IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 22 June 2023, accepted 25 July 2023, date of publication 7 August 2023, date of current version 16 August 2023.

*Digital Object Identifier 10.1109/ACCESS.2023.3302835*

RESEARCH ARTICLE

# Toward Urban Data Governance: Status-Quo, Challenges, and Success Factors

**YUSUF BOZKURT[1,2], ALEXANDER ROSSMANN[2], ASHWINI KONANAHALLI[1], AND ZEESHAN PERVEZ[1], (Senior Member, IEEE)**
[1]Department of Computing, Engineering and Physical Sciences, University of the West of Scotland, PA1 2BE Paisley, Scotland, U.K.
[2]Herman Hollerith Centre, Department of Information Systems, Reutlingen University, 72762 Reutlingen, Germany

Corresponding author: Yusuf Bozkurt (yusuf.bozkurt@reutlingen-university.de)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of the West of Scotland (UWS) Ethics Committee under Case No. 16174.

**ABSTRACT** The benefits of urban data cannot be realized without a political and strategic view of data use. A core concept within this view is data governance, which aligns strategy in data-relevant structures and entities with data processes, actors, architectures, and overall data management. Data governance is not a new concept and has long been addressed by scientists and practitioners from an enterprise perspective. In the urban context, however, data governance has only recently attracted increased attention, despite the unprecedented relevance of data in the advent of smart cities. Urban data governance can create semantic compatibility between heterogeneous technologies and data silos and connect stakeholders by standardizing data models, processes, and policies. This research provides a foundation for developing a reference model for urban data governance, identifies challenges in dealing with data in cities, and defines factors for the successful implementation of urban data governance. To obtain the best possible insights, the study carries out qualitative research following the design science research paradigm, conducting semi-structured expert interviews with 27 municipalities from Austria, Germany, Denmark, Finland, Sweden, and the Netherlands. The subsequent data analysis based on cognitive maps provides valuable insights into urban data governance. The interview transcripts were transferred and synthesized into comprehensive urban data governance maps to analyze entities and complex relationships with respect to the current state, challenges, and success factors of urban data governance. The findings show that each municipal department defines data governance separately, with no uniform approach. Given cultural factors, siloed data architectures have emerged in cities, leading to interoperability and integrability issues. A city-wide data governance entity in a cross-cutting function can be instrumental in breaking down silos in cities and creating a unified view of the city's data landscape. The further identified concepts and their mutual interaction offer a powerful tool for developing a reference model for urban data governance and for the strategic orientation of cities on their way to data-driven organizations.

**INDEX TERMS** Cognitive mapping, data governance, design science research, urban data governance, smart city, expert interviews.

## I. INTRODUCTION

Digitization is not possible without data. Cities face the problem of dealing with various data formats, data sources, and actors as part of their digital transformation to realize beneficial uses of urban data. An integrated, unified, and comprehensive data base is the foundation for implementing data-driven cities, ranging from simple to more complex scenarios (e.g. developing a digital twin) [1], [2]. Cities are considered data factories with diverse data sources

The associate editor coordinating the review of this manuscript and approving it for publication was Edith C.-H. Ngai.

and a growing number of Internet of Things smart city applications [2], [3], [4]. The benefits of urban data cannot be achieved without a political and strategic view of data use. An essential element in this view is data governance. In the corporate context, data governance aligns the corporate strategy in the data-relevant structures and units with data processes, actors, architectures, and data management. Therefore, data governance can establish semantic compatibility among heterogeneous technologies and data silos and connect stakeholders by standardizing data models, processes, and policies [5], [6], [7], [8], [9]. Many cities have so far advertised raw concepts and strategy papers on smart cities and digitalization activities, referring to high-level principles for data handling while not specifying their approach to data governance [10], [11], [12], [13], [14], [15]. Only a few smart city pioneers have recognized the relevance of data and developed data strategies, such as Data Excellence [16], Data for London [17], and City Data Commons [18]. Although the relevance of data in cities is also acknowledged by higher-level policy initiatives, such as in Germany, which published a national data strategy in 2021 [19], critics maintain that it is not specific enough to implement any concrete measures [20]. Despite these state-level initiatives in the area of data governance, cities need to develop stronger coordination and unified implementation of the dimensions of data management. In addition, the roles, actors, and processes in this cross-cutting issue must be clarified further to implement data governance effectively [21]. A systematic literature review [22] on urban data governance reveals that urban data concerns can be categorized into a conceptual framework with eight dimensions of data governance. This literature review shows that cities would benefit from a reference model for urban data governance by applying it as an orientation for the realization of their concrete data governance programs and would achieve their smart city goals more effectively through the proper handling of data [22]. However, research on urban data governance is scarce [5], [22], [23], [24], [25], [26], [27], [28], and a reference model for it is lacking, even though such a model would save development costs in implementing practical solutions by adapting it to the local needs of a city. The absence of a reference model that aligns the overarching aspects of stakeholders and roles, technology, and processes in urban data governance represents a gap in current research and practical implementation [22]. For such a reference model, a design base needs to be established by examining the current state of data governance in cities and identifying the challenges and requirements, which is also hardly addressed in current research.

To fill these research gaps, this study focuses on the following three research questions (RQs):

- RQ1: How is the data ecosystem in cities structured?
- RQ2: What challenges do cities face in governing and managing data?
- RQ3: What measures do cities consider to implement data governance?

To answer these questions, this study evaluates the current state of data governance in cities, identifies problems and challenges with data governance, and delineates factors relevant to a successful implementation. RQ1 articulates how cities are organized in terms of their data ecosystem, to understand the current environment that a reference model is meant to cover. RQ2 elaborates on the city's problems and challenges in handling and managing data. Such problems can relate to technological but also organizational or behavioral issues. Finally, RQ3 aims to understand which measures cities consider relevant for the successful implementation of data governance.

The study is based on the research paradigm of design science research (DSR) [29], [30]. DSR is a design-oriented research paradigm aimed at the output of an artifact and contributes to the problem environment [29], [30]. The goal of this study is to gain knowledge on the problem space and define the scope of the problem to be addressed. It also identifies the requirements for the artifact. Different research methods can be used in the DSR paradigm, and therefore semi-structured expert interviews served as a research method because a qualitative research approach is suitable for answering questions in the exploratory knowledge-building phase. Thus, municipal experts were interviewed to develop a recommendation for the urban data governance reference model in further research.

The remainder of the article proceeds as follows: Section II presents the theoretical background for the basic concepts of smart cities and data governance to enable a common understanding. The study then delves deeper into related work to better understand the research structure around urban data governance and to clarify the contribution to the knowledge base. Section III describes the research design. Section IV presents the results of the expert interviews, followed by a discussion of the implications, limitations, validity, and directions for future research in Section V. Section VI concludes the article.

## II. THEORETICAL BACKGROUND

The following subsections aim to create a common understanding of the core concepts of smart cities and data governance. They also describe the current state of the scientific literature on urban data governance and the resulting research gaps and explain how this article contributes to the body of knowledge by addressing these gaps and outlining its contributions.

### A. CONTEXT

Urban data governance is based on the concept of smart cities and data governance [22]. Therefore, before delving into the results, this article discusses both concepts to create a common understanding.

#### 1) SMART CITIES

The concept of smart cities is multifaceted and multidisciplinary, making it difficult to give a universal definition.

However, an increasingly accepted consensus is that the term "smart city" is a concept that addresses the challenges of growing urbanization [31], [32], [33], [34], [35], [36]. The problems of urbanization include waste management, resource consumption, pollution, traffic congestion, and infrastructure obsolescence, but also technical, material, physical, social, and organizational challenges [37], [38], [39], [40]. These challenges arise from the diverse structure of stakeholders in the city [41], [42] and the interdependencies that complicate urban problems [34], [35]. Cities are a wide organic network consisting of many actors, so an interplay of technology, human capital, and collaboration is required to solve urbanization problems and improve the quality of life [31], [33], [34], [43], [44]. The deployment of information and communication technologies (ICT) plays an enabling role in all smart solutions [31], [45]. In this context, the Internet of Things has attracted a great deal of interest as it adds sensors and generates data to traditional urban infrastructure [32], [35], [36], [44]. In this context, the basic building blocks of data collection, data management, data processing, and application handling are common to any domain used [35]. Batty et al. [32] emphasize that data management constitutes a city's brain, as all smart city applications and data-driven decisions are based on it. A city aims to be smart by automating routine tasks in buildings, waste management, and transport systems; monitoring, analyzing, and planning the city using data; and making strategic and operational decisions based on data [32]. However, dealing with urban data is not straightforward. Many different data sources exist in a city's ecosystem (e.g., sensors, city management systems, cameras, social media, geographic information systems [GIS], statistics, open data), as do various types of data, many different service providers, and many different actors [32], [34], [35], [46], [47], [48]. These myriad sources make dealing with data in a smart city a technical challenge and a multi-layered problem with organizational and cross-collaborative aspects. Addressing the problem requires the interplay of different actors, not a single organization. The stakeholders in a smart city include research and educational institutions, local and regional authorities, utilities, ICT providers, media, citizens, communities, and non-governmental organizations [46], [47], [49], [50]. Marrone and Hammerle [42] provide a more detailed listing of smart city stakeholder perspectives.

#### 2) DATA GOVERNANCE

Governance is a decentralized theory to which different sectors contribute [51]. Chhotray and Stoker [52, p. 214] define governance as the "practice of making collective decisions". They emphasize the multi-layered nature of decision-making with multiple and diverse actors. Similarly, Kooiman [53] describes governance as a set of "interactions" involving public and private actors to solve societal problems or create societal opportunities, including institutions as the context for these interactions and establishing a normative foundation for all these activities.

In the corporate context, several governance domains have emerged, including human resource, financial, risk, information technology (IT), and data governance, with the latter two being comparatively younger streams [54]. However, a closer review of the definitions of data governance reveals that the original characteristics, such as multidisciplinary, collaboration, and coordination, are recognizable. Mahanti [7] defines data governance as "a system of policies, rules, standards, processes, practices and structures, roles and responsibilities, controls and decision rights to oversee the management of data. Data governance should not be confused with the technical management of data" [7, p. 63]. Data governance consolidates all data management activities that enable effective overall data management by establishing standards, processes, roles, and responsibilities across data management, including data warehousing, business intelligence, metadata management, and big data [7], [54]. Seiner's [8] definition is similar and states that data governance is the execution and enforcement of authority over data management and data assets, focusing on shaping behavior in data handling. Drawing on the theory of IT governance, Khatri and Brown [6] describe data governance as the entity that determines who has decision rights and responsibilities for data. Governance and management are distinguished, with management as the implementor of decisions and governance as the decision-maker for effective management. Furthermore, Weber et al. [55] and Weber and Klingenberg [9] define data governance as decision-making and accountability to promote the proper handling of data assets, emphasizing the aspects of data policies and standards that need to be in line with the corporate culture and strategy so that all stakeholders at all hierarchical levels and departments are aligned with a common goal.

Data governance exists in every organization but may not be formalized, resulting in loose and poorly established data policies and practices. As a result, structures that require policies and authority to implement data governance are lacking, which often leads to problems with interoperability, quality, security, lack of trust in data, and poor personal data protection. By contrast, formal data governance comprises established guidelines, practices, and clear roles and responsibilities [54].

### B. RELATED RESEARCH

The concept of data governance in a general context is not new; it has long received attention from practitioners [8], [56], [57], [58] and academics [6], [9], [55], [59], [60]. Therefore, the existing literature on urban data governance was reviewed to understand the research structure in this area and to clarify the contribution of this work. As a first step, a search of IEEE Xplore, EBSCOhost, ACM Digital Library, and Web of Science databases was undertaken by applying the search terms "smart city data governance," "urban data governance," "city data governance," and "municipal data governance" in the titles and abstracts. No filter was set for the period, and only peer-reviewed articles were considered.

The search yielded 58 initial results. The second step analyzed the titles and abstracts of these articles to ensure that the focus was on data governance in an urban context. This review excluded 50 articles. The full text of the remaining eight were analyzed, but three were excluded because they did not focus on data governance but rather on the city's governance and smart city initiatives. Two other articles were included through a backward citation search. The analysis of related work shows that only seven articles contribute to the knowledge base of urban data governance, thus indicating the need for further research on urban data governance. The following paragraphs provide an overview of the seven identified articles.

The work of Thompson et al. [25] provides first indication of the challenges and problems in the public sector regarding data governance. It draws on an audit of the police firearms management system and the department of health information system in Western Australia, which identifies gaps that can be overcome through sound data governance (i.e., data preparation, interface controls, and accountability). The challenge is that individual departments often acquire technical solutions without aligning with the overall strategy, leading to discrepancies and silos. Thus, it is not the business processes or technologies that cause problems but rather the low level of data governance or the complete lack of data governance in a city's agencies. However, with only two case studies, the article draws only general conclusions.

Paskaleva et al. [27] examine how data governance is integrated into the newly emerging sustainability-focused smart city agenda. They provide a conceptual framework for data governance in smart city initiatives built from secondary data. Building on this conceptual framework, they conduct interviews with the cities of Manchester, Eindhoven, and Stavanger to understand their approaches and challenges in the implementation of smart city projects. They conclude that collaboratively working with different stakeholders is a crucial challenge. Data governance can change how data are generated, collected, and used in a smart city to ensure more sustainable value for citizens and the city. For this, data governance measures must be embedded in all stages of the data life cycle, requiring a comprehensive data governance management plan. This work helps understand how data governance measures are applied in smart city projects. The preliminary framework provides insights into which data dimensions are relevant for data governance in cities. However, as the authors note, their work is a first effort and requires further research involving more extensive data collection because they interviewed only three cities in their study. In addition, the interview questions focus on one concrete smart city initiative and therefore are strongly project-related and offer no insights into how data governance takes place outside the project. Nevertheless, that work offered valuable hints for the design of the current study's interview questions to address the issues in a more targeted way.

