in an overall average rating of 5.429, with 6 being the best positive option. The first column of the table provides a label for clarity. The second column contains the recommendations, and the third column presents the average, with the top 12 values highlighted in green. The standard deviation is presented in the last column. In the areas where big data should be utilized, *Smart traffic* was rated the highest, with all experts rating this recommendation as the highest possible value. It is followed by a *Smart industry/economy* and, with the same average, *Smart people* and a *Smart life*. In the main requirements for big data section, *Data storage for sensitive data* is rated the highest, followed by *Scalable data infrastructure* and *Computing infrastructure*.

The highest-rated recommendation in the Smart City strategies section is to Be transparent about data protection and Have a separate document/strategy for data protection. It is followed by the Refer to other standards and laws connected to data protection recommendations. In the last category, Data protection in Smart City should include…, the highest-rated recommendation is Expert or team of experts in data protection that would supervise Smart City projects with an average of 3 and Educating (e.g., training or manuals) employees of Smart City in questions of data protection.

In terms of individual categories, experts rated the area of data protection as more important than the area of big data. Within the subcategories, the category *Main requirements for big data systems in Smart cities should include*, was rated the highest in the area of big data. Within data protection, the category *Smart City strategies should* address how Smart City strategies should approach the issue of data protection, which was considered key, with an average of 3, meaning that all experts agreed on the highest possible rating.

To sum up, effective use of big data often demands extensive personal data collection, raising significant privacy risks, so robust data protection frameworks are essential to enable trust, a prerequisite for successful big data applications. The challenge lies in striking a balance between big data and data protection. By effectively implementing robust data protection measures, Smart Cities can harness the power of big data while safeguarding individual privacy and security. This requires a collaborative effort between policymakers, technologists, and citizens to develop and implement comprehensive data protection strategies.

5. Discussion and limitations

City size significantly influences Smart City challenges and priorities. Larger cities face complexities like dense populations, high infrastructural demands, and diverse data sources, necessitating advanced analytics (Glaeser et al., 2018). Smaller cities, in contrast, often operate with limited resources, data volume, and infrastructure. Stakeholder diversity further complicates matters, requiring collaboration among governments, private entities, academia, non-profits, and citizens (Viale Pereira et al., 2017). Smart City services must address specific local needs and infrastructures (Lnenicka et al., 2024). Governance models – centralized or decentralized – also shape data management, protection, and utilization strategies (Löfgren & Webster, 2020). Additionally, some city officials may lack comprehensive Smart City strategies, relying instead on incomplete or unsuitable measures (World Economic Forum, 2021). Funding remains a critical barrier, especially for smaller cities. Smart City projects demand substantial investments in technology and expertise, making it challenging for cities with limited budgets to implement secure systems or prioritize long-term, impactful initiatives (Das, 2024; Glaeser et al., 2018). So, recommendations aimed at cities with advanced data infrastructures may not be feasible for mid-sized cities or those in developing regions, where financial and technical constraints are common.

Considering the list of recommendations, we can conclude that the selected experts focused more on the issue of big data security than on the general requirements of big data. It suggests that they are likely more familiar with the topic of security and find it more relevant. In the area of big data, they focused on specific Smart City components, i.e., areas of application of smart services such as Smart Transportation, where the practical applications of big data can be more easily envisioned compared to more technologically framed recommendations. Nevertheless, these findings are also valuable, as they can guide further research to focus on specific services that can benefit from big data and only then address the various technological aspects. For concrete applications in the form of services, one can draw inspiration from the work of authors such as Al Nuaimi et al. (2015), Hashem et al. (2016), Lim et al. (2018), or Wu et al. (2018).

It is important to note that some experts raised additional points in the questionnaires that were either too specific or did not fit the topic or structure of the questionnaire. For instance, one expert mentioned the need for organizational and financial considerations, as well as conducting a Smart City readiness analysis, which is not directly related to the issue of big data and data protection. Nonetheless, it is one of the steps that should be implemented within Smart Cities. For example, the expert suggests a study to determine whether the city has sufficient financial, computing, and human resources for such projects, or whether city officials support big data projects. This discussion is also supported by studies such as Al Nuaimi et al. (2015), Barham and Daim (2020), Pivar (2020), Wu et al. (2018) or Yigitcanlar et al. (2022).

