Government Information Quarterly 42 (2025) 102063

Contents lists available at ScienceDirect

Government Information Quarterly

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

Data collaboration in digital government research: A literature review and research agenda

Juliane Schmeling[a,*], Sami al Dakruni[a], Ines Mergel[a,b,c]

[a] Fraunhofer Institute for Open Communication Systems, Department Digital Public Services, Kaiserin-Augusta-Allee 31, 10589 Berlin, Germany
[b] University of Konstanz, Department of Politics and Public Administration, Universitätsstr. 10, 78464 Konstanz, Germany
[c] University of Vaasa, Vaasa, Finland

ARTICLE INFO

Keywords:
Data collaboration
Data ecosystem
Data space
Computational literature review
Structural topic modelling
Digital government

ABSTRACT

Sovereign data infrastructures are a central building block of the European Data Strategy, yet little is known about how public administrations share and collaborate on both open and restricted data. This research addresses the gap by systematically analysing the existing literature on data collaboration within the field of digital government research. We thereby make a methodological contribution to digital government research through a rigorous literature review framework that includes Structural Topic Modelling to understand the different themes of the scientific discussion in the field of digital government. We propose an innovative data collaboration framework that includes the ecosystem, the organisational, and the individual levels, enhancing our understanding of the multidimensional nature of data collaboration. Our analysis reveals that while the emphasis is on innovation and participation, critical aspects like standardisation and data management have a declining topic prevalence, despite their importance in developing federated data ecosystems. This comprehensive analysis not only sheds light on the current landscape but also informs a structured research agenda in digital government, aiming to contribute to the advancement of the field.

1. Introduction

The European data strategy is a pioneering move that significantly propels sovereign data collaboration across federated and trustworthy cloud infrastructures, as outlined by the European Commission (2020). These efforts are part of a paradigm shift towards a democratised and decentralised European data economy. The strategy also seeks to enhance a European Digital Single Market based on European values for fair data exchange (ibid.). The enactment of the General Data Protection Regulation (GDPR) has been pivotal, catalysing discussions on data sovereignty and establishing a foundational framework for an alternative data strategy that diverges from approaches adopted by China and the USA (Calzada & Almirall, 2020). Data sovereignty, as a multifaceted concept, empowers individuals and organisations to maintain control over their data, extending its influence to matters of national protectionism, digital asset management, and individual privacy (Budnitsky, 2022; Couture & Toupin, 2019; Hummel et al., 2021). Recent initiatives by various European institutions, such as the Gaia-X initiative and the European Commission's efforts to establish Common European Data Spaces, underscore a committed move towards developing sovereign and federated data infrastructures (AISBL, 2024; European Commission, 2022). These developments are critical as they shape how stakeholders collaborate on data, highlighting the urgent need for rigorous research into data collaboration within this transformative European context. Recognising the emergent nature of sovereign European data spaces, researchers in digital government must adapt and innovate governance and organisational models for data collaboration. While these initiatives provide a concrete framework and motivation for improved data collaboration, there remains a notable gap in understanding how such collaboration is practically implemented, particularly when it comes to balancing open and restricted data sharing under diverse governance and regulatory frameworks. This gap is even more pronounced given the lack of systematic literature that explores the interplay between technical infrastructures, governance modes, and organisational processes that underpin such collaborations. This study aims to comprehensively examine how current research addresses data collaboration and management, to identify gaps and blind spots that will shape the future research agenda. Through a computational analysis of the existing literature and employing foundational data management frameworks like data collaboratives (Ruijer, 2021; Susha et al., 2017; Verhulst &

* Corresponding author.
E-mail address: Juliane.schmeling@fokus.fraunhofer.de (J. Schmeling).

https://doi.org/10.1016/j.giq.2025.102063
Received 1 July 2024; Received in revised form 4 July 2025; Accepted 7 August 2025
Available online 14 August 2025

Sangokoya, 2015) and data ecosystems (Gelhaar et al., 2017; Susha et al., 2022), this paper offers insights and direction for subsequent research.

This effort is not only crucial for advancing our understanding but also for developing relevant theoretical and practical frameworks that can support the evolving needs of digital governance. Therefore, we set out to explore the following research questions:

1.) How has data collaboration been discussed in digital government research?
2.) What unresolved research problems should future studies focus on?

We used the novel Structural Topic Modelling (STM) technique (Roberts et al., 2016) to transparently analyse data-related topics in peer-reviewed literature. The STM method has not yet received much attention in digital government research, representing a clear methodological gap. A Web of Science query for "Structural Topic Modeling" OR "Structural Topic Modelling" in June 2023 revealed that only 5 out of 200 published records used STM in public administration research, and ten papers were found in the field of political science.

In the following, we will first provide an overview of the current data management concepts and derive a conceptual framework to interpret the results of the analysis. We then present our research design following a mixed-methods approach. Finally, we present the results of the study, discuss the results, and conclude with suggestions for a future research agenda. This comprehensive approach ensures a thorough exploration of the data collaboration landscape, aiming to contribute significantly to the field's development.

2. Data management core concepts as a basis for data collaboration

Data management forms the basis for data collaboration and sharing. Data collaboration is an underdeveloped concept, often outlined through organising principles such as data collaboratives or data ecosystems. Existing data management concepts vary, encompassing both inter-organisational and intra-organisational collaborations.

Data ecosystems can be seen as an umbrella term for any construction to exchange data inter-organisationally. Oliveira et al. (2019) define the data ecosystem as "…a loose set of interacting actors that directly or indirectly consume, produce, or provide data and other related resources (e.g., software, services, and infrastructure). Each actor performs one or more roles and is connected to other actors through relationships, in such a way that actors promote data ecosystems by collaborating and competing with each other". According to the taxonomy for data ecosystems developed by Gelhaar et al. (2017), data ecosystems are classifiable into three meta-dimensions: economic, technical and governance. The economic dimension contains the domain (scientific, government, business, industry), the purpose (innovation, interaction, transaction) and the organisational structure of the data ecosystem (platform-centric, decentralised, marketplace-based, keystone-centric = one major data provider). The technical dimension addresses centralised and decentralised infrastructures, and their degree of openness. The governance dimension outlines aspects of control and interdependencies. While the governance dimension already considers aspects of control and interdependencies, Wareham et al. (2013) emphasise the explicit and dynamic interplay between control and generativity.

Data spaces are often synonymous with data ecosystems, but discuss data collaboration primarily within the framework of the technical meta-dimension. The data space concept emerged 20 years ago in computer science as a decentralised data integration concept without the requirement of physical data integration (Franklin et al., 2005). It is instead based on the semantic layer with shared vocabularies (Halevy et al., 2006). The current understanding of data spaces refers to the collaboration of data in a decentralised infrastructure. It allows trusted data chains to be shared between sovereign data space participants (Otto, 2022).

On the intra-organisational level, a novel concept called data mesh emerged. It is a decentralised organisation of the data in alignment with the internal business units (Goedegebuure et al., 2025). The concept suggests that the specialised units or departmental silos within an organisation are more adept at understanding the data relevant to their specific area of expertise than a centralised data science team. Accordingly, monolithic data architectures should be decentralised in line with data domains relating to the domain-driven design approach (Goedegebuure et al., 2025; Machado et al., 2022). Both the data space and data mesh approaches are based on a decentralised technical infrastructure. However, the current application of data spaces focuses on inter-organisational collaboration, while data mesh addresses intra-organisational technical infrastructure.

As a result, the concept of data collaboratives emerged in digital government research (Susha et al., 2017; Susha et al., 2018; Susha et al., 2022; Verhulst & Sangokoya, 2015). The authors address the collaboration among different stakeholders to share and reuse data across sectors to solve public problems. This endeavour goes beyond the traditional public-private partnership conception. They outline the importance of shared standards and frameworks to enable multi-stakeholder participation. Susha et al. (2022) stressed the critical role of conveners in establishing data ecosystems. They highlight different roles, including connecting, mediating, stimulating, learning catalyst, and infrastructure provision roles. Based on the demand-supply relationship, the taxonomy of data collaboratives for public good uses two dimensions to classify the various ways to organise data sharing: data and collaboration. Fig. 1 illustrates the taxonomy for data collaboratives (Susha et al., 2017). Data sharing on the supply side focuses on various data providers and types, content, and administrative levels of the data to be shared. The demand side pertains to the target user, instruments to attract usage and collaboration among users, and the purpose of use and expected outcomes. The degree of access and facilitation are key mechanisms for matching demand and supply, making them central to the entire process (ibid.).