Lupi [24] presents a concept for a tool that links data governance to a long-term strategy aligned with the city's development agenda. The focus is on data protection and data manipulation in smart cities. For this purpose, the author analyzed data governance plans from the corporate context against general city plans and formulated a blueprint for a city data plan. The key questions on urban data and its management, drawn from the literature, establish the basis for the development of the urban data plan and provide meta-requirements to be fulfilled.

In their study, Franke and Gailhofer [28] structure the debate on data governance using a regulatory perspective. They draw lessons from previous regulatory discussions and propose guidelines for socio-ecological data governance by assessing ideal types of data regulation. However, the strong regulatory aspect leaves the questions of how data governance can be integrated in cities and what challenges cities have to face in daily operations unanswered.

Choenni et al. [5] argue that proper data governance is necessary to use urban data fully. They consider data quality and the establishment of trust the most important functions of data governance in cities. Focusing on these functions, they describe approaches for implementing them in the urban data ecosystem. However, finding a compromise between competing values such as data protection and the fact that data are a public good is a significant challenge. Therefore, one key factor for the success of data governance in cities is the efficient and effective collaboration between the actors in the urban data space, where they align their needs and requirements.

The results of Cuno et al. [26] study with three German cities led to the concept of the urban data space, which enables an ecosystem for data exchange and value creation by using all types of data within the city. The authors provide a classification of urban data, including legal and monetary aspects and consideration of the technical implementation design based on the ICT reference architecture DIN SPEC 91357 [61]. The urban data space comprises a network of different actors, and the authors identify the multi-layered nature of the actors and describe their roles. Furthermore, the study's results highlight the importance of data governance and data sovereignty in the urban data space and provide first insights into data governance roles in this space. Despite only three cities in Germany being assessed, the work is valuable for the design of urban data governance, as stakeholders, data classification, and ICT architecture are key aspects of urban data governance, and thus the results can further help specify an urban data governance reference model.

König's [23] study formulates a citizen-centered data governance framework that addresses the ethical and legitimacy challenges of using citizen data. The author derives ethical and accountability requirements from the stages of data-based value creation for citizens and formulates data governance mechanisms. Although this study considers data governance from an ethics and accountability perspective, cities should consider these two crucial dimensions for their data governance programs and include them in the design of the urban data governance reference model.

Overall, data governance in the urban context has received less attention than data governance in the corporate context, even though the potential of appropriate urban data governance is now more critical than ever. The literature search revealed only seven articles on this topic, while a simple search for "data governance" by title in the Web of Science database yielded more than 300 results. The seven articles mentioned prove the lack of a comprehensive reference model for data governance in cities and provide no comprehensive overview of how cities currently deal with data or how they structure their data landscape. The evidence these articles uncover is based on either secondary or primary data with only a few cases. Another limitation is that these studies took place only in a particular limited region (e.g., covering only German cities or cases from Australia) or their surveys were only for the participating cities of a project, thereby providing evidence only for the duration of the project. All these studies, with their particular scope, provide valuable insights to push research on urban data governance forward.

The current work aims to contribute to theory and practice by expanding the scope and carrying out a more comprehensive primary data collection than the other studies presented here. Furthermore, it does not build on an existing smart city project in which case studies are considered but rather aims to determine the current state of data governance in cities in an independent manner, learn about the data infrastructure of cities, and understand the challenges that cities have in dealing with data. Thus, this study addresses the gaps in (1) a general understanding of the current state of data governance in cities, (2) identification of challenges cities face in dealing with data and factors for the successful implementation of data governance in cities, and (3) creation of a foundation for the development of actions.

## III. RESEARCH DESIGN

This study is part of a DSR project (Fig. 1) in which the results of the following empirical analysis are used to develop an urban data governance reference model.

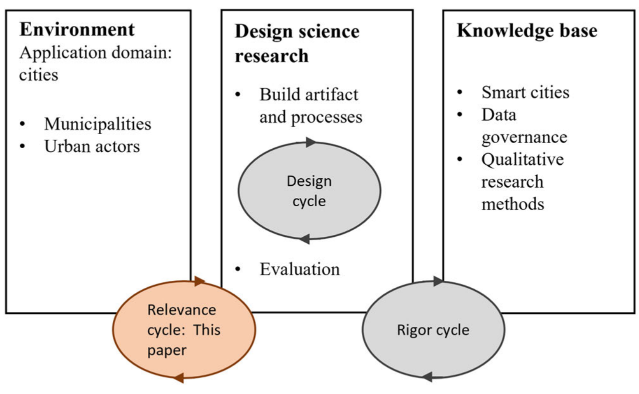



**FIGURE 1. DSR cycles (adapted from [29], [30]).**

The DSR paradigm belongs to design-oriented research, in which the output is an artifact (i.e., reference model, framework, or process) that aims to contribute to the problem environment [30]. Hevner [29] defines three cycles of DSR that must be identifiable in any DSR project. The relevance cycle connects the application domain (environment) with the design science activities of the design cycle (e.g., drawing requirements), while the rigor cycle integrates scientific rigor into the design activities through the scientific foundations, experience, and further theoretical foundation of the knowledgebase. This research aims to gain insight into the problem space and to define the problem's scope to be addressed. It also identifies the requirements for the artifact. Therefore, it is an activity within the relevance cycle highlighted in Fig. 1.

### A. RESEARCH METHODS

Semi-structured expert interviews were conducted to build the database for answering the RQs identified in Section I; a qualitative research approach is appropriate in the exploratory phase of knowledge building. Fink [62] outlines aspects of when qualitative methods are suitable, with the following points applying to this study: (1) the study extracts the knowledge and opinions of experts in a particular field, (2) the study does not intend to limit the formulations of experts by using free formulations, and (3) given the lack of prior knowledge on the topic, use of a standardized survey is infeasible. Especially in a DSR context, interviews can help identify requirements for the solution space. In addition, interviewees are located in the problem space and thus represent the stakeholders of the problem. In this way, the problem space can be correctly addressed before designing the artifact [63]; that is, the design principles are derived from the interviews. The interviews concentrated on open-ended questions, which induced the experts into a narrative mode and left the interviewer enough flexibility to react to unforeseen information. Although the interviews were flexible enough to cover all informants' thoughts, they had a basic structure to make the results of several interviews comparable. To this end, an interview guide was developed to conduct the interviews in a focused and structured way and ensure thematic comparability. The interview guide ensures that no important aspects are left out of the interviews [64]. The thematic focus of the interview guide was developed toward answering the three RQs and is based on a previous systematic literature review [22], related studies [5], [24], [27], [65], and cities' strategies [13], [16], [66], [67].

Furthermore, the study applied structural coding [68], which is particularly well suited as a first-level coding approach, as it forms the basis for further analysis by categorizing the interview into larger segments according to the themes of the interview guide—namely, "current status," "challenges," and "success factors"—which makes the complexity and volume of the interviews more manageable in the analysis [68]. These coded segments eventually formed the basis for creating the cognitive maps and answering the RQs. Moreover, according to Guest et al. [69], MacQueen and Guest [70], and Namey et al. [71], structural coding is highly suitable for standardized or semi-structured data collection (i.e., for interview transcripts of semi-structured expert interviews).

Last, the study applied cognitive mapping [72] for data analysis. A cognitive map makes the essential part of the interview data visually accessible on a single sheet of paper. The use of arrows, which represents the connection, avoids fragmentation of the interview into individual statements isolated from their context and preserves the overall complexity of the contexts expressed by the interviewee. Organizing the data according to these contexts' logic makes it possible to connect conceptually related statements from different interview parts. An arrow from a concept indicates a consequence, and an arrow into a concept indicates an explanation. Thus, each arrow gives an explanatory meaning to one concept and a consequential meaning to another. A negative sign at the head of an arrow indicates negative causality [73], [74]. For better understanding and clarification, section III, part D, explains the described principles of cognitive mapping using an example. Cognitive mapping goes back to George Kelly's theory [75] that one forms a picture of the world to predict what the world will look like and decide how to act or intervene to achieve what one prefers in that world [72], [73]. It is a problem-centered approach that reveals the complexity of the domain and highlights the relationships and interdependencies between concepts [76]. The maps help understand the current state of data governance in cities, such as what concepts exist and how they interact with each other. In addition, challenges become visible by understanding the origin and what situations they cause. Again, the aim is to understand which factors the experts consider necessary and how they can change the current state and address the challenges.

### B. RESEARCH PROCESS

Following the groundwork for the research methods in the previous section, Fig. 2 provides an overview of the research process described in more detail. The first task deals with the selection criteria to choose the cities for empirical analysis. The first criterion was that the cities were in the European Union (EU), to have a comparable group of cities regarding their political and administrative orientation, structures, and processes. The second criterion was that the city was active in the smart city context, for example, by publishing smart city strategies or taking an active role in implementing smart city and digitalization projects. For a systematic screening, the EU Digital Economy and Society Index (DESI) [77] served as a guide. The DESI annually assesses the digital performance of EU countries in connectivity, human capital, use of internet services, integration of digital technology, and digital public services. The focus was on two sections within the DESI: the top four countries (Finland, Denmark, the Netherlands, and Sweden [77]) that provide essential insights into data governance through their advanced experience in digitalization and the median of DESI in Germany and Austria, as these countries have taken up the challenge to handle data through smart city projects and digitalization projects in administration. This division of the sample ensures the spread and diversity of digital maturity levels. As a result, more universal insights can be gleaned from the interviews in total. Finally, being affiliated with a German university, we have direct contact with German cities, which speeds up the acquisition process.

The next step was to develop an interview guide addressing (1) an introduction to the tasks and position of the interviewee; (2) the strategic direction of the city regarding digitalization and data handling; (3) the current state of the city regarding data infrastructure, data sharing, and data governance mechanisms; and (4) the challenges in implementing data governance and data handling, as well as discussing the factors required for successful data governance and a data-driven city.

The guide was reviewed by the research team and colleagues in the research institute to ensure that the questions were understandable and to prevent any misunderstandings. The appendix presents the interview guide.

As the framework of the study was completed, the ethics board was approached to provide an ethical review of the study. The ethics committee reviewed the application in terms of the study objectives, study participants, and data collection and analysis. An informed consent form and a participant information form were also developed. Both documents were reviewed by the ethics committee as well. An ethical review ensured that the study, in particular the data collection and analysis, complied with ethical principles and alleviated any privacy and other concerns.

After completion of the interview preparations and approval from the ethics committee, interview partners were acquired. The first contact was by email, which explained the motivation for the study, its aim, and the conditions for the interview. The purpose of the data use and processing was also explained. After receipt of a positive response, the formalities were clarified and the interview guide sent. Having the interview guide in advance ensured that the person contacted had the necessary qualifications to answer the interview questions. Only in a few cases, after the person initially contacted received the interview guide, were other officers who were more qualified for the interview referred. At the time of the interviews, more interview partners were recruited until theoretical saturation of the results was reached [78, p. 62]. The total number of cities interviewed was 27 of the 85 contacted, resulting in a performance rate of 32%. In addition, some interviews were conducted with more than one informant from each city simultaneously, providing the necessary expertise for the interview. The interviews were conducted online and lasted an average of 42 minutes, resulting in more than 19 hours of interview recordings.

Table 1 summarizes the interview acquisition. At the beginning of the interview, respondents were informed about their participation in the study, asked for their consent to

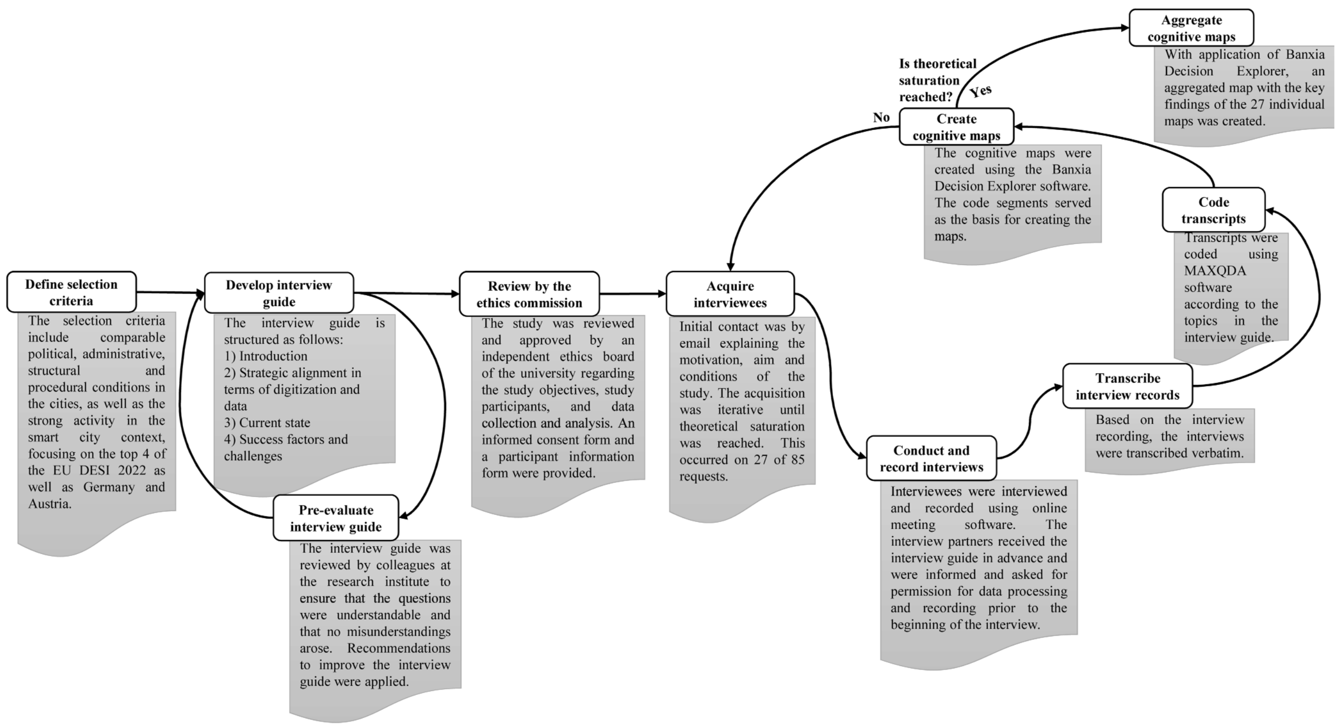



**FIGURE 2. Research process.**

**TABLE 1. Overview of the interview acquisition.**

| | Germany | EU [a] | Total |
|---|---|---|---|
| **Contacted cities** | 45 | 40 | 85 |
| **Conducted Interviews** | 19 | 8 | 27 |
| **Declined** | 5 | 3 | 8 |
| **Ignored** | 21 | 29 | 50 |
| **Interviewees** | 23 | 10 | 33 |
| **Average duration in min.** | 42,68 | 41,75 | 42,22 |
| **Interview language** | German (18), English (1) | English | German (18), English (9) |

[a] Austria, Denmark, Netherlands, Finland, and Sweden.

participate and to be recorded, and given the opportunity to ask questions about the study.