The same expert also mentioned recommendations related to funding and recommendations dealing with stakeholders, their needs, skills, and competencies, which is again one of the steps that cities should pay attention to (Del-Real et al., 2023; Kummitha & Crutzen, 2017). Another expert mentioned the need to conduct an initial analysis before creating a strategy, in which cities would assess within the projects: (1) whether it is just a proposal for consideration, (2) whether it is a specific proposal that includes a detailed clarification of requirements, (3) whether the project is already implemented but should be improved to meet new challenges arising from big data and data protection, and (4) whether the project utilizes existing infrastructure, hardware, and software resources. It is also related to evaluating implemented solutions using different methods, a topic that has often been overlooked in strategies (Barham & Daim, 2020; Monzon, 2015).

Our findings can be further clarified and extended in the context of strategic choices, which can help to select the most suitable approach for Smart City strategy development (Angelidou, 2014). Infrastructure-oriented and sector-based strategies seem to be the approaches that align with the importance of ICT and our findings. We found that the most common big data related context in Smart City strategies is represented by big data analytics. Characteristics of big data analytics frameworks were discussed by Osman (2019), who then proposed a new framework that uses big data value chain operators and functional requirements. AI technologies and big data can also help achieve urban sustainability by optimizing resource use, reducing energy consumption, and improving infrastructure efficiency. Allam and Dhunny (2019) explored the challenges of ICT through big data and AI to develop a framework to improve the liveability of Smart Cities. Since the authors considered only three areas, namely culture, governance, and metabolism, the framework can be extended using 10 smart areas identified in our paper.

Finally, Smart Cities, driven by big data and interconnected technologies, present ethical challenges related to privacy, surveillance, data ownership, and algorithmic governance. Chang (2021) focused on ethical and non-ethical issues in big data analytics applications in Smart Cities and revealed a need to consider individual privacy, data integrity, and social equity. In this regard, addressing these ethical and socio-political implications requires rethinking Smart City strategies to ensure that data protection goes beyond technical security. It involves fostering a human-centered approach to Smart City development, where residents have control over their data, informed consent is a priority, and public engagement shapes how data are used for the common good (Braun et al., 2018; Ziosi et al., 2024).

5.1. Theoretical implications

First, this paper identifies a theoretical contribution in establishing a linkage between big data, data protection, strategy, projects, and Smart Cities concepts. By fostering a unified perspective on these interrelated elements, the paper aims to create a more holistic understanding of their roles within the broader Smart City paradigm. Second, it delves more into the processes of developing and implementing Smart City strategies and contributes to the literature on urban policy and planning by highlighting the importance of big data and data protection measures in Smart City projects. More precisely, it shows how best practices and guidelines can be extracted, merged, transformed into recommendations, and validated to address challenges of not only big data but also other modern ICT and technologies that affect urban policy and planning in Smart Cities.

Third, the paper also utilized the NIST standard and its general requirements for big data systems in its recommendations. Experts evaluated the perceived importance of these requirements for application within the Smart City domain. It contributes to the literature on big data systems, architectures, and respective frameworks by clarifying Smart City components that should be considered in these frameworks. This work additionally explores the specific contexts within which Smart City strategies discuss and potentially utilize big data. The study examines the recurring themes and rationales behind big data adoption in Smart City strategies by employing a qualitative content analysis of policy documents, academic literature, and project reports.

5.2. Practical and policy implications

This research delivers noteworthy contributions to the practical domain. These contributions are represented by the results of data analyses. They show how Smart Cities strategies are structured and which of their elements are the most determining ones in the context of their interrelations. Also, considering the benchmarking efforts, this paper provides contributions resulting from the cluster analysis and suggests the areas on which Smart Cities should focus in urban policy and planning. The most important practical and policy implications result from the validated list of recommendations. It outlines the focus areas for big data, the key requirements for big data systems, and the general requirements that can be applied within the NIST framework. Further, the list provides recommendations on how to proceed with certain aspects of data protection, particularly regarding policy.