Merging the taxonomies of data ecosystems (Gelhaar et al., 2017) and data collaboratives (Susha et al., 2017) yields the following conceptual framework for data collaboration, depicted in Fig. 2. The orange boxes refer to elements of the taxonomy for data collaboratives, and the grey boxes to the taxonomy for data ecosystems. Data collaboration is an umbrella term for managing data exchange on multiple levels. Therefore, we assign facilitation mechanisms for data collaboration to the three meta-dimensions of a data ecosystem: economic, technical and

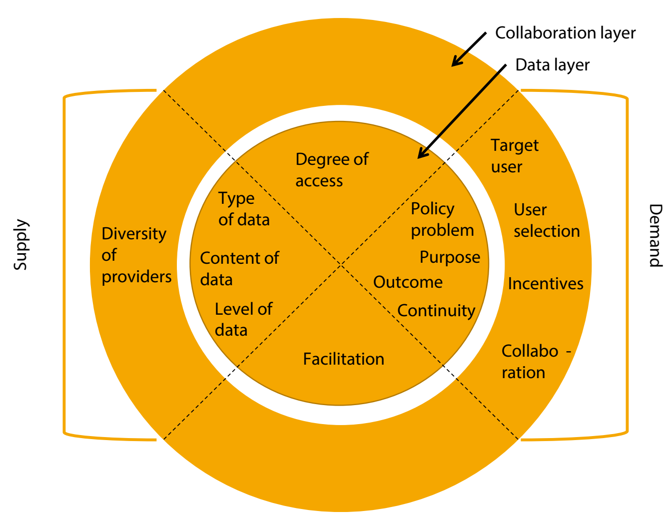

Fig. 1. Taxonomy for data collaboratives according to Susha et al., 2017

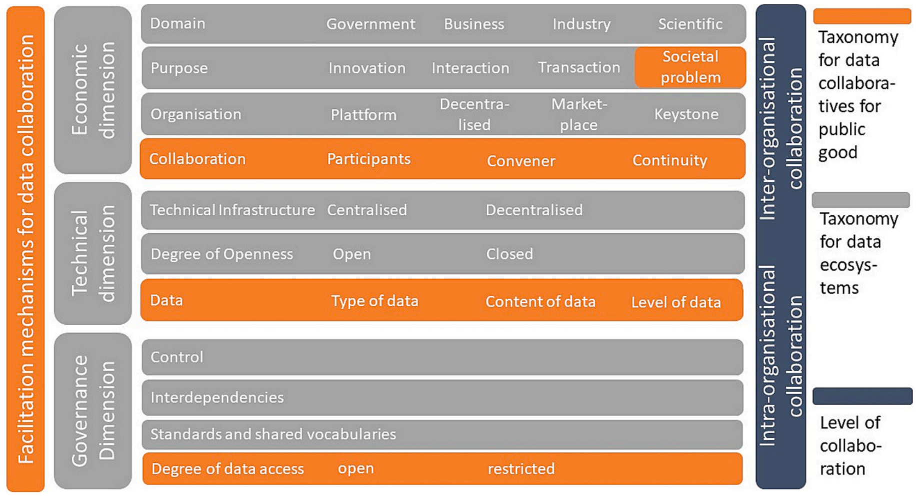

Fig. 2. Data collaboration framework.

governance. The taxonomy for data collaboratives introduces addressing societal problems as an additional purpose for data sharing and collaboration alongside the economic dimension. Further, a new participant-oriented layer for the collaboration among different types of participants, together with the role of a convener that constitutes and drives the ecosystem, comes along. On the technical dimension, the taxonomy for data collaboratives adds the data layer focusing on data types, content and level. On the governance level, the taxonomy for data collaboratives emphasises the degree of access. Each of the dimensions holds for inter-organisational and intra-organisational collaboration, as shown in Fig. 2 below:

This resulting data collaboration framework provides a robust analytical foundation for interpreting and understanding the findings of this study. In the following, we systematically examine how data collaboration was discussed in the digital government literature so far and derive open research questions for future studies.

3. Research design

The research design of this study combines both quantitative and qualitative methods to elaborate on the research questions: How data collaboration has been discussed in digital government research, and what unresolved research problems future studies should focus on? We applied a rather novel computational technique, the Structural Topic Modelling (Roberts et al., 2016) to gain a transparent understanding of the prevalence of data-related topics in the existing peer-reviewed literature.

3.1. STM technique

Automated text mining techniques, like STM, allow for the exploration of topics, themes, and frames from large text corpora, such as news articles, tweets, or other documents, and can support the development of data-driven theories in policy research. Topic models are a broad class of Bayesian generative models that automatically extract rich latent subjects from unstructured text data (Grimmer & Stewart, 2013). These models are sometimes called "unsupervised" techniques because they infer topics without predefined coding or topic schemes, unlike "supervised" approaches, where analysts manually select topics in advance. Compared to clustering algorithms, topic models are based on the assumption that the analysed documents are not only assigned to one category but have a proportional membership to each cluster or topic (Grimmer et al., 2022). STM is based on the Latent Dirichlet Allocation Algorithm (LDA) to uncover and analyse the topic structure of a given text corpus. A substantial benefit of STM over other topic modelling methods is the possibility of including document-level metadata as covariates (Maerz & Puschmann, 2020; Roberts et al., 2014). Although the method invented by Roberts et al. (2014) is novel particularly in political sciences and public administration, several applications in the broader social sciences space can be found in recent years. In the communication domain, STM was used to analyse social media communication, such as Twitter tweets like Zhang et al. (2023) did, to identify patterns in Chinese media communication, specifically in regards to the Hong Kong protests. Chen et al. (2023) applied the STM method to analyse bibliographic data as well as unstructured text data in the field of smart healthcare, in addition to Sharma et al. (2021), who studied the evolution of trend topics in information management research in the last 50 years using bibliographic data to combine bibliometric and topic modelling techniques. Both took advantage of STM, incorporating bibliographic metadata as covariates into the estimation process. Aranda et al. (2021) integrated STM into critical discourse analysis, combining quantitatively derived latent semantic structures with in-depth analysis of exemplary text samples, resulting in a transformative mixed-methods design. Guenduez and Mettler (2022) integrated STM into a qualitative narrative policy framework and enriched latent topics in policy documents with a systematic exploration of policy narratives.

We apply STM in the study to understand the prevalence of "data" in the scientific discussion among digital government researchers. STM helps us discover the relevant topics and their topic prevalence. The Gamma Coefficient (γ) expresses the prevalence. The text corpus consists of 1.441 data-related article abstracts and titles collected in the Digital Government Reference Library. The year of publication serves as the covariate in the analysis of topic prevalence over time.

3.2. Database selection

The Digital Government Reference Library (DGRL) in Version 18.5 served as our database (Scholl, 2022). The peer-reviewed research article collection represents a high-value dataset within the digital government research field. The field of Digital Government Research originates in the late 1990s at the intersection of various traditional academic disciplines. It has evolved into a multi-disciplinary field without a clear "owner", yet benefits from the diverse perspectives of information studies, public administration, and more (Scholl, 2021). The DGRL has played a significant role in shaping this domain by providing an exclusive bibliographic database and supporting bibliometric studies. It has thereby enhanced the field's development and interdisciplinary connections.

3.3. Research procedure

Our research approach consisted of four phases, as shown in Fig. 3. The procedure integrates the quantitative STM approach proposed by Roberts et al. (2019) with the qualitative interpretation framework by Schwartz-Shea and Yanow (2012) and progressed on the following four phases (see, Fig. 3):

3.3.1. Data selection

The DGRL library contains 19.152 articles including abstracts, titles, and bibliographic metadata (Scholl, 2022). We conducted several trials to determine the final text corpus, calculating the STM model with various input corpora to select the articles from the DGRL. This included the entire DGRL, a reduced selection with “cloud” as a keyword, and articles featuring “ecosystem” or “data” in the title. We manually evaluated the Gamma Coefficient, top words for each topic, and conducted diagnostic tests for meaningfulness. The final selection comprised 1.441 articles with ‘data’ in the title, which is too extensive for manual analysis using a supervised approach. Scholl (2022) provides the data in different machine-readable formats that can be loaded directly into R. We cleaned the text data by excluding records without abstracts and titles. This step reduced the number of records from 1.922 to 1.441. Next, we identified and selected variables to retain as document metadata, ultimately choosing three out of 20 variables: text, label, and year. One can easily incorporate additional variables if deemed necessary.

3.3.2. Preparation

To prepare the text data, we stemmed the words to reduce them to their respective root, excluding stop words, and finally transferred every word into one row, maintaining the covariates such as year and author for each term (Roberts et al., 2019). To exclude words with a high frequency, such as “research” or “government”, we applied an iterative process, and the count command was used in each iteration to determine these words. After the preparation, the data set contained 141,176 rows.