After the formalities were completed, the interview was conducted using the interview guide, and an audio recording was made. The interview recordings were transcribed immediately after the interviews. Subsequently, the transcripts were coded by the author team using the structural coding approach with the MAXQDA software [79], and the cognitive maps were created with the Banxia Decision Explorer [80]. No special software configuration other than the default configuration was made for both tools. The continuous analysis allowed for determining when theoretical saturation had occurred to close the interview phase and create an aggregated map with the key findings from the 27 individual interview maps.

## C. SAMPLE DESCRIPTION

As Table 2 shows, German cities accounted for 70.38% of the interviews, Dutch cities for 11.11%, Finnish cities for 7.41%, Danish cities, Austrian cities and Swedish cities for 3.7%. To avoid tracing back which city participated in the interviews, the population is given as a range. All cities have a smart city strategy or a digitalization strategy, but only seven cities have a stand-alone data strategy. In the other cases, the strategic direction on data was either included in the smart city/digitization strategy or was indicated as the next agenda point to develop a data strategy for the city. Except for five interviews, all were one-on-one interviews with only the interviewee and the researcher present. In five interviews, two or three representatives of the city were present, marked (A), (B), and (C) in Table 2. In these combined interviews, particular attention was paid to ensure that each interviewee had enough time to respond and contribute equally. Fifty-six percent of the interviewees hold a managerial position in the city. Among the interviewees, 48% have the term "data" in their job titles. Even if some interviewees do not have it in their job title, all deal with data daily, whether strategically using data or operationalizing data processes in day-to-day business. There is only one chief data officer and three concrete job titles for the data governance lead role. The fact that nine interviewees have the title "smart city manager" shows that the data issues have been assigned to the smart

**TABLE 2. List of conducted expert interviews.**

| # | Country code | Population range | DS * | Position of interviewees | Yrs. in position |
|---|---|---|---|---|---|
| 1 | AT | 200,000 - 250,000 | X | Head of statistics and research | 7 |
| 2 | DK | 40,000 - 50,000 | | Program leader smart cities | 5 |
| 3 | FI | 250,000 - 300,000 | X | Development manager for data | 2 |
| 4 | FI | 200,000 - 250,000 | X | Program manager data-driven city for citizens | 1 |
| 5 | DE | 50,000 - 100,000 | | Chief digital officer | 2 |
| 6 | DE | 450,000 - 500,000 | | IT Architect | 3 |
| 7 | DE | 50,000 - 100,000 | | IT Architect | 7 |
| 8 | DE | 50,000 - 100,000 | | Head of business development and smart cities | 4 |
| 9 | DE | 300,000 - 350,000 | | (A) Chief digital officer, (B) data manager | (A) 3, (B) 1 |
| 10 | DE | 100,000 - 150,000 | | Chief digital officer | 4 |
| 11 | DE | 750,000 - 800,000 | | (A) Chief digital officer, (B) smart city manager | (A) 4, (B) 1 |
| 12 | DE | 550,000 - 600,000 | | Smart city manager | 7 |
| 13 | DE | 1,000,000 - 2,000,000 | | Data governance lead | 2 |
| 14 | DE | 50,000 - 100,000 | | Chief digital officer | 3 |
| 15 | DE | 100,000 - 150,000 | | Smart city and urban data manager | 4 |
| 16 | DE | 300,000 - 350,000 | | (A) Head of IT, (B) smart city manager | (A) 5, (B) 1 |
| 17 | DE | 150,000 - 200,000 | | Data manager | 4 |
| 18 | DE | 550,000 - 600,000 | | (A) Geodata manager, (B) program leader smart cities | (A) 7, (B) 4 |
| 19 | DE | 50,000 - 100,000 | X | Data manager | 2 |
| 20 | DE | 1,000,000 - 2,000,000 | X | Smart city and urban data manager | 6 |
| 21 | DE | 600,000 - 650,000 | | Chief digital officer | 3 |
| 22 | DE | 250,000 - 300,000 | X | Data manager and data governance lead | 2 |
| 23 | DE | 150,000 - 200,000 | | Smart city manager | 3 |
| 24 | NL | 150,000 - 200,000 | | Chief digital officer | 1 |
| 25 | NL | 650,000 - 700,000 | | Program manager "Digital City" | 2 |
| 26 | NL | 550,000 - 600,000 | X | Chief data officer | 4 |
| 27 | SE | 100,000 - 150,000 | | (A) Head of data department, (B) data governance lead, (C) GIS manager | (A) 2, (B) 1, (C) 5 |

*DS: Data strategy

city activities. Interviewees have been in their current position for an average of 3.39 years. Nevertheless, all have many years of professional experience, in some cases more than 20 years.

### D. CODING AND MAPPING

After transcribing the interview using the audio recording, the transcribed data were used for the analyses. To make the large amount of data accessible, the text segments in the categories of the interview guide were coded by the author team, which corresponds to the structural coding approach. This made the relevant text segments directly accessible and ready for mapping. Use of coding and cognitive mapping helped model text passages that fit together in terms of context but may have been mentioned at different stages of the interview into a complex coherent system. For example, interviewees may refer to one concept by another concept, which does not necessarily occur at the same stage in the interview but may have a contextual impact on each other. For the best comprehensibility of the map, the concepts were color-coded: "current state" in green, "challenges" in yellow, and "success factors" in blue. Furthermore, red arrows indicate negative impacts and the green arrows positive impacts.

The following is an example to explain in more detail how the maps were created from the interviews. It should be noted that the given example is taken from a single interview and not from the aggregated map to gain a simplified illustration. Table 3 shows the three codes "Developing a data catalog," "A poor overview of the available data exists," and "Data are isolated and not connected, resulting in 'data silos'" along with their text passages.

**TABLE 3. Examples of text passages and how they were coded and categorized.**

| Text passage | Code | Category |
|---|---|---|
| "Another important point is perhaps the analysis of the current situation. What data do the city have at all? Perhaps in the form of a data dictionary or similar." | Developing a data catalog | Success factor |
| "When I look at my projects, the challenge is first to know who has what data in what format? Because there is no central information about it. You actually have to ask around individually. There is no central place that knows what data are available in which department." | A poor overview of the available data exists | Challenge |
| "However, it turns out that we, like probably many other cities, have different data available, but they are stored in department-specific data silos, and there is no central approach to get an overview of all the data." | Data are isolated and not connected, resulting in "data silos" | Current state |

First, the codes are inserted into the cognitive map. The second step is to model the relationships and interactions between these codes. For the code "Data are isolated and not connected, resulting in 'data silos,'" the statement "there is no central approach to get an overview of all data" means that data silos leads to the challenge of a poor overview of the assets. Moreover, the text passage of the code "A poor overview of the available data exists" suggests that a silo situation exists in which “there is no central place that knows what data are available in which department.” At a later stage, the interviewee mentions that an important point to consider for the future is the analysis of a city’s current state—that is, a data inventory, which the interviewee refers to as a “data directory,” that shows what data are available in the city. The text passage in the code “A poor overview of the available data exists” suggests that the challenge can be met with a central place to determine which data are available in which area. Thus, the data catalog’s negative impact on the challenge “A poor overview of the available data exists” is represented by a red arrow with a negative sign. This is because, on the one hand, a data catalog provides an overview of the existing data assets. On the other hand, data silos promote the challenge of poor overview of data. Although this example is relatively brief and consists of only three concepts, it contains complex relationships. So, the strength of cognitive maps is that complex situations can be represented on a cognitive map in an accessible way, as in Fig 3.

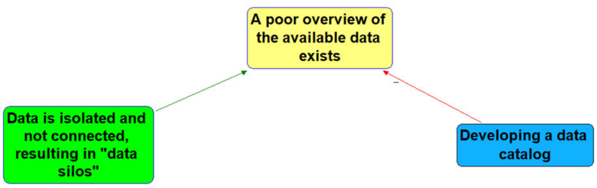



**FIGURE 3.** Example of combining concepts from a single interview.

### E. AGGREGATING MAPS

Following the aforementioned procedure, all 27 interviews were processed and 27 cognitive maps created. Then, these individual maps were combined into an aggregated map. The aggregated map synthesizes separate maps representing information from the individual interviews; it no longer represents the cognition of a single interview. Lavin and Giabbanelli [81] describe an aggregated map as a mental model of a group of interviewees. Thus, an aggregated map allows the drawing of collective conclusions about a group of interviewees, which is an ideal way to answer RQs. For this purpose, the individual maps are iteratively analyzed for their common concepts, and patterns mapped on the aggregated map are formed accordingly. Often, different experts formulate the same ideas and relationships using different terms. It is almost impossible for any two individual maps to be exactly the same, but often common ideas are the basis for aggregation. The aggregation standardizes the concepts of the individual cognitive maps, which are then combined in the aggregated map to form common labels. In addition, relationships between concepts, for example, are not always recognized by all experts, so others enrich this missing information in the aggregated map [73], [76], [82]. A first step entailed listing all the concepts and standardized synonyms, and the second involved analyzing the relationships between the concepts in the individual interviews and gradually integrating them into the aggregated map. To reduce researcher bias, the maps were aggregated within the research team.

## IV. RESULTS

This section presents the results of the analyses. The structural analysis of the aggregated map is first discussed by listing the concepts of the map and depicting it. Then, an in-depth content analysis of the map is conducted to answer RQ1–RQ3 in subsections B, C, and D. Here, all the states and findings were extracted from the expert interviews.

### A. STRUCTURAL ANALYSIS

The aggregated map in Fig. 4 contains 71 concepts, 19 of which represent the current state, 27 relevant challenges, and 25 important success factors. At first glance, it seems that some concepts (e.g., “Data are isolated and not connected, resulting in ‘data silos’”) are more strongly linked than others (e.g., “Increased acceptance of data governance measures”). This is a first indicator of concepts that can be viewed as central elements of the system and where changes have a large-scale impact on the system. Therefore, special attention should be paid to them for decision-making. The link among the concepts is usually such that success factors have a negative impact on challenges. However, challenges can also hinder success factors, which is why there is a negative link between these concept types. All other concept types are linked in a context-specific manner. For example, a challenge can negatively affect a concept of the current state—“reluctance to publish/share data” has a negative impact on the city’s current “Publishing open data” activities. By contrast, the circumstances of the current state can lead to a challenge linked by a green arrow; for example, the concept “Data are isolated and not connected, resulting in ‘data silos’” can complicate the data integration and sharing.

Having 71 concepts in the map challenges the manual analysis of the map. Therefore, the Banxia Decision Explorers analysis tools (i.e., identifying driver concepts, output concepts, centrality analysis, and domain analysis) were used, which resulted in Table 4. The analysis of concept types is particularly well suited for answering decision-making questions [76]. Thus, which concepts are drivers (i.e. concepts that have only outgoing arrows) can be examined. These concepts stimulate (connected with a green arrow) and hinder (connected with a red arrow) other concepts. The outputs (output concept) have only incoming arrows. These concepts result from (connected with a green arrow) and are weakened by (connected with a red arrow) other concepts. Finally, the mixed concepts have incoming and outgoing arrows and influence the other concepts through their connection according to the logic mentioned previously. In this way, how specific interventions affect the entire network and how a particular concept results from the interconnection of the concepts by following the chain of linkages can be analyzed. In addition, structural analyses can help determine which