This study contributes to the practical discourse on Smart City strategies by emphasizing the need for adaptable, context-sensitive recommendations, particularly in the realm of big data and data protection. A key insight emerging from our findings is the importance of tailoring big data and data protection strategies to align with a city's specific resource availability and governance maturity. This adaptive approach ensures that recommendations are actionable for cities across varying levels of budgetary and infrastructural capacity, enhancing their real-world applicability. For example, tiered data protection strategies can provide a scalable framework: smaller cities with limited resources might prioritize foundational measures such as basic encryption and clear data handling policies, while larger cities with advanced governance structures could implement more sophisticated approaches, such as real-time monitoring systems and AI-driven threat detection. This tiered system acknowledges the diversity of urban contexts and enables cities to gradually enhance their data protection capabilities as resources and governance structures evolve. By embedding flexibility into the recommendations, the study addresses potential weaknesses observed in the analysis, ensuring that solutions are not only prescriptive but also practical and scalable.

5.3. Limitations

The process of selecting Smart City strategies for analysis revealed several limitations. While the exploration identified other relevant documents about big data and data protection, these resources were excluded due to constraints imposed by the paper's timeframe and overall scope. Additionally, web-based forms of Smart City strategies were not investigated due to the inherent challenges of establishing a consistent basis for comparison with document-based strategies. Also, the study also faces limitations related to its reliance on document-based analysis. Specifically, the lack of stakeholder input or real-world validation limits the applicability and practical relevance of the findings. Thus, the conclusions drawn may not fully reflect the complexities or feasibility of implementing Smart City strategies in diverse urban contexts.

It is also important to acknowledge the potential limitations associated with this paper's data processing and analysis stages. These limitations may include human error during data entry into the tables. Additionally, challenges were encountered in extracting specific data points from certain documents, particularly publication years and strategy validity periods. Further, the presence of multiple documents about the same city's strategy necessitated a selection process, potentially introducing an element of subjectivity. The research process encountered additional limitations related to document accessibility. Sometimes, translating documents from various languages was necessary, potentially introducing errors during data entry or compromising document selection accuracy. Furthermore, resource constraints limited the translation of certain documents, while others remained entirely inaccessible.

There are limitations related to the use of the Delphi method. The number of experts involved in the process could be enriched with others who could bring other opinions on this topic. Although the Delphi method facilitated the development and validation of the recommendations' list, it is important to acknowledge that the generalizability of these recommendations may be limited. It is because the underlying data analysis primarily drew upon insights gleaned from leading Smart Cities that were identified through established rankings. Consequently, the recommendations may be more applicable to larger urban environments with financial, computing, and human resources to develop and manage projects successfully.

To enhance the generalizability of the findings, it is essential to address these disparities and offer adaptable, scalable solutions for cities of varying sizes. Recommendations could be re-designed to include tiered approaches, allowing smaller cities to prioritize incremental changes that align with their specific resources and needs. For example, instead of building an expansive real-time data analytics system, smaller cities could focus on targeted, low-cost applications of big data in key areas like traffic management or public safety, using open source and cloud-based solutions or shared regional resources. Other barriers may include legislative restrictions, such as outdated privacy laws or fragmented regulatory frameworks, which may inhibit cities from adopting advanced data protection measures. Additionally, cities with limited data governance capacity may lack the expertise or institutional infrastructure necessary to implement and sustain such strategies effectively. Inter-departmental coordination is another limitation. Effective big data and data protection strategies often require collaboration across multiple city departments, each with distinct priorities, operational silos, and resource constraints. Without strong coordination mechanisms, even well-designed strategies may fail to achieve their intended outcomes.

5.4. Future research areas

The current study employed a purposive sample of 40 cities identified through the 2023 SCI ranking, representing a subset of the 141 cities listed in the index. This approach presents an opportunity for future research to leverage the complete dataset and analyze the remaining 101 Smart City strategies, if available. Furthermore, expanding the sample size of experts who evaluated the recommendations using the Delphi technique is another potential avenue for future investigation. However, it is important to acknowledge the associated time constraints and potential cost considerations that may arise when engaging a larger panel of experts.

This study identifies several areas ripe for further theoretical exploration. Notably, the findings suggest a relative need for more research on data security within the domain of Smart Cities. This gap presents a compelling opportunity for a dedicated research initiative focused on data security considerations in Smart City contexts. During the Python-based cluster analysis of the data, an initial exploration aimed to identify recurring patterns within the clusters, particularly regarding big data utilization. However, this avenue of inquiry was not comprehensively explored due to the limitations imposed by the paper's scope. This investigation into big data patterns within clusters presents a promising avenue for future research. A potential avenue for future research lies in exploring additional city documents about big data and data protection. These supplementary resources may offer deeper insights into various cities' specific approaches and considerations in these critical areas.