3.3.3. Estimation

During the estimation phase, the most critical parameter is quantification of the optimal number of topics (Roberts et al., 2019). This setting is usually made manually after various diagnostic tests. Roberts et al. (2014) propose leveraging the combination of semantic coherence and exclusivity to find the best-fitting number of topics (k). We examined a corpus sample with $k = 6, 8, 10$, and 12 topics. Semantic coherence ensures that a topic maintains internal consistency but does not account for similarity between topics. On the other hand, exclusivity assesses how unique a topic is by checking if words that are likely in one topic are unlikely in others (Buhin Pandur et al., 2020; Roberts et al., 2014; Roberts et al., 2019). Fig. 4 shows that the optimal number of topics ranges between 8 and 10, as these topics have the best tradeoff between semantic coherence and exclusivity.

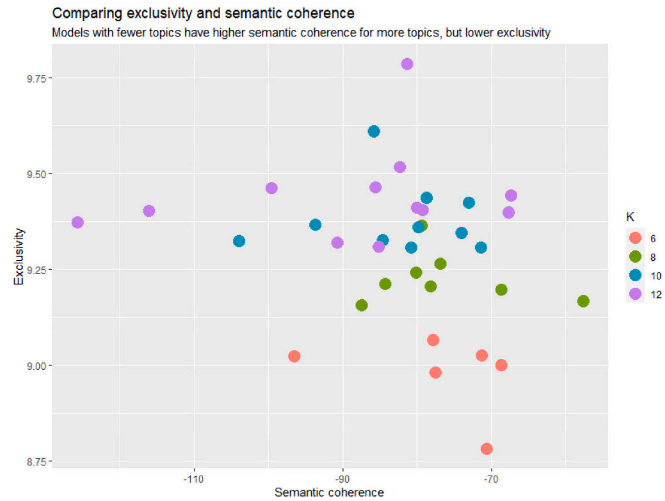

Fig. 4. Semantic coherence and exclusivity.

We performed further diagnostic tests, as represented in Fig. 5. Residuals are lowest for 12 topics, held out-likelihood is highest for ten topics, but semantic coherence is highest for eight. For this reason, the final calculation of topic prevalence was performed with eight topics.

The final result of the STM analysis provides an overview of the prevalence of data-related topics in the DGRL. Topic prevalence is the expected proportion of the text corpus that belongs to each topic (Roberts et al., 2019). The gamma coefficients represent the prevalence, with the highest coefficient occupying the top position. To examine the correlation between time and various topics, we quantified the effect of the year on other topics (Maerz & Puschmann, 2020; Roberts et al., 2014). This method enabled estimating the relationship between time and expected topic proportion between 2005 and 2022.

3.3.4. Interpretation

Lastly, in the interpretation phase, we applied a two-step procedure for labelling the topics as suggested by Chen et al. (2023). First, we conducted an interpretive and qualitative assessment of top words and frex words to complement the statistical analysis (see,

Appendix II). Top words are high-probability words that commonly occur in the respective topic (Roberts et al., 2014). Frex words are the most exclusive terms for each topic (Roberts et al., 2019).

Further, we combined the findings from our rich quantitative model with a qualitative interpretation based on an in-depth literature review

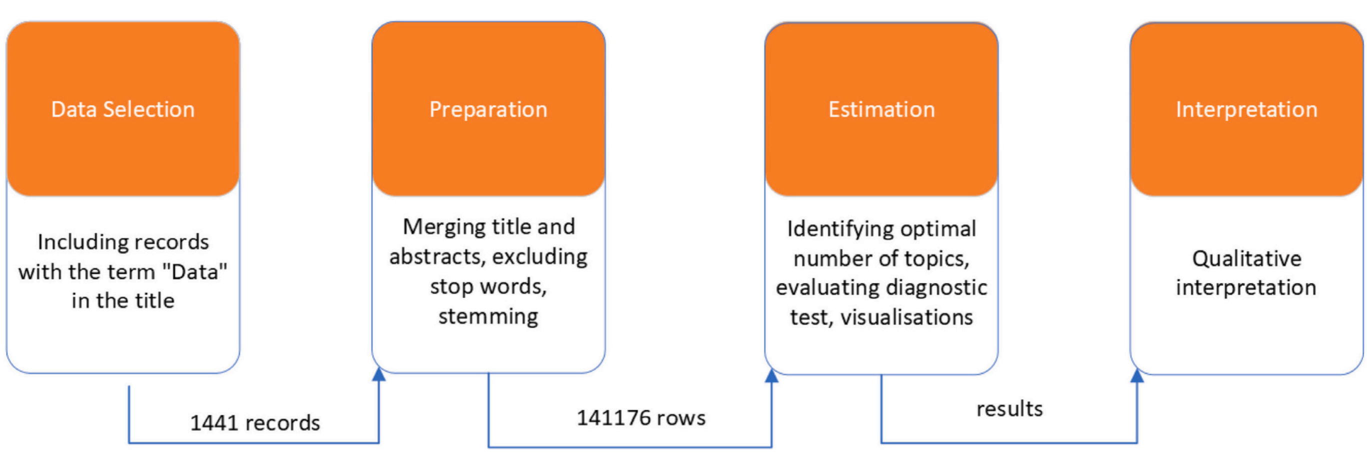

Fig. 3. STM procedure combined with qualitative interpretation.

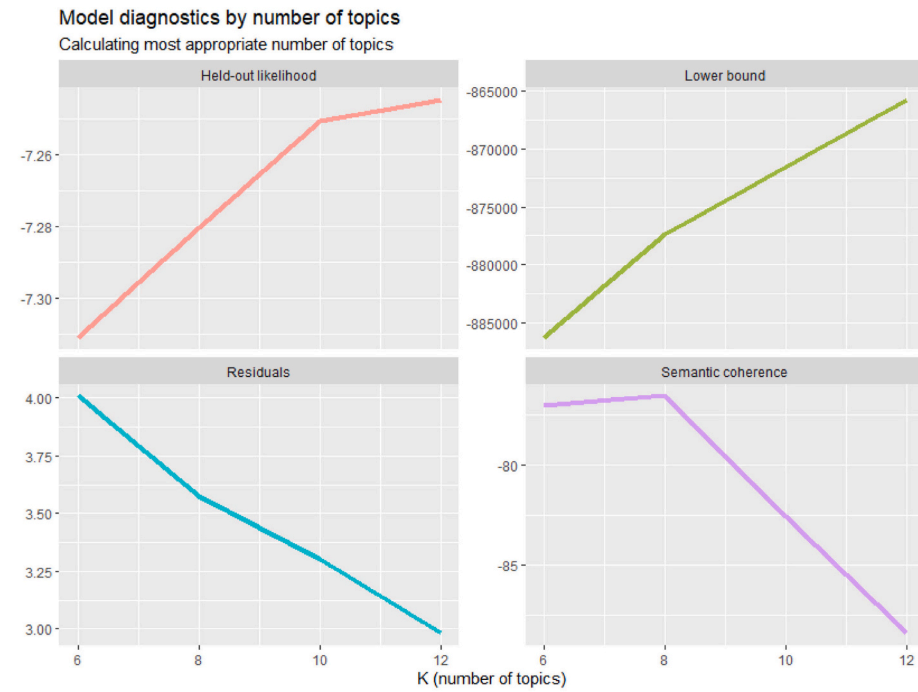

Fig. 5. Model diagnostics by number of topics.

(Schwartz-Shea & Yanow, 2012). To provide an overview of each topic, we summarised and analysed the 10 papers with the highest probability of being associated with the respective topic (see Top ranked references in Appendix I). We applied a content analysis approach that combined both deductive and inductive procedures (Finfgeld-Connett, 2014). The association of articles with the eight predefined topics was conducted in a deductive and automated manner, leveraging the STM results. Subsequently, the content of these articles was analysed inductively, allowing us to assign article content to codes that were generated under the respective main topics. This twofold approach ensured both structured categorisation through automated document-topic association and flexibility in uncovering nuanced insights through inductive interpretation. As a result, this method can be characterised as an innovative, semi-automated content analysis approach that integrates deductive precision with inductive depth. Through this literature analysis, we also identified key future research requirements.

The integration of computational and qualitative analyses played a pivotal role in informing the interpretive phase of the research process. The computational analysis provided a structured, data-driven foundation by identifying key topics across the dataset, offering an objective lens for exploring the underlying themes. Building on this, the qualitative analysis added depth and context, enabling a nuanced understanding of the findings by interpreting them within their broader conceptual and practical frameworks. Together, these complementary approaches ensured a rigorous and well-rounded interpretation of the results, bridging quantitative and qualitative insights to derive meaningful conclusions and actionable research directions.

In the following section, we present the results of our literature analysis.

4. Results: Topic prevalence and interpretation in data-related articles in the DGRL

We first provide the results of the STM analysis, followed by an in-depth interpretation of the identified topics with insights from the most relevant papers.