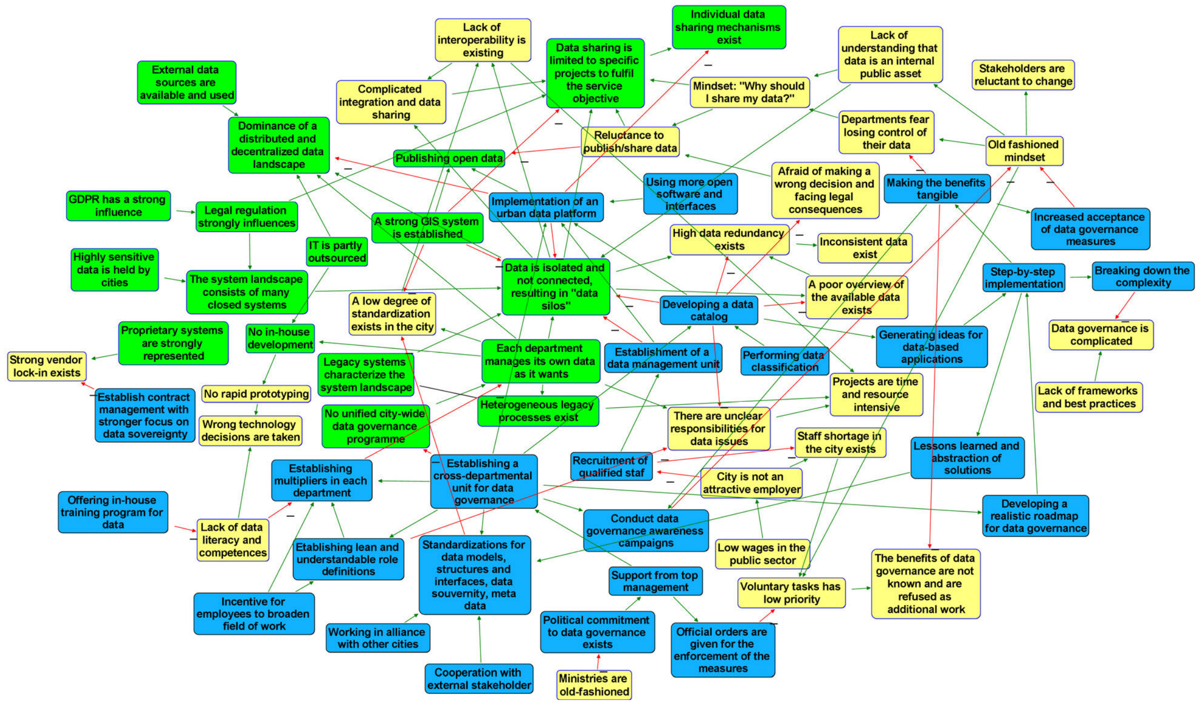



**FIGURE 4.** Aggregated map.

**TABLE 4.** Concepts of the aggregated map.

| Type | Category | Concepts of the aggregated map |
|---|---|---|
| Driver concepts (16) | Current state (7) | A strong GIS system is established, Highly sensitive data are held by cities, Legacy systems characterize the system landscape, External data sources are available and used, Proprietary systems are strongly represented, IT is partly outsourced, GDPR [General Data Protection Regulation] has a strong influence |
| | Challenge (3) | Low wages in the public sector, Lack of frameworks and best practices, Ministries are old-fashioned |
| | Success factor (6) | Performing data classification, Offering in-house training program for data, Cooperation with external stakeholders, Incentive for employees to broaden the field of work, Using more open software and interfaces, Working in alliance with other cities |
| Output concepts (11) | Current state (3) | Dominance of a distributed and decentralized data landscape, Publishing open data, Individual data-sharing mechanisms exist |
| | Challenge (8) | The benefits of data governance are not known and are refused as additional work, Projects are time- and resource-intensive, Strong vendor lock-in exists, A poor overview of the available data exists, Stakeholders are reluctant to change, Inconsistent data exist, Data governance is complicated, Wrong technology decisions are taken |
| Mixed concepts (44) | Current state (8) | The system landscape consists of many closed systems, Data sharing is limited to specific projects to fulfill the service objective, Data are isolated and not connected, resulting in "data silos," Each department manages its own data as it wants, Heterogeneous legacy processes exist, Legal regulation strongly influences, No unified city-wide data governance program |
| | Challenge (17) | Afraid of making a wrong decision and facing legal consequences, City is not an attractive employer, Complicated integration and data sharing, Departments fear losing control of their data, Lack of data literacy and competencies, Lack of understanding that data are an internal public asset, A low degree of standardization exists in the city, Mindset: "Why should I share my data?", Lack of interoperability is existing, No in-house development, No rapid prototyping, Old fashioned mindset, Reluctance to publish/share data, There are unclear responsibilities for data issues, Voluntary tasks have low priority, Staff shortage in the city exists, High data redundancy exists |
| | Success factor (19) | Increased acceptance of data governance measures, Conduct internal and external data governance awareness campaigns, Breaking down the complexity, Establish contract management with stronger focus on data sovereignty, Establishing a cross-departmental unit for data governance, Developing a data catalog, Establishment of a data management unit, Generating ideas for data-based applications, Establishing lean and understandable role definitions, Lessons learned and abstraction of solutions, Making the benefits tangible, Official orders are given for the enforcement of the measures, Political commitment to data governance exists, Developing a realistic roadmap for data governance, Recruitment of qualified staff, Standardization for data model structures and interfaces for data sovereignty metadata, Establishing multipliers in each department, Step-by-step implementation, Support from top management, Implementation of an urban data platform |

concepts have a strong and less potent effect on the system by evaluating the degree of linkage of a concept.

Table 4 is structured according to the concept types "driver concept," "output concept," and "mixed concept." These overarching types are then subdivided into the categories "current status," "challenge," and "success factors," and finally the specific concepts in each category are listed. The mixed concepts are the most represented—this means that the concepts are strongly interwoven and have complex interdependencies, as they are drivers for other concepts but are also driven by other concepts.

The current state represents the concepts that describe the current reality in the city (i.e., in the areas of the data landscape, data sharing, data management, and data governance, such as "Highly sensitive data are held by cities"; "Data are isolated and not connected, resulting in 'data silos'"; "Dominance of a distributed and decentralized data landscape"; or "Each department manages its own data as it wants").

The concepts in the challenges category indicate the hurdles cities face when working with data and managing and governing data (e.g., "The benefits of data governance are not known and are rejected as additional work"). The challenges are diverse and mainly belong to the type of mixed concepts. This means that the challenges in a city often lead to different outcomes; however, they do not stand alone but are also caused by certain circumstances in the city. The map explains what causes a particular challenge, what situation the challenge leads to.

Finally, concepts in the success factors category describe what is necessary to be successful with data governance and to be more data-driven (e.g., "Offering in-house training program for data"). Success factors are either a mixed concept or a driver concept.

To answer the RQs in this study adequately, the map depicted in Fig. 4 was broken down into smaller parts by first presenting the overall domain map of the current state, challenges, and success factors. This is followed by a focused analysis of the most central five concepts of each (listed in Table 5), as discussing all 71 concepts would exceed the length of this article. Summarizing the map in this way is appropriate to organize the analysis of such large maps [83]. For this purpose, Banxia Decision Explorer helped calculate the central concepts. Each concept is scored for its degree of linkage; the more strongly a concept is linked, the higher is the value. However, not only is the first degree of linkage taken into account, but the indirect linkages across several levels are also included in the calculation to determine the central concept in the entire system and not only locally limited, strongly linked concepts, with the calculation factor decreasing with each subsequent level [84].

### B. CURRENT STATE IN CITIES

To answer RQ1, this subsection aims to explore how cities are organized in terms of their data ecosystem to understand the current data environment better. With respect to the current state, a filter is applied to the aggregated map that shows only the concepts of the current state, as shown in Fig. 5. The current state of data infrastructure, data exchange, and data governance in the city can be briefly described as follows: the system landscape in cities is characterized by many specialized applications that are outdated, closed and proprietary, and often outsourced. Many of the interviewees noted the GIS as they contain a great deal of data, and the city's affinity for data is mainly in these departments. Moreover, the data infrastructure is very distributed and decentralized. Many experts noted that some data pools are redundant. Data silos characterize the general picture of the data infrastructure in cities. This also affects data exchange, such that data are mostly exchanged only to fulfill the legal administrative mandate and individual, project-dependent ways, from automated application programming interface (API) calls to sending flat files by email. The departments have their data governance mechanisms and approaches for data management and exchange based on the legal parameters and the GDPR, as no city-wide data governance program provides city-wide guidelines and policies.

**TABLE 5. Most essential concepts in the aggregated map.**

| Category | Concept |
|---|---|
| Current state | (1) Data are isolated and not connected, resulting in "data silos"<br>(2) Each department manages its own data as it wants<br>(3) Data sharing is limited to specific projects to fulfill the service objective<br>(4) Dominance of a distributed and decentralized data landscape<br>(5) No unified city-wide data governance program |
| Challenges | (1) Lack of understanding that data are an internal public asset<br>(2) There are unclear responsibilities for data issues<br>(3) Lack of interoperability is existing<br>(4) A low degree of standardization exists in the city<br>(5) High data redundancy exists |
| Success factors | (1) Establishing a cross-departmental unit for data governance<br>(2) Developing a data catalog<br>(3) Implementation of an urban data platform<br>(4) Establishing multipliers in each department<br>(5) Establishment of a data management unit |

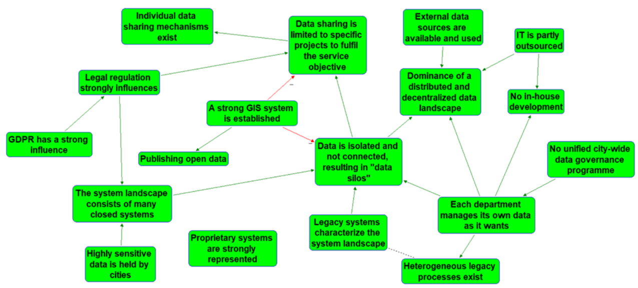



**FIGURE 5. Section of the aggregated map focusing on the "current state."**

Next, analyzing the relationships gives deep insight into the central concepts of the current state. For this purpose, the central concepts were calculate with Banxia Decision Explorer (for calculation details, see part A of section IV and [84]).

The first concept, "Data are isolated and not connected, resulting in 'data silos,'" has the highest centrality, not only in the current state concepts but also in the whole map (depicted in Fig. 6). Interviewees describe that cities—with few exceptions—store a high amount of data in their domain-specific vertical data silos. The data silos are so structured that little to no interaction exists between the departments holding the data. These data silos are technically designed to operate independently; therefore, metadata and database structures are defined exclusively within these silos [85], [86], [87]. These vertical silos perform well within their scope and fulfill their task. However, combining data from different siloed databases has a high potential for combined analysis, which is a major challenge due to the poor interoperability of these silos. These siloed structures are found not only in legacy administrative systems but also in smart city solutions, such as smart parking, smart lights, and so on. A closer review of the map shows that the reason for the data silos in the cities is partly cultural, so there is simply no understanding that urban data are a common good at all, and partly technological, because some systems are still old and cannot communicate given the lack of interfaces. In addition, legal requirements prevent the exchange of data (e.g., health, religious, and police data are protected specially). Therefore, these departments keep their data in separate systems. As a result, data are infrequently shared and only then to fulfill the mandate of the service in just a few projects. Such projects are highly complicated due to the lack of interoperability. In addition, many data silos represent a "black box" for a holistic view of the city's data assets, which means a relatively poor overview of the data and leads to redundant data assets. Such are the challenges arising from data silos as a concept. To break down the data silos, the interviewees mentioned, for example, developing a data catalog, which goes hand in hand with establishing a city data platform facilitated by a data management unit. In addition, GIS is a key enabler in overcoming data silos as it centrally provides a great deal of shared data between departments in a standardized form.

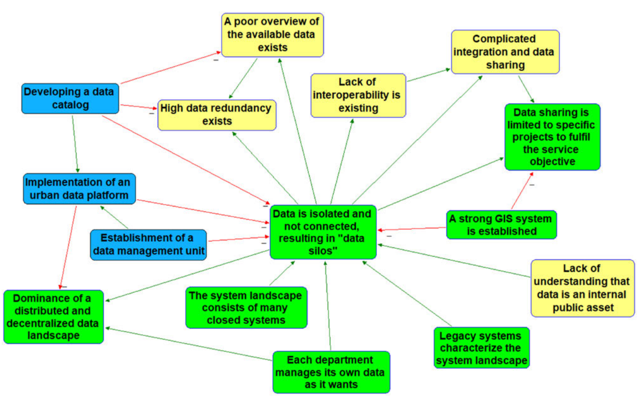



**FIGURE 6. Outline of the concept of "Data are isolated and not connected, resulting in 'data silos'" and its relationship to other concepts.**

The second central concept is that each department manages its own data as it wants (Fig. 7). As a direct consequence, data silos are emerging. A city has several departments, depending on its size and structure, that usually manage their data in their own way. They decide on applications, interfaces, and data models. In addition to data silos, the departments act independently, leading to a distributed and decentralized data landscape and heterogeneous and outdated processes. The greatest challenge, however, is that a city has little standardization. To counteract this, the experts suggested that so-called multipliers be designated to each department. These multipliers are appointed in their department and form a city-wide community integrated into the cross-sectional unit of data governance.

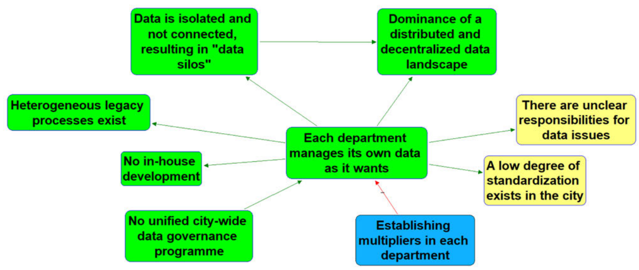



**FIGURE 7. Outline of the concept of "Each department manages its own data as it wants" and its relationship to other concepts.**

The third key concept is that data sharing is mainly limited to specific projects to fulfill the service mandate (Fig. 8). Highly individual data exchange mechanisms are used, ranging from automated API calls and direct access to databases to sending PDF files via email. One reason for this, however, is that legal conditions constrain data exchange. In most cases, however, the legal framework is used as an excuse for the lack of understanding of data sharing. Often, not even non-personal data are shared for cross-departmental analysis by the responsible department. The mindset in cities tends to be, "why should we share data?" The benefits of sharing and open data are often still unclear, and people are afraid that by sharing their data, they might lose control over their data or face consequences due to, for example, the law or poor data

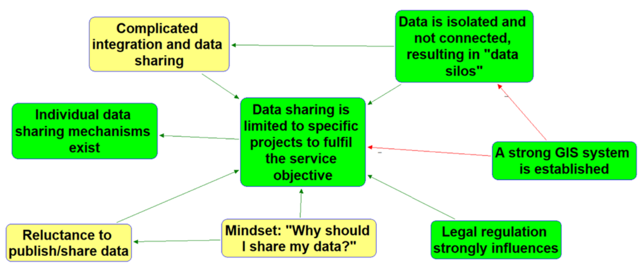



**FIGURE 8. Outline of the concept of "Data sharing is limited to specific projects to fulfill the service objective" and its relationship to other concepts.**

quality. This generally leads to the challenge that data sharing is more limited and reserved and that data are not considered a common good, but rather as hard-won valuable data that only belong to one department exclusively.

Fourth, departments manage their data, leading to a highly distributed and decentralized data landscape in the city (Fig. 9). This landscape is further strengthened by external data sources, such as industry, federations, and municipal subsidiaries, but also by the high degree of outsourcing, as cities often work with municipal service providers hosting multiple IT applications for the city. This also means that agile work, such as rapid prototyping, is impossible, resulting in poor technology decisions from which the city cannot benefit. To centralize the data landscape, the experts mentioned the development of an urban data platform, leading to a centralized approach, as a success factor.