Finally, future research should focus on identifying best practices for overcoming limitations presented in the previous section, such as developing modular policy frameworks adaptable to varying legislative environments or providing capacity-building programs to enhance data governance expertise in resource-constrained cities. Additionally, exploring mechanisms for fostering inter-departmental collaboration – such as centralized data governance committees or inter-agency task forces – could provide actionable insights for policymakers and practitioners. Another research avenue could be focus on aligning big data and data protection policies with new technology-driven risks and opportunities. Investigation of the cost-benefit dynamics of investing in advanced big data and data protection measures, particularly in cities with budgetary constraints could help policymakers justify and prioritize expenditures on data governance.

6. Conclusions

As the complexity of big data grows, so do the risks related to data privacy, security vulnerabilities, and ethical considerations, emphasizing the need for robust frameworks and strategies to ensure that data protection measures keep pace with technological advancements and the expanding data ecosystem. This paper examined big data and data protection measures used in Smart Cities strategies in 28 cities worldwide. It aimed to identify specific recommendations for developing Smart City projects focused on big data and data protection. A systematic approach was followed to ensure trustworthy and reliable results. It involved a series of steps for in-depth data analysis and evaluation. Different methods were used throughout the process, including finding relevant resources, analyzing content, and using the Delphi method to validate the gathered data.

A sample of 40 cities was identified based on the SCI 2023 report. Among these, only 30 possessed documented Smart City strategies. However, due to data availability limitations, the final analysis included

28 strategies. Examining these strategies revealed that a substantial majority (68 %, $n = 19$) explicitly incorporated the term *Smart* within their titles. Furthermore, the term *Digital* emerged as a frequently used term within these strategies. Half of the selected cities mentioned big data in their strategies, with the most common contexts being big data analytics and transportation. Cluster analysis, a key component of the Python model built for this study, focused on grouping cities based on their big data usage patterns. While a detailed examination of each cluster revealed certain commonalities, the findings were deemed inconclusive based on the sample size.

Informed by the insights gleaned from both the literature analysis and analysis of Smart City strategies, a comprehensive list of recommendations was subsequently formulated. This list was then subjected to a rigorous validation process employing the Delphi method across three iterative rounds, with the participation of seven experts. These recommendations focus on creating a comprehensive, integrated approach to data protection, recognizing that while big data holds great potential for optimizing resources, enhancing infrastructure, and facilitating real-time, data-driven urban management, it also requires strict safeguards to prevent misuse and protect residents' privacy. Firstly, implementing these data management strategies could alleviate privacy concerns by establishing clear guidelines for data governance and security, addressing issues such as those noted by Ziosi et al. (2024), who warn against the risks of pervasive surveillance and loss of autonomy in data-driven urban systems. By incorporating measures to protect personal data, Smart City initiatives can gain public trust and foster inclusivity, mitigating some of the ethical dilemmas surrounding data collection and surveillance.

Furthermore, the Delphi recommendations suggest that such strategies could enable more equitable integration of technologies like IoT, AI, and cloud computing to deliver urban services fairly, a crucial step in addressing the e-gentrification concerns described by Middha and McShane (2022). Ensuring that digital infrastructures benefit all citizens, including marginalized groups, would help prevent the social displacement associated with digital inequities in Smart Cities. Additionally, while the recommendations are tailored primarily to larger cities, they acknowledge potential applicability to smaller urban areas, which could help address the regional disparities highlighted by Mykhnenko (2023). By extending these data protection and integration strategies beyond major urban centers, Smart City models could support underserved or peripheral areas, fostering more balanced urban development and preventing the concentration of resources and innovation in already thriving city cores. Finally, future research could explore modular policy frameworks, mechanisms for inter-departmental collaboration, and the cost-benefit dynamics of investing in big data and data protection measures, especially for resource-constrained cities.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Martin Lnenicka: Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Conceptualization, Writing – review & editing. **Petr Hervert:** Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Oldrich Horak:** Writing – original draft, Supervision, Methodology, Conceptualization.

Acknowledgments

The authors would like to express their gratitude to the University of Pardubice that supported this work financially and materially through the project of the Student Grant Competition - SGS_2024_017 (Information technology and data analytics as a means of supporting the development of a smart region).