4.1. Topic prevalence and trends

The STM analysis resulted in an automatised generated topic structure comprising eight distinct topics, each accompanied by relevant top words, as illustrated in Fig. 6. The topics are displayed according to their decreasing prevalence. The topic of innovation and participation has the highest prevalence, at 20,4 %. It is followed by the topics of standardisation and data management, dynamic capabilities, and open government data international. Behind topics 1 and 4, there is a significant leap in topic prevalence. Smart cities hold a topic prevalence of 10.5 %, closely followed by the topic of privacy, security, and trust. The last two positions are occupied by data analytics techniques and geospatial data.

Figs. 7 and 8 illustrate the temporal development of the topics, highlighting their trends over time and providing insight into how their prevalence has shifted throughout the observed period between 2005 and 2022. Topic prevalence increased for topics 1, 3, and 4, was stable for topic 7, and decreased for topics 2, 5, 6, and 8.

4.2. In-depth topic description and interpretation from an ecosystem lens

We provide a detailed discussion of each topic shown in Fig. 6, examining their trends in digital government research as illustrated in Figs. 7 and 8. For each topic, we include a comprehensive summary of the ten most relevant papers, accompanied by an interpretive discussion of collaboration within data ecosystems. Appendix I presents the top-ranked articles, showing the Gamma coefficient for each paper and the full reference. All publications are cited in the main text, with full references available in the reference list. Each topic description follows a strict structure:

1.) Topic label and development of topic prevalence.
2.) Comprehensive summary of the top 10 papers for each topic.
3.) Interpretation through alignment with the data collaboration framework introduced in Fig. 2, highlighting the key findings within the relevant areas of the framework.

Fig. 6. STM topic prevalence.

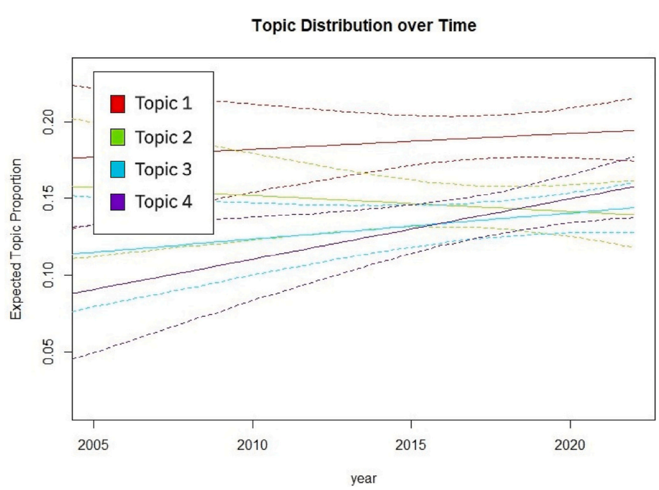

Fig. 7. Topic prevalence 2005–2022, 1–4.

4.2.1. Innovation and participation

The first topic with the highest topic prevalence includes articles focusing on innovation, citizens, and participation. It underlines the participation-oriented investigation of data in digital government research. The topic also contains the top word "ecosystem", an exclusive top word representing a central concept in data sovereignty. Topic prevalence has increased constantly since 2005.

The topic addresses the involvement of different domains and participants in exchanging data inter-organisationally in data ecosystems. Data ecosystems build a socio-technical network of individual and corporate participants, technology, and governance structures to exchange data for creating data value (Iftikhar Hussain Shah et al., 2021). Puposes to exchange data are elaborated in manifold ways and specify further the purpose of creating data values. Slobodova (2020) extracted five significant objectives for contributing to open data initiatives from her literature review: economic growth, efficiency, transparency and accountability, participatory democracy, and social solidarity economy. More general dominant puposes are public sector innovations (Kassen, 2020), citizen engagement (Cordasco et al., 2017; Kassen, 2019) and data collaboration for the public good (Susha et al., 2019). Susha (Susha et al., 2019, p. 228) referred to the concept of data collaboratives as "a novel form of cross sector and public-private partnerships catalyzed by the big data trend and the need to address complex societal challenges collectively".

Overall, this topic deals with the economic dimension in our conceptual framework for collaboration in data ecosystems (see Fig. 2). On the domain layer of our framework, the digital government literature considers ecosystem participants in government, business, industry, and science domains. Overarching societal questions, like sustainable development goals, are pressing policy issues and can be seen as higher-level objectives that involve several sectors. In addition to the four domains in the taxonomy, the literature considers citizens as a further domain, specifically important for public administration and politics in the context of empowerment and participation of citizens to solve public problems and to drive positive societal impact. On the purpose layer, creating added value from data can be seen as a superordinate purpose to share and collaborate on data. Innovation, interaction, and transaction, as defined by the taxonomy of data ecosystems, seem to be mechanisms aimed at achieving this higher-level purpose. According to this, the research on digital government has mainly focused on innovation and participation mechanisms, while the transaction mechanism has received less attention.

4.2.2. Standardisation and data management

The second topic is assigned to standardisation and data management, stemming from top words such as ontologies, metadata, or schema. Despite Europe's efforts to establish sophisticated data infrastructures for sovereign data exchange, the topic has shown a decline in prevalence.

The topic summarises standardisation efforts in the context of the open data movement, which is discussed in detail in section 4.2.4. Poor

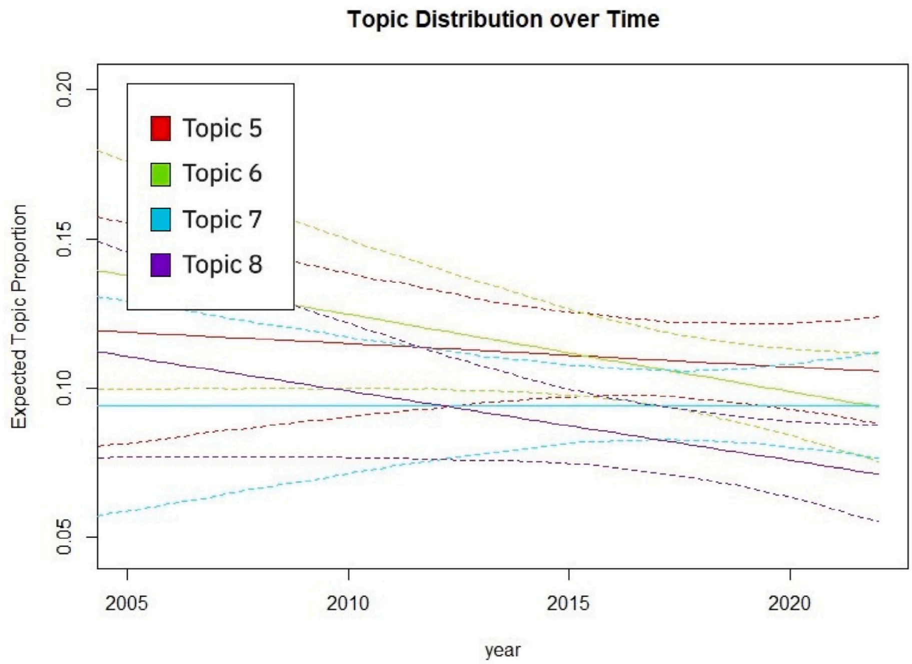

Fig. 8. Topic prevalence 2005–2022, 5–8.

data quality is a common concern due to inconsistent formats and standards for creating and storing open government data (Bulazel et al., 2016; Missier et al., 2003). Diverging ontologies create obstacles for data-sharing initiatives and challenge digital public and corporate services (Musyaffa et al., 2018; Stranacher et al., 2013). Web portals like national open data portals and the European Data Portal provide storage and sharing capabilities for open data, which can be accessed, processed, refined, and distributed freely (Kirstein et al., 2019a). To ensure interoperability between European open data portals, a metadata standard called Data Catalogue Application Profile (DCAT-AP) is used based on the Resource Description Framework (RDF) (Musyaffa et al., 2018). This approach, known as Linked Data, structures data to create interlinked schemas that enhance semantic queries and promote semantic interoperability. The approach involves using Uniform Resource Identifiers (URI) to name common concepts in open government data (Bulazel et al., 2016). URI combine Uniform Resource Locators and Uniform Resource Names to identify and link entities uniquely. Bulazel et al. (2016) highlighted the Comprehensive Knowledge Archive Network (CKAN) as a web application that facilitates the building of data catalogues and provides a Linked Data interface. The CKAN extension DCAT-AP enables the integration of metadata from other catalogues using RDF documents serialised with DCAT.

Overall, the topic relates to the governance meta-dimension in our data collaboration framework (see Fig. 2). It addresses metadata standards in the open data environment applied by public open data portals on national and supranational levels in Europe. The linked data approach aims to improve data management and data governance for open government data in the public sector by drawing upon improved data availability and data cross-linkage and promoting standardised ontologies. Here, digital government research can pay much more attention to contributing to standardised ontologies as a basis for open and closed data descriptions. Linked data as a schema to structure data depends on the semantics used to describe Uniform Resource Names. These terms must originate from the organisational and functional systems underlying these resources, rather than being derived solely from a technical perspective.