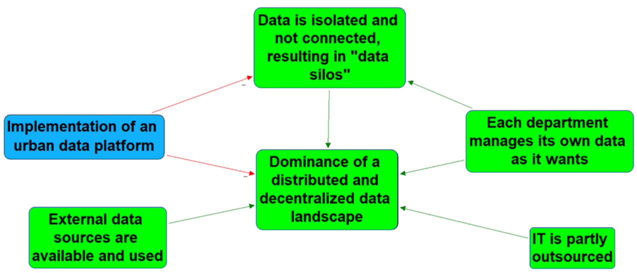



**FIGURE 9.** Outline of the concept of "Dominance of a distributed and decentralized data landscape" and its relationship to other concepts.

The fifth key concept in the current state is the lack of a unified city-wide data governance program (Fig. 10), which results in each department managing its data in its own way. This leads to the emergence of data silos (as described previously). Thus, the absence of a city-wide data governance program can lead to what some interviewees called "chaos" and a proliferation of data silos, applications, and hidden data that are not interoperable. With the help of a cross-cutting unit for data governance, the experts agreed that these successive disadvantages can be tackled by introducing a city-wide data governance program.

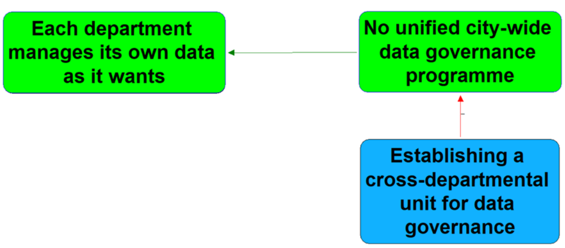



**FIGURE 10.** Outline of the concept of "No unified city-wide data governance program" and its relationship to other concepts.

### C. CHALLENGES

To answer RQ2, this subsection analyzes the cities' data-related problems and challenges in the same way as described

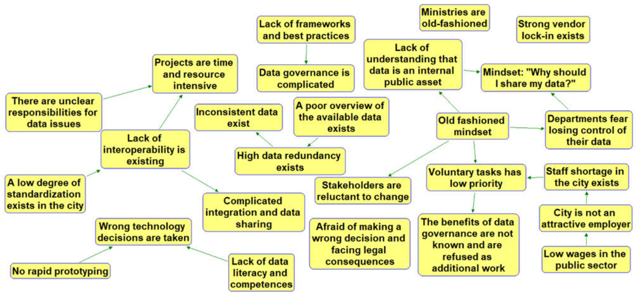



**FIGURE 11.** Section of the aggregated map focusing on the "challenges."

before. Fig. 11 shows the section of the aggregated map focusing on challenges. The interlinked challenges have a positive causal relationship. For example, an "old-fashioned mindset" leads directly to the challenges of "Stakeholders are reluctant to change," "Voluntary tasks have low priority," "Departments fear losing control of their data," and "Lack of understanding that data are an internal public asset." However, in contrast with the "current state" concepts, there is less interconnectedness among the challenges. For example, concepts such as "Strong vendor lock-in exists" are only linked to concepts not stated as challenges. That means that the challenges are also often caused by the current state concepts.

The challenges of cities in working with data are multifaceted. The experts described technological challenges such as "lack of interoperability," cultural challenges such as "old-fashioned way of thinking," and also organizational challenges such as "lack of staff." Cities face the problem that data interoperability is poorly developed, which makes data integration projects complicated and time-consuming. Furthermore, unclear responsibilities promote the problem of integrating data. There is also a lack of an overview of data and many redundancies, leading to inconsistent data. An organization's culture strongly affects how individuals work and behave. Often the culture is historically developed and cannot be changed from one day to the next without considerable effort. Usually, an "old-fashioned mindset" still prevails in cities, which means that, on the one hand, making changes is not welcome because "things have been working that way for years" and, on the other hand, there is simply a lack of understanding that data are a common good, which is why they are not willingly shared. Data governance activities are seen as additional workload in daily business, and because they are often not mandatory, they are not performed or are performed poorly. The city does not seem to be an attractive employer for talented young people, so the culture change is also slow. The experts desired to hire more qualified employees to strengthen and increase human resources. After all, the lack of skilled staff results partly in wrong technology decisions and greater dependence on service providers. The cities also desire more support from the ministries, but even these are partially unable to support the cities on their way to becoming more data-oriented as they are also challenged by their old-fashioned processes and ways of thinking. Finally, the experts stated that the high complexity of data governance in the city is a challenge for them. They do not know where to start or which measures are necessary for a realistic road map—they lack best practices and frameworks.

The most central concept of the challenges is the ''Lack of understanding that data are an internal public good'' (Fig. 12). That data are not considered a city-wide asset in many people's minds is due to the ''old-fashioned mindset'' that leads to the current state of data silos with many consequences. Given the many consequences of data silos, as described previously, addressing this concept is a high priority, according to the interviewees. The lack of understanding also leads to data not being shared, ultimately driving silo thinking.

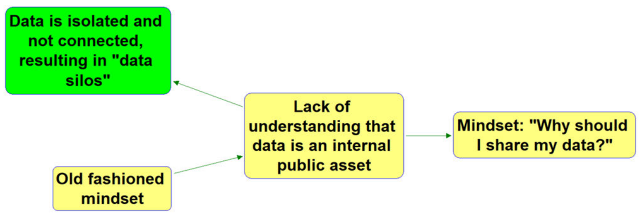



**FIGURE 12.** Outline of the concept of ''Lack of understanding that data are an internal public asset'' and its relationship to other concepts.

The next central concept is unclear responsibilities (Fig. 13). According to the interviewees, it is not obvious who is the right contact person for certain data matters. Often the responsibilities are not defined, and the information often lies in the heads of the individuals. For example, in smart city projects involving more than one department, novice city personnel often need to rely on the knowledge of an experienced colleague who, based on years of experience alone, knows who might be responsible for which data. This is not an optimal approach as, on the one hand, the know-how may be lost in the future; on the other hand, unclear responsibilities lead to slow processes and lack of quality, resulting in projects that inevitably exhaust more resources than is optimal. The lack of clarity in responsibilities is because the departments are currently completely independent in working with data, with no mutual agreement or overview. Therefore, the experts recommended the development of a data catalog as a success factor that can decrease the unclear responsibilities (shown in Fig. 13 as a negative causality). Furthermore, the need for

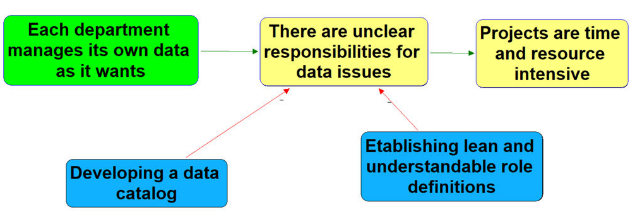



**FIGURE 13.** Outline of the concept of ''There are unclear responsibilities for data issues'' and its relationship to other concepts.

lean and understandable role definitions can counteract the problem of unclear responsibilities.

The lack of interoperability is the next key concept in the challenges (Fig. 14). Interoperability is not only technical. The lack of standardization affects all levels of interoperability, such as the organization, semantics, technology, and process. Thus, integration and data-sharing projects become highly complex and time- and cost-intensive. Data silos strongly drive the lack of interoperability through their isolated nature, so various data models have been built and used in isolated organizations, causing semantic interoperability issues. Interoperability is also weakened by the high heterogeneity of the processes and, in some cases, departments' mindset. Therefore, experts listed interoperability as a significant challenge.

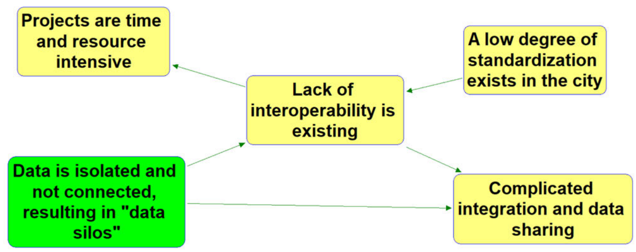



**FIGURE 14.** Outline of the concept of ''Lack of interoperability is existing'' and its relationship to other concepts.

The next concept is the low level of standardization (Fig. 15), which is also one of the causes of the lack of interoperability. Nevertheless, not all domains experience a low level of standardization. GIS again resides at the forefront in terms of the establishment of international standards such as the KML data format, INSPIRE metadata standards, and other services (e.g., WMS, WPS). However, standardization is a technical and procedural challenge. Suppose each city and each department interprets its processes differently and manages its data in a self-defined, individual way. In this case, cross-departmental or cross-city services can only be realized with great effort. As such, the experts noted a great need for standardizing sensor-based services, which are increasingly finding their way into cities as part of their smart city activities. Unfortunately, the providers of services and sensors define their own data models and interfaces, which are often incompatible with those of the cities and cannot

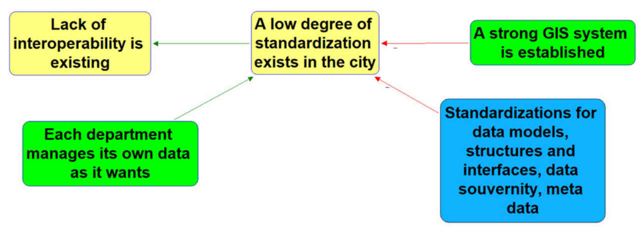



**FIGURE 15.** Outline of the concept of ''A low degree of standardization exists in the city'' and its relationship to other concepts.

be integrated with third-party tools for better data workflows. The experts viewed common open standards for data models, data structures, interfaces, and metadata as cornerstones and success factors to overcome the standardization challenge.

Finally, the data have a high level of uncontrolled redundancy (Fig. 16). This challenge stems from the existing data silos and the absence of data overview, which also derives from the data silos. Most cities still have many black boxes and hidden data, so they do not know what data they have, as no central point is charged with bundling and describing the data. So, for projects, a great deal of effort is invested in determining what data exist, how they are structured, where they are located, and who the contact person is. A common occurrence is the recording of even the same data sets by several different departments, causing a high degree of redundancy, which, in the worst case, can lead to inconsistent data. For example, several departments record the same data sets in Excel files, and a single point of truth is absent. The experts agreed that a data catalog can help overcome this challenge by performing a data inventory and making it available to the city's stakeholders as a catalog.

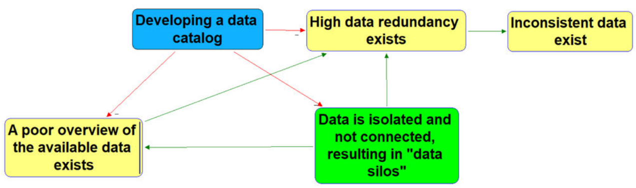



**FIGURE 16.** Outline of the concept of "High data redundancy exists" and its relationship to other concepts.

## D. SUCCESS FACTORS

Finally, the last category covers RQ3, which provides an understanding of the measures required for successful data governance. Fig. 17 shows a focused view of the success factors, or the concepts the experts outlined as being necessary to move forward with data governance in the city. These concepts are rather interconnected and often mutually beneficial. Typically, a success factor has such an output that either a challenge is targeted or a current state concept is affected. The experts indicated great potential for improvement in technology, organization, and culture. For example, a requirement is for an urban data platform with open software and interfaces set up by a data governance and data management unit. These units are also linked to the organizational aspect of recruiting qualified staff for these units and obtaining support from top management for the data governance activities. In addition, the motivational aspect is significant in the success factors. The experts recommended a realistic roadmap for a step-by-step implementation, as introducing data governance in a complex environment is not a task that can occur overnight. Instead, it is about breaking down the complexity; making the benefits of data governance activities understandable for everyone, especially by putting them into practice in the

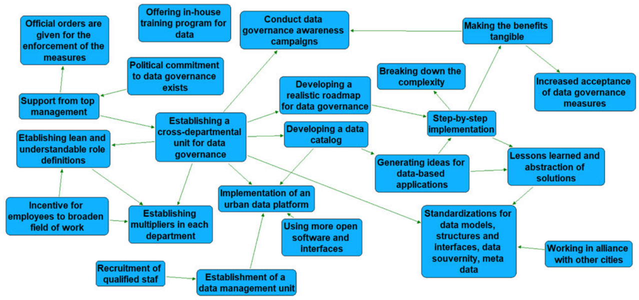



**FIGURE 17.** Section of the aggregated map focusing on the "Success factors."

daily business of individuals, which fosters acceptance; and using the lessons learned from such step-by-step developments to adapt the solution to other domains, which leads to city-wide standardization in the long run. Data applications usually do not stop at city borders, so many experts indicated a desire to work with other cities to promote cross-city standardization. They also viewed role definitions and responsibilities as success factors but agreed that these should be kept streamlined and understandable, especially to avoid challenging the old-fashioned mindset with fuzzy and new terms. Data governance activities cannot be performed alone, so so-called multipliers must be established in each department. The experts recommended that these people be trained internally and act as liaisons between their own department and the city-wide data governance and data management activities.

First, the concept with the highest central value in the success factors is the establishment of a cross-sectional unit for data governance (Fig. 18). The interviewees referred to this concept with different names, such as "data office," "data factory," "data house," and "urban data governance unit," but in actuality, it refers to a cross-sectional organizational unit that bundles the activities of data governance. Ideally, this unit consists of full-time employees who do not conduct tasks other than their work and who focus on data governance full-time. It is often the case that data governance activities are transferred to part-time staff. Thus, the establishment of such a unit can have a significant impact; to do so, resources and organizational structures of the city must be expanded, and the unit should be located centrally in an interdisciplinary

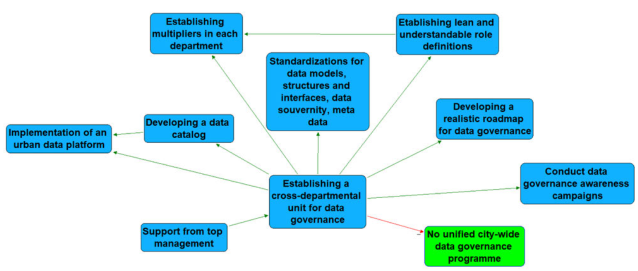



**FIGURE 18.** Outline of the concept of "Establishing a cross-departmental unit for data governance" and its relationship to other concepts.

office of the city. A major project such as this needs to attain enough backing from the city's top management, ideally with the political support of the city council.