Appendixes

Table A1
Description of selected Smart Cities and their strategies (source: own elaboration).

SCI Rank	City HDI	City	Population (2023)	Country	Country HDI	Strategy format	Language	Strategy title	Year of publication	Validity of the strategy	No. of pages	Does it include big data?	No. of mentions of big data	Applications of big data
1	0.989	Zurich	1431 538	Switzerland	0.962	Text	English and German	Smart city strategy	2018	2035 (part of another strategy)	No			
2	0.98	Oslo	1085 992	Norway	0.961	N/A								
3	0.98	Canberra	472 304	Australia	0.951	Text	English	Digital strategy	2020	2022	31	No		
4	0.967	Copenhagen	1381 005	Denmark	0.948	Presentation	English and Danish	Smart city	2015	Was not specified	50	Yes	10	Platform, transportation, urban planning, waste management
6	0.973	London	9648 110	United Kingdome	0.929	Text	English	Smart city plan	Was not specified	2020	54	Yes	5	Data analytics, transportation
7	0.939	Singapore	6080 859	Singapore	0.939	Text	English	Smart project	2018	Was not specified	34	Yes	1	Big data analytics
8	0.96	Helsinki	1337 786	Finland	0.94	N/A								
10	0.972	Stockholm	1700 066	Sweden	0.947	Text	Swedish	Smart city strategy	2021	2040	39	Yes	10	Transportation, platform, analytics, AI, machine learning, real-time
11	0.972	Hamburg	1787 520	Germany	0.942	Text	English	Digital strategy	2020	Was not specified	31	Yes	5	Big data analytics, IoT, urban objectives
12	0.907	Beijing	21766 214	China	0.768	Presentation	Chinese	Smart city	2019	Could not translate	41	Could not translate		
13	0.911	Abu Dhabi	1566 999	United Arab Emirates	0.911	Presentation	English and Arabic	Digital transformation plan	2019	2029+	79	Yes	5	Big data analytics, big data management, big data engines
14	0.96	Prague	1323 339	Czech Republic	0.889	Presentation	Czech	Concept smart city	2017	2030	88	Yes	20	Tourism, big data analytics, social media, companies, transportation, energy
15	0.962	Amsterdam	1174 025	Netherlands	0.941	Text	Dutch	Digital town	2019	2025	24	No		
16	0.952	Seoul	9988 049	South Korea	0.925	Presentation	English	Smart city status and strategies	Was not specified	Was not specified	21	Yes	1	E-government
19	0.946	Hong Kong	7684 801	Hong Kong	0.949	Text	English/ Chinese	Smart city blueprint	2020	Was not specified	36	Yes	3	Big data analytics, real-time
21	0.938	New York	7888 121	USA	0.921	Text	English	Building smart city	2015	Was not specified	24	No		
22	0.951	Auckland	1673 220	New Zealand	0.937	N/A								
26	0.959	Reykjavik	137 618	Island	0.959	Presentation	Icelandic	Employment and innovation strategy	2022	2028	27	No		
29	0.916	Taipei City	2754 196	Tai wan	0.916	Text	Chinese and English	Smart city	Was not specified	Was not specified	35	Yes	4	Big data analytics, AI, police, water supply management
30	0.9	Riyadh	7682 430	Saudi Arabia	0.875	Only national strategy								
34	0.949	Boston	617 459	USA	0.921	N/A								
39	0.94	Washington D.C.	631 693	USA	0.921	N/A								
40	0.943	Ottawa	1010 391	Canada	0.936	Text	English	Smart city	2017	Was not specified	30	Yes	3	Data analysis, big data analytics

(*continued on next page*)

Table A1 (*continued*)