4.2.3. *Dynamic capabilities*

The topic in the third position links to organisational capabilities, considering top words like capability, organisation, or local. Its prevalence increased in the last two decades.

This topic focuses on advancing data-sharing initiatives, which involve improving administrative architecture and analytic capabilities to meet the growing demand for data usage (Gong & Janssen, 2021; Sayogo & Pardo, 2011; Wanckel, 2022). Academic sources also discuss the importance of absorptive capacity or dynamic capabilities, which refers to the ability of data organisations to adapt their data processes to changing demands and external environments (Cohen & Levinthal, 1990; Huber et al., 2020; Zahra & George, 2002; Zeleti & Ojo, 2019). Qualitative research suggests that successful data infrastructure requires responsive capacities and capabilities to accommodate diverse contexts, as no single structure is suitable for every scenario (Allard Scott et al., 2018; Choi et al., 2021; Wanckel, 2022). Academic research highlights the benefits of sharing data and using analytics to inform policy decisions. However, there needs to be more understanding of how governments can effectively implement these practices and overcome obstacles to realise their potential value (Allard Scott et al., 2018; Wanckel, 2022). Qualitative fieldwork suggests that the ease of linking and sharing data resources and the administrative structures of government institutions are key factors in realising the potential value of data sharing and analytics (Allard Scott et al., 2018; Sayogo & Pardo, 2011). Various studies focused on integrating data sharing and analytics technology in organisational decision-making processes. These studies explored not only policy-related decisions but also local actions. The practical applications of this integration include, for example, combating cyber threats (Bihl et al., 2020), improving trade customs administration supervision methods (Rukanova et al., 2021), enhancing management in social, educational, and health services (Campana et al., 2019), and establishing new networks in environmental sciences (Choi et al., 2021; Sayogo & Pardo, 2011; Zeleti & Ojo, 2019).

Overall, the topic focuses on developing organisational capabilities to enhance data analytics and sharing in government agencies and public sector organisations. It highlights the need for responsive capacities and capabilities, the value of data sharing and analytics in policy design, and practical applications in various domains. Dynamic capabilities are not included in our data collaboration framework (see Fig. 2); instead, they pertain to the capabilities of each participant within the ecosystem. This indicates that digital government research extends beyond the ecosystem approach, highlighting the necessity of robust and agile participants as a prerequisite for successful collaboration in data ecosystems.

4.2.4. Open government data international

The fourth topic revolves around open government data (OGD), one of the main application fields for data management, with increasing topic prevalence since 2005. As such, open data is an ambitious initiative that seeks to integrate all social and political actors in a coherent information ecology. Governments worldwide are developing strategies to incorporate open data in their jurisdiction. Related articles on this topic are a collection of open data activities worldwide, including Indonesia (Fitriani et al., 2019), Pakistan (Khurshid et al., 2022), Canada (Marczak & Sieber, 2018), Gulf countries (Mutambik et al., 2022; Saxena, 2018), and Korea (Seo et al., 2021). Scholars have found that the effective implementation of open data programmes largely depends on the legislative framework that is already in place within the relevant jurisdiction, whether it is at the national (Saxena, 2018), subnational (Marczak & Sieber, 2018), or other level.

In addition to the apparent impact of technical infrastructure (Mutambik et al., 2022), the effectiveness of open data programs is influenced by human capital and national cultural dimensions (Krishnana & Lymmb, 2016; Saxena, 2018). Sociological theory-based literature suggests that citizens' perceptions are crucial in determining their willingness to interact with digital platforms. Research reveals that users are more likely to engage with digital platforms if they perceive the system as easy-to-use and practically useful (Khurshid et al., 2022). Furthermore, people's trust in government initiatives can influence their intentions to use such platforms, and successful open government data schemes can reinforce trust in national institutions (Fitriani et al., 2019; Purwanto et al., 2020). The quality of OGD is vital for its widespread use. It includes architectural capabilities, robustness, and other factors. Scholars suggest assessing legislation that mediates disputes over licenses for data provision on OGD platforms and comparing metrics and methodologies to measure governments' progress in OGD initiatives. New quality control approaches aim to create a comparative framework that combines different criteria to determine OGD quality (Seo et al., 2021; Zuiderwijk et al., 2021).

The topic explores factors that influence the successful implementation of open government initiatives, such as legislative frameworks, information infrastructure, human capital, and cultural dimensions. Additionally, it highlights the role of trust in user engagement and the importance of quality control in several countries worldwide. In our data collaboration framework (see Fig. 2), human capital, cultural dimensions, and legislative frameworks are not addressed. These aspects are more related to the capabilities of individual and organisational participants in the ecosystem. This suggests a differentiation between the ecosystem, organisational, and individual perspectives in the scientific discussion. Concerning the mentioned information infrastructure as a further success factor, the technical meta-dimension of the data collaboration framework was addressed. The technical architecture involves questions on the technical ecosystem design and decisions about centralised or decentralised systems that should be aligned with organisational and functional requirements and therefore be made by interdisciplinary decision boards. This relates to the question of which administrative-level (European, national, federal, local, and organisational) should be involved in data ecosystem design processes and who can take on the most promising role of a convener to constitute data ecosystems in the government domain. The latter demonstrates how closely architectural design decisions are related to the purpose and the composition of participating stakeholders.

4.2.5. Smart cities and internet of things

The subsequent topic can be labelled "Smart city and IoT", with a topic prevalence that is slightly decreasing. Smart cities use technology to improve urban life, sustainability, and equality and bring two different disciplines on the same page: the technology-related, which deals with embedding the hardware and software in existing infrastructures, and the organisation theory-related one that addresses the impact of these technologies on urban structures and living environments (Batty, 2020).

A pertinent element is the augmentation of computational big data networks reliant on Internet of Things (IoT) infrastructures (Martinez-Gil et al., 2020; Parameswaran et al., 2022; Rajendran et al., 2022; Sudhakar Yadav et al., 2022). Information and Communication Technologies have led to a wide range of innovative ideas for urban governance and infrastructure management, like smart wastewater governance (Du et al., 2019), disaster management (Akerkar, 2018; Gheisari & Esnaashari, 2018), electric vehicle integration (Hussain et al., 2020), and fire department dispatches (Parameswaran et al., 2022). Smart villages aim to improve infrastructure management and planning in rural areas through digitalisation (Martinez-Gil et al., 2020). Integrating electric vehicles in transport-oriented smart cities powered by smart grids to reduce the carbon footprint is also discussed (Hussain et al., 2020). The sensors of ubiquitous devices used for smart control systems generate an enormous variety of data. Different data frequently have contradictory processing means. As such, appropriate investigative methods, depending on the data type and modalities, are also discussed. Real-time processing and Big Data analytics techniques are essential in processing data and forecasting (Gheisari & Esnaashari, 2018). An example of cutting-edge research relating to finding solutions to traffic congestion is the framework Decision-Making System using Ensemble Learning, which employs extreme machine learning and random forest algorithms for the decision-making process (Sudhakar Yadav et al., 2022).

From the ecosystem perspective, smart cities are data ecosystems that focus on rural or local data. Constructing data ecosystems for smart cities is a widely discussed application case, which suggests the concrete scope of data ecosystem boundaries to be within cities or rural areas. The concrete application scenarios range from wastewater and disaster management to electric vehicle integration and demonstrate concrete societal purposes aimed at solving local policy problems through data collaboration. On the economic dimension of our data collaboration framework in Fig. 2, these application scenarios indicate the involvement of service providers of general public interest, such as water companies, waste disposal, or fire brigades in the government or industry domain. The predominant data type is sensor data or data provided in IoT infrastructures. Building these data collaboration infrastructures on a sovereign base and integrating sovereignty into the design process of smart city data ecosystems seems to be a promising future research field. This is already the practice case, but should be recognised also in digital government research.

4.2.6. Privacy, security, and trust

The sixth topic pertains to privacy and security issues with falling topic prevalence development. The topic highlights the risks of cyberattacks and security breaches on digital data structures. Policymakers are working to protect data privacy and combat digital criminal activity. State institutions deal with confidential citizen information, so privacy is a significant concern. The General Data Protection Regulation (GDPR) forces robust anonymisation and pseudonymisation technologies to ensure compliance with Europeans' right to confidentiality (Karasek-Wojciechowicz, 2021; Schulz et al., 2020). Internationally, other states have followed suit with their data protection ventures in digital transformation. Under the top ten articles, China and its Personal Information Protection Law adopted in 2021 (He, 2022) and Australia's new enhanced data protection regime (Alazab et al., 2021) were discussed. Similarly, state actors need to demonstrate their ability to satisfy data protection concerns coming from civil society. Civil society campaigns express apprehension regarding data collection methods and sharing between public and private actors (Hallinan et al., 2020; Lowry & Sexton, 2020; Trenwith & Venter, 2019). Anonymisation and pseudonymisation technologies for data protection can increase the risk of criminal activity. Policymakers need to address this while protecting the privacy of legitimate users (Alazab et al., 2021; Chertoff, 2018; Karasek-Wojciechowicz, 2021).