This success factor can overcome the current state of "no unified city-wide data governance program," which, as mentioned previously in the description of the current state, is the cause of many concepts, such as each department managing its data according to its own policies and data silos creating significant challenges. The data governance unit must ensure a high level of multidisciplinary activities to address all stakeholders' needs. Therefore, multipliers from the individual domains are essential for its work; however, these multipliers do not develop on their own. They must be driven by the data governance unit, with the goal to elaborate understandable and lean role definitions.

As Fig. 18 shows, the data governance unit drives many other success factors. It is significantly involved in developing the data catalog and implementing the urban data platform; however, this unit cannot implement both activities alone and thus requires a multidisciplinary approach in collaboration with other stakeholders. The data governance unit is instrumental in the standardization process through its central role. This can be achieved by having a decision-making body within the data governance unit, alongside the data governance unit staff, with representatives from each city's departments and domains meeting regularly to define city-wide decisions on standards, guidelines, and processes. Finally, the challenges have exposed the impact of cultural barriers; therefore, awareness campaigns for internal and external stakeholders coordinated and organized by the data governance unit are necessary. Many measures are critical for the success of data governance in the city, bringing high complexity when a city wants to handle it all at once. As such, having a realistic roadmap aligned with the city's overall strategy toward a step-by-step implementation of data governance activities is essential.

The second key concept of the success factors is developing a data catalog (Fig. 19). Cities have a wide variety of data distributed in different data silos. The data differ in format, access rights, validity, location, and many other

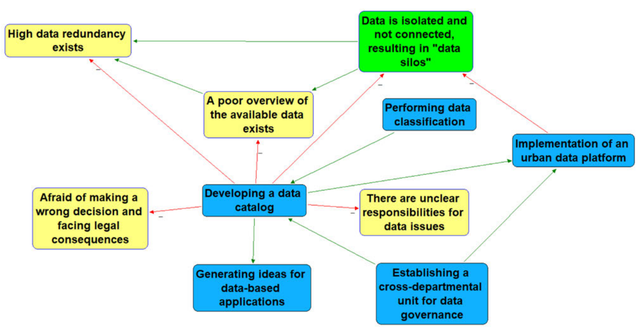



**FIGURE 19. Outline of the concept of "Developing a data catalog" and its relationship to other concepts.**

attributes. This mass of data, combined with the silo situation of the city, brings the challenge of a poor overview of the existing data, and responsibilities are unclear. With the help of a data catalog, this heterogeneous data can be classified, and its meta-information described. Broadly defined, a data catalog comprises an inventory of the data assets and thus enables the discovery, description, and organization of data sets through their metadata. Data catalogs make an essential contribution to the FAIR principles, which entail the availability, accessibility, interoperability, and reusability of data [88], [89].

Open data catalogs already exist in urban contexts. Still, they need further development by including internal data. Furthermore, the classification of a data set should be specified at the attribute level so that a personal attribute, for example, does not prevent the entire data set from being shared. Furthermore, the experts perceived great potential in including the data model of the data set in the data catalog, which would make working with the data set both technically and semantically easier. Finally, the data catalog is an essential building block for an urban data platform providing stakeholders with the necessary information.

The urban data platform is the third central concept of the success factors (Fig. 20). It is driven by data governance and the data management unit and based on open software and interfaces. Through an urban data platform, the cities hope for an improved and standardized data exchange in which the urban data platform functions as a data hub for internal, external, and open data. By linking individual data repositories of the departments to the urban data platform, data silos are avoided and a more centralized landscape achieved.

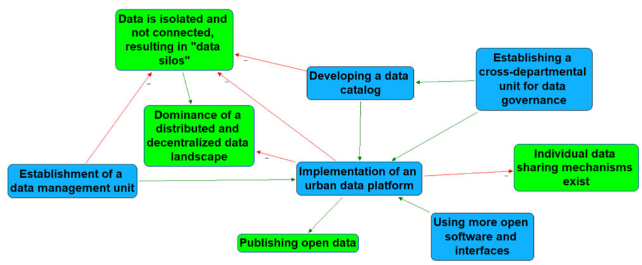



**FIGURE 20. Outline of the concept of "Implementation of an urban data platform" and its relationship to other concepts.**

The fourth concept of the central ranking in the success factors is the establishment of multipliers in the departments (Fig. 21). As noted, these multipliers are experts in their field who act as allies to the data governance unit in each department, driving the agenda of data governance issues in their departments and serving as a contact point for data issues within their team for colleagues. The multipliers provide an interface function between departments to prevent them from defining their own data governance and management, thereby creating shadow governance mechanisms. However, the experts emphasized the difficulty in

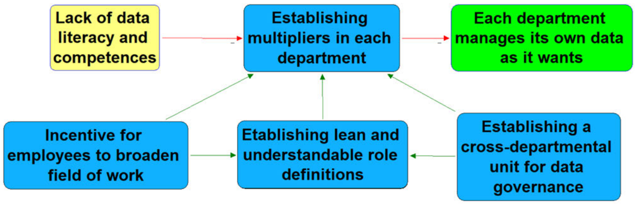



**FIGURE 21.** Outline of the concept of "Establishing multipliers in each department" and its relationship to other concepts.

establishing this role in cities because, first, data literacy must still be built up properly and, second, employees suitable for this role must be given the incentive to take it on by receiving benefits in the form of professional development and higher salaries for the expansion of their tasks. However, experts' description of the multipliers fits the role of a data steward [7], [8], [9], [54], [55], [56], though are not explicitly named as such.

Finally, the last core top five concept of the success factors is establishing a data management unit (Fig. 22). The difference between the data management unit and the data governance unit is that the former is more concerned with the implementation of data governance decisions. How the data management unit is set up depends heavily on a city's organizational structure and resources. For example, the experts suggested that the data governance and management units could be organized in teams under an umbrella unit that work closely together but are separate in terms of tasks. Nevertheless, such a data management unit requires recruiting skilled IT staff to build the competences in-house; even if implementation houses are brought in for larger implementation projects, the in-house expertise in the cities is indispensable.

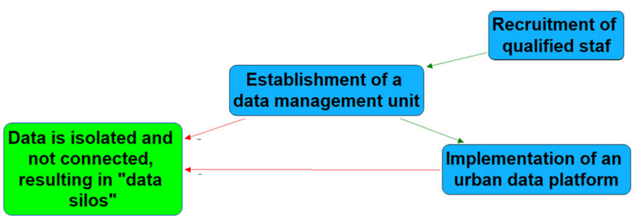



**FIGURE 22.** Outline of the concept of "Establishment of a data management unit" and its relationship to other concepts.

## V. DISCUSSION

The presented findings contribute to the knowledge base, fill the research gap targeted herein, and answer RQ1–RQ3 on the current state, challenges, and success factors. Compared with related works, this study portrays urban data governance in such a comprehensive view for the first time. As noted, each city department usually defines data governance independently, and no unified approach exists. The lack of standards leads to significant heterogeneity, forming a siloed data architecture in cities. Not only is the silo situation due to the technical nature of the architecture, but the systems are also isolated from each other and constrained by strong silo thinking embedded in the culture of a city. From technical and cultural perspectives, data sharing is difficult, and thus interoperability suffers. Ultimately, a city-wide data governance organization that has an overarching function and works closely with the other actors can help break down the silos in a city. The data catalog is an essential tool that contains the meta-information of data and standardizes the data models and data interfaces for high interoperability. However, required are not only technological measures but also considerable effort to convince, motivate, and educate city authorities to move toward a data-driven city with appropriate data governance. Data governance is not a one-time project with a fixed duration but an ongoing activity that aligns with the overall city strategy and comes to fruition through a firmly established secondary organization of data concerns. A realistic roadmap with progressively defined smaller goals is necessary to anchor data governance in the daily work of all actors through targeted incentives and tangible benefits.

The proper handling of data has recently gained prominence, but the development has so far been generic and detached from a city's overall strategy. Smart city initiatives have brought the aspect of handling data to the attention of cities. After all, data are not only an IT issue but also a business issue. However, it is also clear that a single team or department cannot introduce data governance for the entire city; instead, it requires firm conviction and support from the city leadership. Moreover, resources and structural changes in the organizational structure are necessary. Political initiatives, as described in the introduction, show that the relevance of data has also reached the top level, thus giving hope that cities will obtain the political tailwind to push their data governance initiatives. In all interviews, it was clear that the relevance of data governance is given a high priority but that the implementation of measures is at different stages.

This study provides a comprehensive understanding of urban data governance that yields theoretical contributions and practical implications. From a theoretical perspective, future research on urban data governance can use the findings of this study to develop a reference model for urban data governance. Beyond the given research gap regarding the reference model, the interviewees also clearly expressed the need for more guidance from the scientific community. Cities need significant support in designing their data governance practices. Although data governance frameworks already exist in the corporate context, the experts believed that a city cannot adopt these frameworks unchanged; instead, it requires a tailored tool in the form of a data governance reference model. Thus, this work provides a foundation for developing such a reference model by shedding greater light on the environment for such a model and delineating the challenges that need to be addressed. The cognitive maps show many complex causal relationships within the system, emphasizing the multifaceted nature of a city compared with a company, which should help future research examine the extent to which data governance concepts from companies can be adapted to cities. The results of this work not only provide the basis for developing a reference model for urban data governance but also have implications for the smart city body of knowledge by providing a better understanding of the data ecosystem. Thus, they are important for research on urban data platforms and data-driven urban planning and smart governance.

Finally, the problems and challenges identified highlight potential areas for future research. In addition, the practical implication of the findings is that cities can use them to define initial measures by comparing their situation with the current findings to identify their strengths and weaknesses and initiate targeted actions. In this process, the cognitive maps are a valuable aid, as they show the interdependencies of the concepts. This study should raise awareness and help incorporate data governance into the design of overall urban strategies. Practitioners can also use the findings in their future projects to ensure that data standards, interoperability, and data sharing are considered from the beginning and included in the design of procurement criteria. The findings may also help practitioners identify which data aspects they need to consider in their projects to avoid becoming further trapped in silos and to prevent additional challenges. Finally, this study signals to policy makers that cities need political and financial support to develop their data governance, which would ultimately benefit other projects.

This work also has a few limitations. First, 27 expert interviews were conducted, after which theoretical saturation was reached. Nevertheless, many interview requests were declined or ignored, thus possibly leaving out cities that are much further along in data governance and that could have contributed valuable additional insights to the work. Second, providing a detailed overview of the whole city is difficult for one person, so the interviews with multiple interviewees provided even more insights. However, exploring a city more intensively by interviewing several departments and stakeholders would also be worthwhile to understand the complex system better. Smart cities do not consist solely of city government representatives. They are much more complicated, with various stakeholders not directly sitting in the city government. Despite this, the interview partners had a good overview of the smart city activities and, thus, of the active stakeholders, making them qualified contacts. Third, the study used only qualitative data in the form of expert interviews, which may influence the data analysis through personal assessment. To minimize this bias, the research team discussed and agreed on the individual phases of the data analysis. Especially in cognitive mapping, not only one model describes a problem. Each map is unique and can be customized by a researcher. Therefore, there is no one ultimate map of a problem [72]. Thus, the research team evaluated the map to ensure its generality as much as possible. Despite the collection of a high amount of data, this study represents only a sample of the data needed to answer the RQs. Fourth, the data set also shows, that there are cities that are much more advanced in their data governance activities. These cities have a different cognitive map because they currently have a different setup; that is, they already have a data governance unit and face other challenges. However, these progressive cities also once had the same starting point as the less advanced cities and faced many similar challenges and problems. However, they have already begun putting the discussed success factors into practice. Therefore, despite omitting the maturity level, the aggregated map abstracts the big picture well.

To ensure the validity and reliability of the work, the research team relied on the framework proposed by Runeson and Höst [90]. Construct validity describes the extent to which the subject of the study matches the researcher's understanding and RQs [90]. This means, for example, that the interviewees interpret the interview questions the same way as the researcher. To ensure this, the interviewees were informed in advance about the study's background, scope, and objectives. In addition, the interview guide was provided to the interviewees to give them the opportunity to ask questions about the guide in particular and the study in general to eliminate ambiguities or misunderstandings. Internal validity describes the credibility of the results [90].

To ensure the credibility of our work, the interview questions were based on a prior literature review [22], related work [5], [24], [27], [65], and cities strategies [13], [16], [66], [67]. This means that the results are rooted in a sound body of knowledge. In addition, compared with [27], a larger number of interviews were conducted, and more than only one country or one project [25], [26], [27] were evaluated. Finally, the data analysis was not conducted by a single researcher but always in close consultation with the research team to minimize researcher bias in the final results.

The external validity aspect describes the generalizability of the results [90]. As mentioned previously, interviews were conducted until theoretical saturation occurred. Furthermore, the focus was not on just one country or select cities based on their size. Nevertheless, governance forms and cultures of countries will vary, especially for non-EU cities, which is why our results represent an explanatory model for urban data governance and not definitive evidence. Reliability describes the possibility of reproducing the results [90]. To that end, the interviews were dictated by an interview guide based on related work and the body of knowledge. Thus, other research teams can use the interview guide to replicate the data collection.

The reproducibility of the work was also ensured through a detailed description of the research design. Finally, the bias of the individual researchers was reduced through collaborative analysis.