SCI Rank	City HDI	City	Population (2023)	Country	Country HDI	Strategy format	Language	Strategy title	Year of publication	Validity of the strategy	No. of pages	Does it include big data?	No. of mentions of big data	Applications of big data
42	0.944	Vancouver	672 857	Canada	0.936	Presentation	English	Digital strategy	2013	2016+	36	No		
48	0.943	Toronto	2903 456	Canada	0.936	Text	English	Digital infrastructure: strategic framework	2022	Was not specified	108	No		
50	0.931	Los Angeles	3769 485	USA	0.921	Text	English	Smart city	2020	2028	54	No		
53	0.942	Denver	699 288	USA	0.921	Text	English	Smart city challenge	2016	Was not specified	33	No		
55	0.94	Seattle	725 487	USA	0.921	Text	English	Smart city	2020	Was not specified	32	Yes	3	Big data analytics, big data systems, transportation
59	0.855	Doha	658 344	Qatar	0.855	N/A								
61	0.929	Chicago	2608 425	USA	0.921	Presentation	English	IT Strategic plan	2021	Was not specified	167	No		
72	0.951	Tokyo	37194 105	Japan	0.925	Presentation	Japanese	Smart city implementation strategy	2019	2040	69	Yes	5	Economy, social problems, urban development, local problem solving, transportation, research
88	0.839	Bangkok	11069 982	Thailand	0.8	Text	Thai	Development with digital technologies	2018	2022	12	No		
89	0.858	Kuala Lumpur	8621 724	Malaysia	0.803	Presentation	English	Smart city plan	2021	2025	21	No		
91	0.919	Tel Aviv	4420 855	Israel	0.919	N/A								
96	0.816	Muscat	1650 319	Oman	0.816	Only national strategy								
100	0.744	Hanoi	5253 385	Vietnam	0.703	N/A								
102	0.759	Jakarta	11248 839	Indonesia	0.705	Presentation	English	Building smart city	Was not specified	Was not specified	20	Yes	2	Digital technology and innovation, analysis
105	0.73	Delhi	32941 309	India	0.633	N/A								
117	0.896	Nicosia	412 156	Cyprus	0.896	Presentation	Greek	Smart City strategie	2018	2028	19	No		
118	0.757	Medellin	4102 308	Colombia	0.752	Presentation	Spanish	Smart city	2021	2030	31	No		

References

Al Nuaimi, A. J. J., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications, 6*, 1–15.

Allam, Z., & Dhunny, Z. A. (2019). On big data, artificial intelligence and smart cities. *Cities (London, England), 89*, 80–91.

Angelidou, M. (2014). Smart city policies: A spatial approach. *Cities (London, England), 41*, S3–S11.

Arockia, P., Varnekha, S. S., & Veneshia, K. A. (2017). The 17 V's of Big Data. *International Research Journal of Engineering and Technology, 4*(9), 329–333.

Barham, H., & Daim, T. (2020). The use of readiness assessment for big data projects. *Sustainable Cities and Society, 60*, Article 102233.

Barlow, M., & Levy-Bencheton, C. (2018). *Smart cities, smart future: Showcasing tomorrow*. John Willey and Sons.

Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography, 3*(3), 274–279.

Bawany, N. Z., & Shamsi, J. A. (2015). Smart city architecture: Vision and challenges. *International Journal of Advanced Computer Science and Applications, 6*(11), 1–11.

Bellanova, R., & De Goede, M. (2022). The algorithmic regulation of security: An infrastructural perspective. *Regulation & governance, 16*(1), 102–118.

Berrone, P., & Ricart, J. E. (2022). *IESE cities in motion index*. Business School - University of Navarra Available at: Https://www.iese.edu/media/research/pdfs/ST-0633-E.pdf .

Bibri, S. E., Alexandre, A., Sharifi, A., & Krogstie, J. (2023). Environmentally sustainable smart cities and their converging AI, IoT, and big data technologies and solutions: An integrated approach to an extensive literature review. *Energy Informatics, 6*(1), 1–39.

Bokolo, A. J. (2023). Data driven approaches for smart city planning and design: A case scenario on urban data management. *Digital Policy, Regulation and Governance, 25*(4), 351–367.

Braun, T., Fung, B. C., Iqbal, F., & Shah, B. (2018). Security and privacy challenges in smart cities. *Sustainable cities and society, 39*, 499–507.

Camerin, F., Álvarez-Del-Valle, L., Díez-Bermejo, A., & Rodríguez-Suárez, I. (2024). Towards the Spanish local urban Agenda. The evolution of urban regeneration in Spain (2014-2022). *TeMA-Journal of Land Use, Mobility and Environment, 17*(1), 51–70.

Chang, V. (2021). An ethical framework for big data and smart cities. *Technological Forecasting and Social Change, 165*, Article 120559.

Chauhan, S., Agarwal, N., & Kar, A. K. (2016). Addressing big data challenges in smart cities: A systematic literature review. *info, 18*(4), 73–90.