As previously implied, digital data initiatives must demonstrate their willingness to listen to civil society's concerns about wavering trust. Trust is essential for compliance in an open data digital environment (McLeod, 2020), which is ultimately vital for the success of public services reliant on data resources (Lokhorst & van Eechhoud, 2020). Moreover, a democratic society needs to guarantee freedom to its subjects (Hallinan et al., 2020).

Overall, the topic highlights the importance of safeguarding privacy and security in the digital era. It emphasises the role of regulations in protecting individuals' privacy rights and discusses the need to combat cybersecurity threats. Building trust through addressing civil society's concerns is crucial for ensuring compliance and the success of digital data initiatives. However, the topic's prevalence is decreasing, but it addresses exactly the European shift to develop secure and privacy-sensitive data infrastructures based on sovereignty and trust. This highlights the necessity for interdisciplinary research on data collaboration, as it bridges technological and governance-related aspects within the context of our data collaboration framework (see Fig. 2).

4.2.7. Data analytics techniques

The second-to-last topic is linked to data analytics, and topic prevalence develops stably.

The topic covers the importance of data analytics methodologies in data initiatives. Data analytics helps data scientists to structure, evaluate, and gain insights from data. Clustering data groups is based on defining characteristics for more precise analysis. Machine learning is used to automatically analyse large amounts of data, identifying behavior patterns, for example, to prevent cyberattacks (Abdullayeva, 2022; Banerjee et al., 2021) or even real-world smuggling operations (Wena et al., 2012). Other patterns of behavior frequently analysed by researchers relate to politics (Clark et al., 2018) and the media's influence on it (W. H. Lo et al., 2021), commerce (Rao & Chandran, 2013) or finance (Xuemiri et al., 2006). Social media appears to be an ever more prominent source of data, mainly text data, since social media serves as a platform for expressing diverse political and social opinions, which are readily available for researchers to access. Hence, social media will continue to affect social research and the direction of analytics profoundly (Li et al., 2022; Yaqub et al., 2018, 2020).

In sum, the topic addresses specific, particularly analytical capabilities, encompassing methodological data analytics expertise or competence as well as respective software and hardware availability for data analyses. Individual employees possess the data expertise necessary at the personal level, highlighting the need for an individual layer in our data collaboration framework. On one hand, software and hardware can be utilised as a service, with cloud service providers playing a significant role in enhancing analytical capabilities within the data ecosystem. On the other hand, the development of analytical capabilities addresses the organisational level, where strategies must be formulated and operationalised to create conditions conducive to data-driven processes and decision-making. As a prominent source of data, social media plays a relevant role in literature as a specific data type on the technical dimension in our data collaboration framework in Fig. 2.

4.2.8. Geospatial data

The final topic 8 indicates the prominence of spatial data in the data-related academic discussion, but topic prevalence is decreasing slightly.

Geospatial data focuses on Earth's surface and its characteristics. It is essential in modernising rural areas and developing land policies. Information technologies can facilitate an extensive land administration system. The quality of such systems is of utmost importance as it has significant consequences for industry (Esperanca & Samet, 2000), census data, land and forest rights (Aydinoglu & Bovkir, 2017; Choudhury & Behera, 2017), trade, transport (Chomkhamsri et al., 2017; Chua, 2003), real estate development (Aydinoglu & Bovkir, 2017), agriculture (Huffman et al., 2018), tourism (Lo Duca & Marchetti, 2019), and last but not least, combating crises caused by climate change (Chomkhamsri et al., 2017). Beyond commerce and civilian infrastructure, geospatial data is salient in military initiatives to enable quick, responsive and flexible collaboration of telemetry systems (McGee & Graves, 2008). Thus, researchers found it essential that policymakers integrate policies which operate by considering the latest discoveries in Information Science, i.e. using state-of-the-art design patterns in URI strategies or specialised scripting languages (Abbas & Ojo, 2014; de Assis-Rodrigues et al., 2016; Esperanca & Samet, 2000).

In conclusion, the topic portrays geospatial data as a valuable resource for effective land administration, development, and decision-making in various sectors. In the data collaboration framework in Fig. 2, geospatial data can be classified as data type on the technical level. The government can be a provider and user for several land-related policy problems, and the topic demonstrates a broad discussion in digital government research.

5. Discussion and research agenda

This study offers a systematic computational analysis of prevalent topics in digital government research, which were analysed algorithmically and interpreted manually. In the first part of the discussion in Section 5.1, the topics identified are systematically mapped onto the data collaboration framework, offering a comprehensive overview. This mapping not only refines and extends the conceptual framework but also highlights existing research gaps that warrant further exploration. In Section 5.2, we develop the future research agenda, structured around the three meta-dimensions of the conceptual framework, providing a clear path for addressing these gaps and advancing the field.

5.1. Expanded data collaboration framework

The results show that data collaboration in digital government research impacts multiple research domains and is positioned at the intersection of the three meta-dimensions of a data ecosystem: economic, technical, and governance. Fig. 9 presents our expanded framework, in which aspects discussed in the literature are marked with an asterisk (*), while gaps in the literature are indicated by the absence of an asterisk. Both aspects are discussed in detail in the following Section 5.2.

Additional aspects that expand the framework are indicated with a hash (#). The literature on various topics influences not only the ecosystem perspective, but also the organisational and individual perspectives. This, in turn, broadens our conceptual framework for data collaboration. Specifically, topic 3 "Dynamic Capabilities" focuses on the organisational level, highlighting capabilities that enhance an organisation's maturity to engage in intra- or inter-organisational data collaboration. Authors discuss intensively the cultural factors that shape organisations. Furthermore, the literature emphasises human capital as a vital asset, particularly regarding data expertise within the technical dimension. For these reasons, we have added two vertical layers to our framework that specifically address the organisational and individual levels.

5.2. Future research agenda

The three meta-dimensions of the data collaboration framework demonstrated stability throughout our analysis, thus providing a robust foundation for structuring our discussion on a future research agenda for data collaboration. For each meta-dimension, we examine the current state of knowledge as reflected in the literature and identify potential unresolved research challenges, which we visualise through a focused view within the expanded Data Collaboration Framework. The

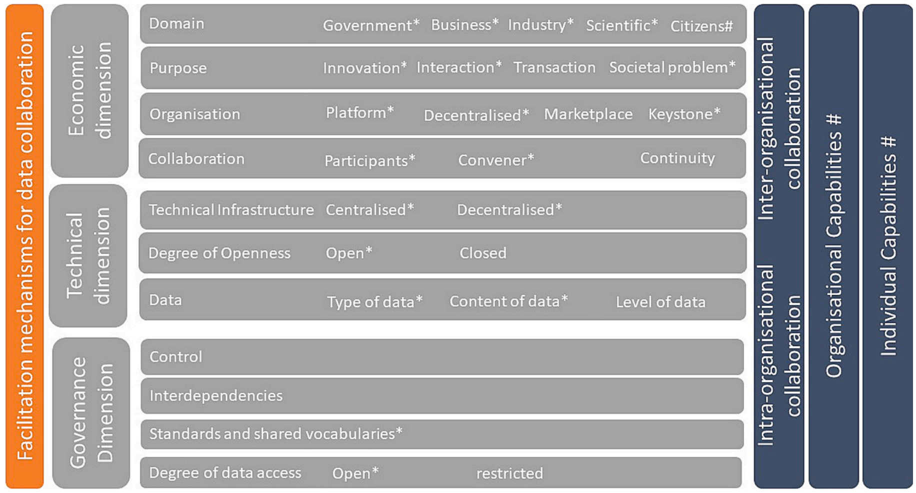

Fig. 9. Expanded Data Collaboration Framework.

discussion conclude with the overview Tables 1-3 that illustrate the progression from established findings in the literature to unexplored areas, which we emphasise through exemplary research questions. These research areas are intended to highlight gaps in the field. The derived research questions do not claim to be comprehensive. Moreover, these research areas can be examined from various methodological approaches, including, for example, explorative, descriptive, explanatory, or theory-testing angles.

5.2.1. *Economic dimension*

The economic dimension outlines different domains such as government, business, industry, and scientific sectors (see Fig. 10). Although the literature on data collaboration predominantly draws on concepts of inter-organisational or cross-boundary data sharing and ecosystem theory—rarely employing the term "data space"(Dawes et al., 2012; Gil-García & Sayogo, 2016; Susha & Gil-Garcia, 2019; Yang & Maxwell, 2011) —the term has gained practical significance. The European Union plans to establish common European data spaces for major application domains at a cross-national level, reflecting the practical adoption of the concept beyond its limited use in administrative sciences. The literature acknowledges the role of citizens, which we identify as an important additional domain that complements our data collaboration framework (Cordasco et al., 2017; Kassen, 2019; Slobodova, 2020).