Future research should focus on developing a reference model for urban data governance. It can follow the research paradigm of design science and use the results of this study to develop an initial design for a reference model for urban data governance as part of the design cycle. In doing so, several strands contribute to the development of the reference model and to the knowledge base on urban data governance. For example, the different maturity levels of cities could be studied in more detail by creating a data governance maturity model for cities and benchmarking on this basis. Furthermore, research could explore whether the general digitization of a country (i.e., a high ranking in the DESI index) is reflected in a city's maturity level of data governance. Another possible research avenue is to investigate whether data governance frameworks from the corporate context are suitable for use in cities. In this way, the suitable approaches can be adopted in a urban data governance reference model. Finally, in the further development of a reference model, other urban actors should be included in the analysis, as smart cities ultimately do not consist only of the city administration; they are a complex system with multiple actors, such as civic organizations, research and educational institutions, mobility services, energy companies, and many other non-municipal actors that provide a large part of urban data.

## VI. CONCLUSION

Smart cities strive to improve quality of life, economic growth, and sustainability through the use of various technologies. Data management is the brain of a smart city, enabling data-driven decisions, planning, and applications. The findings show that cities face a siloed data landscape, leading to various challenges such as integration and interoperability issues. In addition, a general lack of data literacy and an old-fashioned mindset lead to a reluctance to become more data-driven. Data governance helps create a unified framework for data management by establishing policies, standards, and structures for all departments in line with a city's overall strategy to drive effective and efficient data management. In addition, data do not stop at city boundaries. Therefore, urban data governance promotes collaboration with city stakeholders and between cities to maximize value across city boundaries. A reference model for urban data governance can help cities implement a concrete data governance program. This article lays the groundwork for developing a reference model for urban data governance by presenting a comprehensive survey of 27 EU cities to identify the current urban data landscape, challenges cities face in data governance and management, and success factors. The findings reveal that many cities have yet to make progress in data governance, even though data are the backbone for the future of a city and the implementation of smart cities. Unfortunately, cities do not know where to start or what actions to take. By drawing on the current findings, future work could adopt the design cycle of the DSR (Fig. 1) to develop an urban data governance reference model. City leaders could also use the findings to shape their strategy and better understand the complexities of their data ecosystem.

## APPENDIX
## INTERVIEW GUIDE

a) Introduction

1) What is your current position, how long have you been in that position, and which responsibilities does it include?

b) The digitization of the city and the role of data usage

2) Does the city have a digitalization strategy, and what role does data usage play?

3) How do you define urban data?

c) Current status in handling data

4) How is your data infrastructure/landscape structured?

5) Is there data integration/exchange within and outside the city?

6) What governance structures and processes do you currently have to manage data?

7) How are decisions on data-related issues made and in which organizational unit?

8) Are the current data processes developed in-house or based on data governance frameworks?

d) Challenges, influencing factors, and outlook

9) What technological challenges do you experience in operating and developing your data infrastructure?

10) What are the organizational and procedural challenges you experience in managing data?

11) What factors do you think are necessary for successful data governance?

## ACKNOWLEDGMENT

The authors would like to thank the interview partners and the people who put them in touch with the respective interviewees. Despite their busy schedules, the interviewees took the time to support this work by sharing valuable firsthand insights, for which they are very grateful. Most notably, they made clear in the interviews that cities need more support from the academic world to become more data-driven.

## REFERENCES

[1] H. Wache and B. Dinter, "The digital twin—Birth of an integrated system in the digital age," in *Proc. Annu. 53rd Hawaii Int. Conf. Syst. Sci.*, 2020, pp. 5452–5461, doi: 10.24251/HICSS.2020.671.

[2] D. Puiu, P. Barnaghi, R. Tönjes, D. Kümper, M. I. Ali, A. Mileo, J. X. Parreira, M. Fischer, S. Kolozali, N. Farajidavar, F. Gao, T. Iggena, T.-L. Pham, C.-S. Nechifor, D. Puschmann, and J. Fernandes, "CityPulse: Large scale data analytics framework for smart cities," *IEEE Access*, vol. 4, pp. 1086–1108, 2016, doi: 10.1109/ACCESS.2016.2541999.

[3] S. Barns, "Smart cities and urban data platforms: Designing interfaces for smart governance," *City, Culture Soc.*, vol. 12, pp. 5–12, Mar. 2018, doi: 10.1016/J.CCS.2017.09.006.

[4] V. Moustaka, A. Vakali, and L. G. Anthopoulos, "A systematic review for smart city data analytics," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–41, Dec. 2018, doi: 10.1145/3239566.

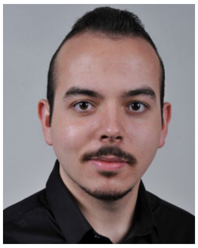
[5] S. Choenni, M. S. Bargh, T. Busker, and N. Netten, "Data governance in smart cities: Challenges and solution directions," *J. Smart Cities Soc.*, vol. 1, no. 1, pp. 31–51, Feb. 2022, doi: 10.3233/scs-210119.

[6] V. Khatri and C. V. Brown, "Designing data governance," *Commun. ACM*, vol. 53, no. 1, pp. 148–152, Jan. 2010.

[7] R. Mahanti, *Data Governance and Compliance*. Singapore: Springer, 2021, doi: 10.1007/978-981-33-6877-4.

[8] R. Seiner, *Non-Invasive Data Governance: The Path of Least Resistance and Greatest Success*, 1st ed. Sedona, AZ, USA: Technics Publications, 2014.
[9] K. Weber and C. Klingenberg, *Data Governance—Der Leitfaden für Die Praxis*. München, Germany: Hanser, 2021.
[10] Hamburg Department of IT and Digitalization. (2020). *Digital Strategy for Hamburg*. Accessed: Apr. 27, 2022. [Online]. Available: https://www.hamburg.de/contentblob/14924946/e80007b350f1abdc455cf aea7e8cd76c/data/download-digitalstrategie-englisch.pdf?userVariant=
[11] Digital Glasgow Board. (2018). *Digital Glasgow Strategy*. Accessed: Apr. 27, 2022. [Online]. Available: https://www.glasgow.gov.uk/ CHttpHandler.ashx?id=43572&p=0
[12] Digitales Und IT (DIGIT) Stadt Freiburg im Breisgau. (2019). *Freiburg. Digital. Gestalten*. Accessed: Apr. 27, 2022. [Online]. Available: https://digital.freiburg.de/digitalstrategie
[13] E. und B. Senatsverwaltung für Wirtschaft. (2018). *Digitalstrategie Berlin—Grünbuch für die Digitalisierungsstrategie Des Landes Berlin*. Accessed: Apr. 27, 2022. [Online]. Available: https://digitalstrategie.berlin.de/documents/17/Digitalisierungsstrategie_ Berlin_-_Grunbuch_-_update030521.pdf
[14] S. Böblingen. (2021). *Digitalstrategie des Landkreises Böblingen*. Accessed: Apr. 27, 2022. [Online]. Available: https://www.boeblingen. de/site/Boeblingen-Responsiv/get/params_E-2036796128/19101682/ Digitale-Strategie.pdf
[15] Stadtrat Stadt Zürich. (2018). *Strategie Smart City Zürich*. Accessed: Apr. 27, 2022. [Online]. Available: https://www.stadt-zuerich.ch/portal/de/index/politik_u_recht/stadtrat/weitere-politikfelder/ smartcity/strategie/publikationstrategie.html
[16] B. Lutz and Magistratsddirektion der Stadt Wien. (2019). *Data Excellence Wien*. Accessed: Apr. 27, 2022. [Online]. Available: https://digitales. wien.gv.at/wp-content/uploads/sites/47/2019/03/Data-Excellence.pdf
[17] L. Romualdo-Suzuki, "Data for London: A city data strategy," London, U.K.: Gerater London Authority, Tech. Rep., 2016, doi: 10.13140/RG.2.2.17632.79362.
[18] Barcelona City Council. (2022). *City Data Commons | Barcelona Digital City*. Accessed: Apr. 27, 2022. [Online]. Available: https://ajuntament. barcelona.cat/digital/en/digital-transformation/city-data-commons
[19] The Federal German Government. (2021). *Data Strategy of the Federal German Government*. Berlin. Accessed: Jun. 20, 2023. [Online]. Available: https://www.bundesregierung.de/breg-de/suche/data-strategy-of-the-federal-german-government-1950612
[20] S. Döring. (2021). *Die Datenstrategie der Bundesregierung—Erste Eindrücke*. Accessed: Mar. 24, 2022. [Online]. Available: https://muenchen. digital/blog/diedatenstrategie-der-bundesregierung-erste-eindruecke/
[21] K. Schlüter, *Die Stadt der Zukunft Mit Daten Gestalten: Souveräne Städte—Nachhaltige Investitionen in Dateninfrastrukturen*. Berlin, Germany: Deutscher Städtetag, 2021.
[22] Y. Bozkurt, A. Rossmann, and Z. Pervez, "A literature review of data governance and its applicability to smart cities," in *Proc. Annu. 55th Hawaii Int. Conf. Syst. Sci.*, 2022, pp. 2680–2689.
[23] P. D. König, "Citizen-centered data governance in the smart city: From ethics to accountability," *Sustain. Cities Soc.*, vol. 75, Dec. 2021, Art. no. 103308, doi: 10.1016/j.scs.2021.103308.
[24] L. Lupi, "City data plan: The conceptualisation of a policy instrument for data governance in smart cities," *Urban Sci.*, vol. 3, no. 3, p. 91, Aug. 2019, doi: 10.3390/urbansci3030091.
[25] N. Thompson, R. Ravindran, and S. Nicosia, "Government data does not mean data governance: Lessons learned from a public sector application audit," *Government Inf. Quart.*, vol. 32, no. 3, pp. 316–322, Jul. 2015, doi: 10.1016/j.giq.2015.05.001.
[26] S. Cuno, L. Bruns, N. Tcholtchev, P. Lämmel, and I. Schieferdecker, "Data governance and sovereignty in urban data spaces based on standardized ICT reference architectures," *Data*, vol. 4, no. 1, p. 16, Jan. 2019, doi: 10.3390/data4010016.
[27] K. Paskaleva, J. Evans, C. Martin, T. Linjordet, D. Yang, and A. Karvonen, "Data governance in the sustainable smart city," *Informatics*, vol. 4, no. 4, p. 41, Nov. 2017, doi: 10.3390/informatics4040041.
[28] J. Franke and P. Gailhofer, "Data governance and regulation for sustainable smart cities," *Frontiers Sustain. Cities*, vol. 3, pp. 1–12, Dec. 2021, doi: 10.3389/frsc.2021.763788.
[29] A. Hevner, "A three cycle view of design science research," *Scandin. J. Inf. Syst.*, vol. 19, no. 2, pp. 1–12, 2007.
[30] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quart.*, vol. 28, no. 1, pp. 75–105, Mar. 2004.
[31] V. Albino, U. Berardi, and R. M. Dangelico, "Smart cities: Definitions, dimensions, performance, and initiatives," *J. Urban Technol.*, vol. 22, no. 1, pp. 3–21, Jan. 2015.
[32] M. Batty, K. W. Axhausen, F. Giannotti, A. Pozdnoukhov, A. Bazzani, M. Wachowicz, G. Ouzounis, and Y. Portugali, "Smart cities of the future," *Eur. Phys. J. Special Topics*, vol. 214, no. 1, pp. 481–518, Dec. 2012.
[33] A. Caragliu, C. D. Bo, and P. Nijkamp, "Smart cities in Europe," *J. Urban Technol.*, vol. 18, no. 2, pp. 65–82, 2011.
[34] H. Chourabi, T. Nam, S. Walker, J. R. Gil-Garcia, S. Mellouli, K. Nahon, T. A. Pardo, and H. J. Scholl, "Understanding smart cities: An integrative framework," in *Proc. 45th Hawaii Int. Conf. Syst. Sci.*, Jan. 2012, pp. 2289–2297.
[35] B. N. Silva, M. Khan, and K. Han, "Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities," *Sustain. Cities Soc.*, vol. 38, pp. 697–713, Apr. 2018, doi: 10.1016/j.scs.2018.01.053.
[36] Y. Sun, H. Song, A. J. Jara, and R. Bie, "Internet of Things and big data analytics for smart and connected communities," *IEEE Access*, vol. 4, pp. 766–773, 2016.
[37] M. A. Kuddus, E. Tynan, and E. McBryde, "Urbanization: A problem for the rich and the poor?" *Public Health Rev.*, vol. 41, no. 1, pp. 1–4, Jan. 2020, doi: 10.1186/S40985-019-0116-0.
[38] B. Güneralp, R. I. McDonald, M. Fragkias, J. Goodness, P. J. Marcotullio, and K. C. Seto, "Urbanization forecasts, effects on land use, biodiversity, and ecosystem services," in *Urbanization, Biodiversity and Ecosystem Services: Challenges and Opportunities*. Dordrecht, The Netherlands: Springer, Jan. 2013, pp. 437–452, doi: 10.1007/978-94-007-7088-1_22.
[39] M. Shahidehpour, Z. Li, and M. Ganji, "Smart cities for a sustainable urbanization: Illuminating the need for establishing smart urban infrastructures," *IEEE Electrific. Mag.*, vol. 6, no. 2, pp. 16–33, Jun. 2018, doi: 10.1109/MELE.2018.2816840.
[40] D. G. Costa, J. P. J. Peixoto, T. C. Jesus, P. Portugal, F. Vasques, E. Rangel, and M. Peixoto, "A survey of emergencies management systems in smart cities," *IEEE Access*, vol. 10, pp. 61843–61872, 2022, doi: 10.1109/ACCESS.2022.3180033.
[41] V. Fernandez-Anez, "Stakeholders approach to smart cities: A survey on smart city definitions," in *Proc. Int. Conf. Smart Cities*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics. Cham, Switzerland: Springer, 2016, pp. 157–167, doi: 10.1007/978-3-319-39595-1_16.
[42] M. Marrone and M. Hammerle, "Smart cities: A review and analysis of stakeholders' literature," *Bus. Inf. Syst. Eng.*, vol. 60, no. 3, pp. 197–213, Jun. 2018, doi: 10.1007/s12599-018-0535-3.
[43] A. Meijer and M. P. R. Bolívar, "Governing the smart city: A review of the literature on smart urban governance," *Int. Rev. Administ. Sci.*, vol. 82, no. 2, pp. 392–408, Jun. 2016, doi: 10.1177/ 0020852314564308.
[44] A. Kirimtat, O. Krejcar, A. Kertesz, and M. F. Tasgetiren, "Future trends and current state of smart city concepts: A survey," *IEEE Access*, vol. 8, pp. 86448–86467, 2020, doi: 10.1109/ACCESS.2020. 2992441.
[45] Y. Bozkurt, R. Braun, A. Rossmann, and D. Hertweck, "Smart cities in research: Status-quo and future research directions," *IADIS Int. J. WWW/Internet*, vol. 18, no. 1, pp. 121–138, 2020.
[46] D. Ahlers, L. W. M. Wienhofen, S. A. Petersen, and M. Anvaari, "A smart city ecosystem enabling open innovation," *Commun. Comput. Inf. Sci.*, vol. 1041, pp. 109–122, Jan. 2019, doi: 10.1007/978-3-030-22482-0_9.
[47] M. Stone, J. Knapper, G. Evans, and E. Aravopoulou, "Information management in the smart city," *Bottom Line*, vol. 31, nos. 3–4, pp. 234–249, Nov. 2018, doi: 10.1108/BL-07-2018-0033.
[48] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014, doi: 10.1109/JIOT.2014.2306328.
[49] L. Broccardo, F. Culasso, and S. G. Mauro, "Smart city governance: Exploring the institutional work of multiple actors towards collaboration," *Int. J. Public Sector Manage.*, vol. 32, no. 4, pp. 367–387, May 2019, doi: 10.1108/IJPSM-05-2018-0126.