Crisp, J., Pelletier, D., Duffield, C., Adams, A., & Nagy, S. U. E. (1997). The delphi method? *Nursing research, 46*(2), 116–118.

Czupich, M. (2019). The role of ICT in the smart city concept. *Olsztyn Economic Journal, 14*(1),
63–74.

Escolar, S., Villanueva, F. J., Santofimia, M. J., Villa, D., del Toro, X., & López, J. C. (2019). A multiple-attribute decision making-based approach for smart city rankings design. *Technological Forecasting and Social Change, 142*, 42–55.

Das, D. K. (2024). Exploring the symbiotic relationship between digital transformation, infrastructure, service delivery, and governance for smart sustainable cities. *Smart Cities, 7*(2), 806–835.

Del-Real, C., Ward, C., & Sartipi, M. (2023). What do people want in a smart city? Exploring the stakeholders' opinions, priorities, and perceived barriers in a medium-sized city in the United States. *International Journal of Urban Sciences, 27*(sup1), 50–74.

Favaretto, M., De Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PloS one, 15*(2), Article e0228987.

Giffinger, R., Fertner, C., Kramar, H., Kalasek, R., Pichler-Milanović, N., & Meijers, E. (2007). *Smart cities. ranking of european medium-sized cities. final report*. Centre of Regional Science (SRF), Vienna University of Technology.

Glaeser, E. L., Kominers, S. D., Luca, M., & Naik, N. (2018). Big data and big cities: The promises and limitations of improved measures of urban life. *Economic Inquiry, 56*(1), 114–137.

Hacker, M. E., & Binz, C. (2021). Navigating institutional complexity in socio-technical transitions. *Environmental Innovation and Societal Transitions, 40*, 367–381.

Halegoua, G. (2020). *Smart cities*. MIT Press.

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., et al., (2016). The role of big data in smart city. *International Journal of Information Management, 36*(5), 748–758.

Holubová, I., Kosek, J., Minařík, K., & Novák, D. (2015). *Big Data a NoSQL databáze*. Profes!onal. Praha, Grada.

Hussein, A. A. (2020). Fifty-six big data V's characteristics and proposed strategies to overcome security and privacy challenges (BD2). *Journal of Information Security, 11*(4), 304–328.

IMD Business School. (2023). *IMD Smart City Index Report 2023*. Available at: Https://www.imd.org/wp-content/uploads/2023/04/smartcityindex-2023-v7.pdf.

Isarsoft GmbH. *Smart city strategy: The foundation of every smart city*. Available at: Https://www.isarsoft.com/article/smart-city-strategy-the-foundation-of-every-smart-city.

Ismagilova, E., Hughes, L., Rana, N. P., & Dwivedi, Y. K. (2022). Security, privacy and risks within smart cities: Literature review and development of a smart city interaction framework. *Information Systems Frontiers, 24*, 393–414.

Javed, A. R., Shahzad, F., ur Rehman, S., Zikria, Y. B., Razzak, I., Jalil, Z., et al., (2022). Future smart cities: Requirements, emerging technologies, applications, challenges, and future aspects. *Cities (London, England), 129*, Article 103794.

Khan, Z., Anjum, A., & Kiani, S. L. (2013). Cloud based big data analytics for smart future cities. In *2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing* (pp. 381–386). IEEE.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal, 79*(1), 1–14.

Kummitha, R. K. R., & Crutzen, N. (2017). How do we understand smart cities? An evolutionary perspective. *Cities (London, England), 67*, 43–52.

Lai, C. M. T., & Cole, A. (2023). Measuring progress of smart cities: Indexing the smart city indices. *Urban Governance, 3*(1), 45–57.

Lilja, K. K., Laakso, K., & Palomäki, J. (2011). Using the Delphi method. In *2011 Proceedings of PICMET'11: Technology Management in the Energy Smart World (PICMET)* (pp. 1–10). IEEE.

Lim, C., Kim, K. J., & Maglio, P. P. (2018). Smart cities with big data: Reference models, challenges, and considerations. *Cities (London, England), 82*, 86–99.

Li, D., Cao, J., & Yao, Y. (2015). Big data in smart cities. *Science China. Information Sciences, 58*(10), 1–12.

Lnenicka, M., & Komarkova, J. (2019). Big and open linked data analytics ecosystem: Theoretical background and essential elements. *Government Information Quarterly, 36*(1), 129–144.