Participation and innovation stand out as having the highest prevalence and increasing popularity in the analysed literature. This pertains to the economic dimension of the data collaboration framework, which involves the participation of various stakeholders. These participants actively share and exchange data, facilitating a more integrated approach to create added value (Slobodova, 2020). Additionally, innovation is highlighted as a key purpose of data collaboration, contributing significantly to the economic aspects by fostering new opportunities and solutions (Kassen, 2020). On the local level, data ecosystems and data spaces appear indirectly in the context of the smart city concept and urban theory (Batty, 2020; Martinez-Gil et al., 2020). Future research should therefore focus on how data ecosystems can be constructed and interconnected to encourage active participation, incorporating public administrations at various administrative levels. Gaps in the economic dimension centre on transaction mechanisms, various organisational models, and questions regarding the continuity or boundaries of a data ecosystem for data collaboration (see Fig. 10). It is not clear what the focus of a data ecosystem should be, and how different purposes impact the scope of a data ecosystem. Additionally, the transaction patterns are not well-defined, and future digital government research could contribute to elaborating specific transaction mechanisms based on different data-sharing purposes.

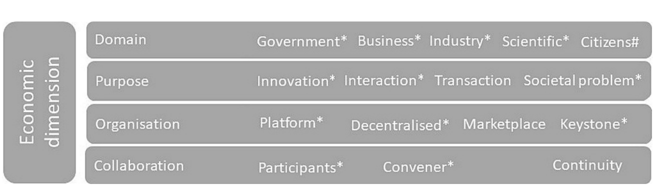

Fig. 10. Focused view on the economic dimension.

Table 1
Research agenda for data collaboration on the economic dimension.

Existing knowledge	Unresolved research problems	Exemplary research questions
- Data ecosystems are designed to create added value. - Data ecosystems can contribute to the creation of public value. - Participation and innovation are extensively discussed in the literature, including citizens as an additional domain. - Inter-organisational and cross-boundary stakeholder collaborations form the foundational framework of data ecosystems. - Smart City captures the discussion often at the local level.	- The process of generating added value as a unifying objective for data collaboration under conditions of data sovereignty remains insufficiently examined in current research. - The concept and scope of data collaboration purposes within data ecosystems lack a clear definition and consensus. - Aspects influencing the scope, boundaries, and interconnections of data ecosystems have not been widely examined in the analysed literature. - Stakeholder participation beyond local administrative levels is underrepresented in data ecosystem integration efforts.	- How does data sovereignty affect the public value creation through data collaboration? - How can the boundaries of data ecosystems be effectively designed? - How is the impact of different purposes on the scope and boundaries of data ecosystems? - How can data ecosystems be implemented and interconnected to maximise the participation of public administrations across different administrative levels?

5.2.2. *Technical dimension*

The technical dimension involves the infrastructural nature of data ecosystems, whether they are centralised or decentralised, how open they are for participation, what kind of data can be shared and collaborated on The analysed literature focuses on several data types, particularly on social media data (Li et al., 2022; Yaqub et al., 2018, 2020), sensor data (Martinez-Gil et al., 2020; Parameswaran et al., 2022;

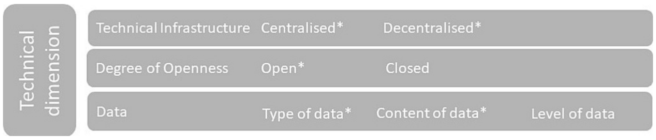

Fig. 11. Focused view on the technical dimension.

Table 2
Research agenda for data collaboration on the technical dimension.

Existing knowledge	Unresolved research problems	Exemplary research questions
- The discourse predominantly centers on the dichotomy between centralised or decentralised technical infrastructures in data ecosystems.	- The interdependencies between stakeholder organisations and the architectural design of data ecosystems remain insufficiently understood. - The degree of openness remains largely unexamined in the analysed literature, owing to the prevailing focus on open data portals.	- How do different typologies of relationships between participant groups influence architectural design decisions? - How can varying degrees of openness in data ecosystems be conceptualised?

Rajendran et al., 2022; Sudhakar Yadav et al., 2022), and geospatial data (topic 8), but underrepresents the degree of openness of a data ecosystem and the administrative level of data, indicating a need for further research in these areas (see Fig. 11).

Further, technical infrastructure design decisions are significantly impacted by the organisational structure of the data ecosystem. So far, discussions have focused on either centralised platform approaches (Alhawawsha & Panchenko, 2020; Bagheri et al., 2021; Bonina & Eaton, 2020; Linåker & Runeson, 2021) or decentralised systems (Hawig et al., 2019; O. Lo et al., 2022; Soundarabai et al., 2021). Depending on the relation between various ecosystem participants, a data ecosystem should be designed in a centralised or decentralised manner. We recommend that future research in the field of digital government focus on typologies of relationships between different participant groups in order to address the dependencies between organisational and technical aspects from an administrative science perspective.

5.2.3. Governance dimension

The governance dimension emphasises control, interdependencies, standards, and varying degrees of data access from open to restricted. Issues of control, interdependencies, and the use of restricted data as a governance mode are less intensively addressed in the analysed literature (see Fig. 12). Data sovereignty is an integral part of data governance, as it is closely related to the control and independence of data management. The concept of data sovereignty doesn't occur as an associated top word in the context of data collaboration in digital government research, but receives rising attention in European and federal data strategies. The research discussion has revolved so far more around privacy and security (Agbozo et al., 2018; Bargh et al., 2016; Graham et al., 2016; Habibzadeh et al., 2019; Hallinan et al., 2020). Ultimately, research must consider the role of sovereignty in facilitating effective collaboration in line with European legislation (Data Act, Digital Market Act, Data Governance Act, Open Data Directive). The considered literature in this area tends to predominantly focus on open data or open

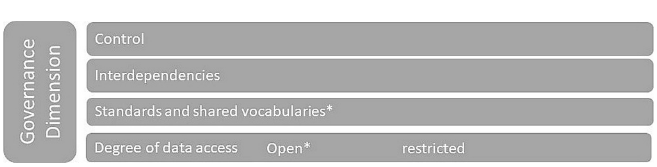

Fig. 12. Focused view on the governance dimension.

Table 3
Research agenda for data collaboration on the governance dimension.

Existing knowledge	Unresolved research problems	Exemplary research questions
- A significant emphasis is placed on the role and utilization of open (government) data within data ecosystems. - Research discussions primarily focus on issues related to privacy and security in data ecosystems. - There is a decreasing emphasis on topics such as standardisation, data management, and the interconnected themes of privacy, security, and trust. - Metadata standards are a key area of exploration, particularly in the context of open data portals.	- Effective facilitation mechanisms for data sharing under restricted access conditions are not yet well-defined. - Data sovereignty is underrepresented in research discussions, despite being highly emphasised in legislative efforts surrounding data collaboration. - The development of standardisation and data management practices within advanced sovereign data infrastructures remains insufficiently explored. - The establishment and application of metadata standards for restricted data lack comprehensive investigation.	- What does public administration require to facilitate the sovereign management of restricted data within data ecosystems? - How does the concept of data sovereignty affect control mechanisms and interdependencies within data ecosystems? - How is European legislation—such as the Data Act, Digital Markets Act, Data Governance Act, and Open Data Directive—implemented in data collaboration within public administration?How do the technical data space concept and the organisational data mesh approach adapt to and manage environmental dynamics? - How should semantic ontologies and controlled vocabulary be designed to describe data resources meaningfully?

government data, while providing limited attention to restricted data access (Huber et al., 2020; Kassen, 2020; Purwanto et al., 2020; Zeleti & Ojo, 2019). This suggests that there is a research gap that demands further exploration, with emphasis on facilitation mechanisms within restricted data access conditions. Future research should focus on whether federated data ecosystems, based on the technical data space concept and the organisational data mesh approach, can efficiently and effectively facilitate restricted data. It should also examine whether the infrastructure and governance framework are agile enough to cope with environmental dynamics.

Furthermore, following the STM analysis, the current trend of decreasing prevalence of topics related to standardisation and data management, as well as privacy, security, and trust, is going against the practical movements towards federated data ecosystems with sophisticated sovereign infrastructures. The European Union is pushing for the implementation of federated data infrastructures driven by practical applications rather than digital government research. In this dynamic technological environment, the role of standards should be investigated in depth. So far, the main focus lies on the metadata structure and implementation of metadata schemes in open data portals (Barthelemy et al., 2022; Bulazel et al., 2016; Kirstein et al., 2019b) A fundamental part of the data asset descriptions is semantic descriptions, often expressed through controlled vocabulary to ensure semantic interoperability. Future research has to focus on how semantic ontologies and controlled vocabulary should be designed to describe data resources meaningfully.