[50] N. S. Jayasena, H. Mallawaarachchi, and K. G. A. S. Waidyasekara, "Stakeholder analysis for smart city development project: An extensive literature review," in *Proc. MATEC Web Conf.*, vol. 266, Feb. 2019, p. 06012, doi: 10.1051/matecconf/201926606012.
[51] M. Bevir, *A Theory of Governance*. Berkeley, CA, USA: Univ. California Press, 2013.
[52] V. Chhotray and G. Stoker, *Governance: From Theory to Practice*. London, U.K.: Palgrave Macmillan, 2009, pp. 214–247.
[53] J. Kooiman, *Governing as Governance*. Newbury Park, CA, USA: SAGE, 2003.
[54] R. Mahanti, *Data Governance and Data Management Contextualizing Data Governance Drivers, Technologies, and Tools Data Governance*. Cham, Switzerland: Springer, 2021.
[55] K. Weber, B. Otto, and H. Österle, "One size does not fit all—A contingency approach to data governance," *J. Data Inf. Quality*, vol. 1, no. 1, pp. 1–27, Jun. 2009.
[56] *DAMA-DMBOK: Data Management Body of Knowledge*, Dama International, Technics Publications, Sedona, AZ, USA, 2017.
[57] S. Soares, *The IBM Data Governance Unified Process: Driving Business Value with IBM Software and Best Practices*. Boise, ID, USA: MC Press, 2010.
[58] The Data Governance Institute. (2023). *Data Governance Framework & Components—The Data Governance Institute*. Accessed: Feb. 8, 2023. [Online]. Available: https://datagovernance.com/data-governance-framework-components/
[59] L. Kuan Cheong, V. Chang, and L. Kuan, "The need for data governance: A case study," in *Proc. 18th Australas. Conf. Inf. Syst.*, 2007, pp. 999–1008.
[60] A. Castro, V. A. Villagrá, P. García, D. Rivera, and D. Toledo, "An ontological-based model to data governance for big data," *IEEE Access*, vol. 9, pp. 109943–109959, 2021, doi: 10.1109/ACCESS.2021.3101938.
[61] DIN SPEC 91357. (Dec. 2017). *Reference Architecture Model Open Urban Platform (OUP)*. Accessed: Apr. 2, 2023. [Online]. Available: https://www.beuth.de/en/technical-rule/din-spec-91357/281077528
[62] A. Fink, *The Survey Handbook*. Newbury Park, CA, USA: SAGE, 2003.
[63] D. Siemon. (2022). *Methods in Design Science Research*. Accessed: Apr. 27, 2022. [Online]. Available: https://design-science-research.de/en/post/methods-in-dsr/
[64] R. Edwards and J. Holland, *What is Qualitative Interviewing?* New York, NY, USA: Academic Press, 2013.
[65] D. Eke and O. E. John, "The role of data governance in the development of inclusive smart cities," in *Proc. Int. Conf. Ethical Social Impact ICT*, 2020, pp. 603–619.
[66] S. Köln and D. Blauhut, "Daten 'machen' strategie," City Cologne, Cologne, Germany, Tech. Rep. I-D2, 2021. [Online]. Available: https://buergerinfo.stadt-koeln.de/getfile.asp?id=821964&type=do
[67] S. Soest, "Datenstrategie soest," City Soest, Soest, Germany, Tech. Rep. vers.1.0, 2021. [Online]. Available: https://digital-soest.de/images/AIDW/Datenstrategie_Final_2021-05-14.pdf
[68] J. Saldaña, *The Coding Manual for Qualitative Researchers*, 2nd ed. Newbury Park, CA, USA: SAGE, 2013.
[69] G. Guest, K. MacQueen, and E. Namey, *Applied Thematic Analysis*. Newbury Park, CA, USA: SAGE, 2012, doi: 10.4135/9781483384436.
[70] K. M. MacQueen and G. Guest, "An introduction to team-based qualitative research," in *Handbook for Team-Based Qualitative Research*, K. M. MacQueen and G. Guest, Eds. Lanham, MD, USA: AltaMira Press, 2008, pp. 3–19.
[71] E. Namey, G. Guest, L. Thairu, and L. Johnson, "Data reduction techniques for large qualitative data sets," in *Handbook for Team-Based Qualitative Research*. Lanham, MD, USA: AltaMira Press, 2008, pp. 137–162.
[72] F. Ackermann, C. Eden, and S. Cropper, "Getting started with cognitive mapping," in *Proc. 7th Young OR Conf.*, Apr. 1992, pp. 65–82.
[73] C. Eden, "Cognitive mapping," *Eur. J. Oper. Res.*, vol. 10, pp. 1–13, Jan. 1988, doi: 10.1016/0377-2217(88)90002-1.
[74] C. Browne, "The application of cognitive map analysis to semi-structured depth interviews with adults who have received formal adult literacy tuition," *Stud. Educ. Adults*, vol. 21, no. 2, pp. 140–149, Oct. 1989, doi: 10.1080/02660830.1989.11730527.
[75] G. A. Kelly, *The Psychology of Personal Constructs*. New York, NY, USA: Norton, 1955.
[76] H. S. Firmansyah, S. H. Supangkat, A. A. Arman, and P. J. Giabbanelli, "Identifying the components and interrelationships of smart cities in Indonesia: Supporting policymaking via fuzzy cognitive systems," *IEEE Access*, vol. 7, pp. 46136–46151, 2019, doi: 10.1109/ACCESS.2019.2908622.
[77] European Commission. (2022). *Digital Economy and Society Index (DESI) 2022*. Accessed: Jan. 20, 2023. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/desi
[78] B. Glaser and A. Strauss, *Grounded Theory? Strategien Qualitativer Forschung*, vol. 3. Bern, Switzerland: Hans Huber, 2010.
[79] MAXQDA. (2022). *All-In-One Tool for Qualitative Data Analysis & Mixed Methods—MAXQDA*. Accessed: Feb. 10, 2023. [Online]. Available: https://www.maxqda.com/
[80] BANXIA Software. (2023). *BANXIA Decision Explorer*. Accessed: Feb. 10, 2023. [Online]. Available: https://banxia.com/dexplore/
[81] E. A. Lavin and P. J. Giabbanelli, "Analyzing and simplifying model uncertainty in fuzzy cognitive maps," in *Proc. Winter IEEE Simulation Conf. (WSC)*, Dec. 2017, pp. 1868–1879, doi: 10.1109/WSC.2017.8247923.
[82] T. Bouzdine-Chameeva, F. Durrieu, and T. Mandják, "Cognitive mapping methodology for understanding of business relationship value competitive paper," in *Proc. 17th IMP Conf.*, Oslo, Norway, 2001, pp. 1–20.
[83] U. Ozen and F. Ulengin. (2001). *Analyzing Strategic Thoughts of Corporations Based on Cognitive Map*. Accessed: May 31, 2023. [Online]. Available: https://www.infoman.com.tr/dosyalar/dosya/Yayinlar_-_Analyzing_Strategic_Thoughts_by_Cognitive_Mapping.pdf
[84] *Decision Explorer User's Guide*. Banxia Software Ltd., Kendal, U.K., 2017.
[85] R. Jain, "Out-of-the-box data engineering events in heterogeneous data environments," in *Proc. 19th Int. Conf. Data Eng.*, 2003, pp. 8–21, doi: 10.1109/ICDE.2003.1260778.
[86] J. Kim, H. Ha, B. G. Chun, S. Yoon, and S. K. Cha, "Collaborative analytics for data silos," in *Proc. 32nd Int. Conf. Data Eng. (ICDE)*, May 2016, pp. 743–754, doi: 10.1109/ICDE.2016.7498286.
[87] S. Jeong, S. Kim, and J. Kim, "City Data Hub: Implementation of standard-based smart city data platform for interoperability," *Sensors*, vol. 20, no. 23, p. 7000, Dec. 2020, doi: 10.3390/s20237000.
[88] C. Labadie, C. Legner, M. Eurich, and M. Fadler, "FAIR enough? Enhancing the usage of enterprise data with data catalogs," in *Proc. IEEE 22nd Conf. Bus. Informat. (CBI)*, Jun. 2020, pp. 201–210, doi: 10.1109/CBI49978.2020.00029.
[89] K. Czajkowski, C. Kesselman, and R. Schuler, "ERMRest: A collaborative data catalog with fine grain access control," in *Proc. IEEE 13th Int. Conf. E-Sci.*, Oct. 2017, pp. 510–517, doi: 10.1109/ESCIENCE.2017.83.
[90] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," *Empirical Softw. Eng.*, vol. 14, no. 2, pp. 131–164, Apr. 2009, doi: 10.1007/s10664-008-9102-8.



**YUSUF BOZKURT** received the B.Sc. and M.Sc. degrees in information systems from Reutlingen University, Germany, in 2017 and 2019, respectively. He is currently pursuing the Ph.D. degree in data governance and data management in smart cities with the University of the West of Scotland. He did the B.Sc. thesis on "In-Memory Databases and Their Impact on Data Warehousing" and in the master's thesis he analyzed how the business intelligence landscape of a utility distribution network operator in Germany, could be redesigned for future requirements, such as real-time analytics. Since 2019, he has been a Research Fellow with the Herman Hollerith Centre, Information Systems Research Institute, Research Laboratory for Digital Business, Reutlingen University, where he involved in smart city projects with local stakeholders and provides scientific support to cities.

**ALEXANDER ROSSMANN** received the master's degree from the University of Tübingen and The State University of New York and the Ph.D. degree from the University of St. Gallen. He is currently a Professor of digital business with Reutlingen University and the Project Director of the University of St. Gallen. He is also leading the Research Center for Digital Business, Reutlingen University, and the Herman Hollerith Center for Computer Science in Böblingen. Prior to this, he was ten years of experience as the Managing Director of a leading consultancy firm. He has a track record of more than 80 peer-reviewed publications in relevant conferences proceedings and scientific journals. He has more than 20 years of research and industry experience in the area of digital transformation and the application of digital technologies in corporations. His research interests include applied machine learning, digital business, smart cities, and organizational agility. He served as a programme committee member for more than 50 conferences and a reviewer for more than 100 journals submissions.

**ASHWINI KONANAHALLI** received the Ph.D. degree from Queen's University Belfast, Belfast, U.K. She is currently the Program Leader of the Civil Engineering Graduate Apprenticeship Undergraduate Course with the University of West of Scotland (UWS), U.K. Prior to joining UWS, she was a Lecturer with Queen's University Belfast. In the current role, she teaches extensively and leads various undergraduate and post graduate courses. She was previously a practicing Architect and involved in several architectural projects in India and Northern Ireland. She has experience of securing knowledge transfer partnerships (KTP) in the digital built environment. This research track record has been achieved through sustained engagement and collaboration with globally leading contributors from both industry and universities.

**ZEESHAN PERVEZ** (Senior Member, IEEE) is currently the Director of the Data Analytics for Intelligent and Autonomous Systems (DeLTA) and a Professor of computer science with the University of the West of Scotland (UWS). He is an ACM Distinguished Speaker, a Senior Fellow of the Higher Education Academy, and a Full Member of EPSRC Peer-Review College. He has a strong track record of securing research, industry, and capacity-building funding from the Innovate U.K., Scottish Funding Council, Innovation Centers, European Commission, Erasmus+, Microsoft Research, and international research institutes. He has more than 18 years of research and industry experience in addressing technological and societal challenges and designing and developing enterprise-ready solutions. He has outstanding performance and success in securing substantial research funding, enterprise engagement, high-quality research outputs, and associated wider research/societal impact. He has successfully supervised to completion: 29 postdoctoral fellows, postgraduate research/taught, and undergraduate students (as a lead and a co-supervisor). His areas of expertise are data science, applied ML/AI, cybersecurity, Industry 4.0, and edge/fog computing. His research interests include Industry 4.0, smart cities, social housing, predictive maintenance, facilities management, healthcare, data science, and ICT4D. He has served as a TPC member for more than 100 conferences and a regular reviewer for more than 30 IEEE Transactions, Elsevier, and Springer journals. He was a recipient of the 2019 UWS STARS Award—Staff Appreciation and Recognition Scheme. He has chaired/evaluated more than 28 doctoral exams for the U.K. and international higher education institutes. He has been regularly invited by national and international research and academic institutes to deliver talks on topics ranging from the Internet of Things to data science, cybersecurity, and cloud computing.