Lnenicka, M., Nikiforova, A., Clarinval, A., Luterek, M., Rudmark, D., Neumaier, S., et al., (2024). Sustainable open data ecosystems in smart cities: A platform theory-based analysis of 19 European cities. *Cities (London, England), 148*, Article 104851.

Löfgren, K., & Webster, C. W. R. (2020). The value of Big Data in government: The case of 'smart cities. *Big Data & Society, 7*(1), 1–14.

Middha, B., & McShane, I. (2022). E-gentrification: Digital community engagement, urban change and digital rights to the city. In *Citizen participation in the information society: Comparing participatory channels in urban development* (pp. 141–165). Springer International Publishing.

Monzon, A. (2015). Smart cities concept and challenges: Bases for the assessment of smart city projects. In *2015 international conference on smart cities and green ICT systems (SMARTGREENS)* (pp. 1–11). IEEE.

Mullen, P. M. (2003). Delphi: Myths and reality. *Journal of Health Organization and Management, 17*(1), 37–52.

Mykhnenko, V. (2023). Smart shrinkage solutions? The future of present-day urban regeneration on the inner peripheries of Europe. *Applied Geography, 157*, Article 103018.

Osman, A. M. S (2019). A novel big data analytics framework for smart cities. *Future Generation Computer Systems, 91*, 620–633.

Pelton, J. N., & Singh, I. B. (2018). *Smart cities of today and tomorrow: Better technology, infrastructure and security*. Springer International Publishing.

Pivar, J. (2020). City management support and smart city strategy as success factors in adopting big data technologies for smart cities. In *Smart governments, regions and cities* (pp. 167–183). University of Rijeka.

Rani, S., Mishra, R. K., Usman, M., Kataria, A., Kumar, P., Bhambri, P., et al., (2021). Amalgamation of advanced technologies for sustainable development of smart city environment: A review. *IEEE Access: Practical Innovations, Open Solutions, 9*, 150060–150087.

Rosati, U., & Conti, S. (2016). What is a smart city project? An urban model or a corporate business plan? *Procedia-Social and Behavioral Sciences, 223*, 968–973.

Sharifi, A. (2019). A critical review of selected smart city assessment tools and indicator sets. *Journal of Cleaner Production, 233*, 1269–1283.

Sharifi, A., Allam, Z., Bibri, S. E., & Khavarian-Garmsir, A. R. (2024). Smart cities and sustainable development goals (SDGs): A systematic literature review of co-benefits and trade-offs. *Cities (London, England), 146*, Article 104659.

Smart Nation and Digital Government Office. (2018). *Smart nation: The way forward – executive summary*.

Stockholms Stad. (2021). *Strategi för stockholm som smart och uppkopplad stad: Bilaga 1. strategi*. Stockholms stad.

Sun, L., Zhang, H., & Fang, C. (2021). Data security governance in the era of big data: Status, challenges, and prospects. *Data Science and Management, 2*, 41–44.

Viale Pereira, G., Cunha, M. A., Lampoltshammer, T. J., Parycek, P., & Testa, M. G. (2017). Increasing collaboration and participation in smart city governance: A cross-case analysis of smart city initiatives. *Information Technology for Development, 23*(3), 526–553.

Williams, S. P., Hardy, C. A., & Holgate, J. A. (2013). Information security governance practices in critical infrastructure organizations: A socio-technical and institutional logic perspective. *Electronic Markets, 23*, 341–354.

World Economic Forum. (2021). *Governing smart cities: Policy benchmarks for ethical and responsible smart city development*. World Economic Forum, Geneva.

Wu, S. M., Chen, T. C., Wu, Y. J., & Lytras, M. (2018). Smart cities in Taiwan: A perspective on big data applications. *Sustainability, 10*(1), 106.

Wu, Y. J., & Chen, J. C. (2021). A structured method for smart city project selection. *International Journal of Information Management, 56*, Article 101981.

Yigitcanlar, T., Degirmenci, K., Butler, L., & Desouza, K. C. (2022). What are the key factors affecting smart city transformation readiness? Evidence from Australian cities. *Cities (London, England), 120*, Article 103434.

Ziosi, M., Hewitt, B., Juneja, P., Taddeo, M., & Floridi, L. (2024). Smart cities: Reviewing the debate about their ethical implications. *AI & Society, 39*, 1185–1200.