5.3. Limitations of the study

Our study is not without limitations, which include, in our case, mainly the data quality and our data interpretation approaches.

First, using the STM approach, we identified the latent topic structure, which indicates the prevalence of certain topics in the Digital Government Reference Library. It allowed us to understand the academic discussion on data-related topics in the digital government research field in a specific period. One limitation of STM is that the output depends on the quality of the input data provided to the algorithmic model (Guenduez & Mettler, 2022). No database is complete, and in this case, there might be a selection bias based on the individual decisions made by the creator of the database, and studies might be missing. Given the focused and widely recognised nature of the articles within the DGRL data source, which are specifically tailored to our field of interest, we decided to accommodate this limitation. As a further development of our research approach, future studies could expand the scope by including additional journals to enrich and complement the analysis of digital government research, thereby offering a broader and more diverse perspective on the field.

A second limitation of our approach relates to specificity and data quality (Guenduez & Mettler, 2022). Topics with high political or societal relevance may not appear if a critical proportion of the text corpus does not address them, meaning topic modelling can only depict the main content. While Chen et al. (2023) confirmed that topic modelling identifies key topics and provides timely, objective results, it faces challenges such as topic overlaps and spurious terms, potentially leading to gaps in coverage. This limitation is common in topic modelling and systematic analyses, where manual evaluation of large datasets is labour-intensive and time-consuming. Sharma et al. (2021) highlighted that bibliometric studies produce results that evolve over time, complicating reproducibility, particularly in dynamic fields like information management and technology. This applies to our study, where evolving topics in digital government, such as the rise of artificial intelligence, underscore the need for periodic analyses.

Finally, in our research design, we combined the computational systematic approach to data collection and analysis with our qualitative interpretation of the topics. On the one hand, interpreting words in isolation from the textual material they appear in may lead to misinterpretations, and, therefore, more in-depth analysis must be conducted to validate and augment STM results through qualitative text interpretation (Guenduez & Mettler, 2022). On the other hand, calculated topics can also contribute to exploratory research approaches, which are traditionally more akin to qualitative research (Stebbins, 2001). We aimed to overcome our interpretation limitations of the resulting huge dataset by triangulating the different data points (Bryman, 2006; Della Porta & Keating, 2008; Takahashi & Araujo, 2019).

6. Conclusions

This study addressed the critical gap in understanding how government entities share and collaborate on both open and restricted data, a fundamental aspect of the European Data Strategy. By systematically analysing the literature on data collaboration within the digital government field, we provided a deeper exploration of this topic. Our proposed data collaboration framework, which incorporates the ecosystem, the organisational, and the individual layers, underscores the necessity of a multidimensional approach in addressing data collaboration challenges.

The prevailing focus on innovation and participation in the analysed literature reveals a significant trend towards participatory approaches in digital government research. However, we also identified a concerning decline in discussions around standardisation and data management as well as privacy, security, and trust, highlighting a disconnect between the academic attention and the actual advancements in the practice of federated data ecosystems. Furthermore, our findings indicate that the predominant emphasis on open data has resulted in the neglect of critical data that is subject to restricted access.

As a result of our comprehensive computational and qualitative analysis, we are contributing an advanced research framework for the study of data collaboration, including a set of future research questions that should be addressed using empirical evidence. This study highlights the need for future research to investigate how various purposes shape the focus and scope of data ecosystems. Additionally, exploring participant relationships and assessing the efficiency of federated data ecosystems will provide valuable insights for managing restricted data and adapting to environmental changes. Understanding the role of sovereignty in facilitating collaboration within the framework of European legislation, along with developing meaningful semantic ontologies and controlled vocabularies, will enhance the clarity and usability of data resources in these ecosystems.

Our study contributes both theoretically and practically, and it addresses a significant methodological gap in political science and public administration by applying the novel Structural Topic Modelling technique to the analysis of data-related topics in peer-reviewed literature. Despite its potential for uncovering nuanced thematic structures in large text corpora, STM has remained largely underutilised in these disciplines. By demonstrating the value and transparency of STM for systematic literature analysis, our work contributes both to methodological innovation and to the advancement of research practices in political science and public administration.

Overall, our comprehensive approach aims to significantly advance the field and to encourage further exploration of underrepresented research problems in data collaboration.

CRediT authorship contribution statement

Juliane Schmeling: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Sami al Dakruni:** Writing – original draft, Formal analysis, Data curation. **Ines Mergel:** Writing – review & editing, Supervision, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Grammarly in order to improve readability and language. After using this tool, the authors reviewed and edited the content manually as needed and take full responsibility for the content of the publication.

Funding

This work was supported by the Federal Ministry for Economic Affairs and Climate Action, Germany [grant number 68GX21009B].

Declaration of competing interest

This project received funding from the Federal Ministry for Economic Affairs and Climate Action, Germany [grant number 68GX21009B]. The grant was used to develop a reference implementation for Gaia-X data space specifications in the public sector. Gaia-X is an initiative aimed at creating shared data infrastructures across Europe. The literature review laid the groundwork for implementing data spaces in public administrations. The study sponsor had no active role in the study design, in the collection, analysis, and interpretation of data, in the writing of the report, and in the decision to submit the paper for publication.

Appendix A. Appendix I

Top-ranked references for each topic.
Topic 1 – Innovation & Participation.

Top-ranked references for topic I

Gamma	Reference
0,98,428,286	Kassen, M. (2020). Open data and its peers: Understanding promising harbingers from Nordic Europe. Aslib Journal of Information Management, 72(5), 765–785. https://doi.org/10.1108/AJIM-12-2019-0364
0,95,400,092	Kassen, M. (2019). Open data and e-government - related or competing ecosystems: A paradox of open government and promise of civic engagement in Estonia. Information Technology for Development, 25(3), 552–578. https://doi.org/10.1080/02681102.2017.1412289
0,93,780,301	Iftikhar Hussain Shah, S., Peristeras, V., & Magnisalis, I. (2021). Government big data ecosystem: Definitions, types of data, actors, and roles and the impact in public administrations. ACM Journal of Data and Information Quality, 13(2), [1–25] 8. https://doi.org/10.1145/3425709
0,93,343,576	Susha, I., & Gil-Garcia, J. R. (2019). A Collaborative Governance Approach to Partnerships Addressing Public Problems with Private Data. In Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS-52), Wailea, Hawaii, US. http://scholarspace.manoa.hawaii.edu/handle/10125/59726https://scholarspace.manoa.hawaii.edu/handle/10125/59726
0,93,180,916	Susha, I., Rukanova, B. D., Gil-García, J. R., Hua Tan, Y., & Gasco, M. (2019). Identifying Mechanisms for Achieving Voluntary Data Sharing in Cross-sector Partnerships for Public Good In Y.-C. Chen, F. Salem, & A. Zuiderwijk (Eds.), dg.o 2019 (pp. 227–236). ACM. https://doi.org/10.1145/3325112.3325265
0,92,534,096	Slobodova, O. (2020). Motivations of civil society actors to reuse open data. Case study of discrete actors in Lyon. Bitacora Urbano Territorial, 30(3), 55–70. https://doi.org/10.15446/bitacora.v30n3.80740
0,91,942,847	Virkar, S., Viale Pereira, G., & Vignoli, M. (2019). Investigating the Social, Political, Economic and Cultural Implications of Data Trading. In I. Lindgren, M. Janssen, H. Lee, A. Polini, M. P. Rodríguez Bolívar, H. J. Scholl, & E. Tambouris (Eds.), Lecture Notes in Computer Science: Vol. 11,685 (pp. 215–229). Springer International Publishing. https://doi.org/10.1007/978-3-030-27325-5_17
0,91,688,941	Cordasco, G., Donato, R. de, Malandrino, D., Palmieri, G., Petta, A., Pirozzi, D., Santangelo, G., Scarano, V., Serra, L., Spagnuolo, C., & Vicidomini, L. (2017). Engaging Citizens with a Social Platform for Open Data. In (pp. 242–249). ACM. https://doi.org/10.1145/3085228.3085302
0,91,451,948	Susha, I., Schiele, J., & Frenken, K. (2022). Business-to-government data sharing for public interests in the european union: Results of a public consultation. In M. Janssen, C. Csáki, I. Lindgren, E. Loukis, U. Melin, G. Viale Pereira, M. P. Rodríguez Bolívar, & E. Tambouris (Eds.) (pp. 515–529). Springer International Publishing. https://doi.org/10.1007/978-3-031-15086-9_33
0,9,141,531	Kassen, M. (2017). Open data in Kazakhstan: Incentives, implementation and challenges. Information Technology and People, 30(2), 301–323. https://doi.org/10.1108/Itp-10-2015-0243

Topic 2 – Standardisation.

Top-ranked references for topic II

Gamma	Reference
0,97,633,637	Stranacher, K., Krnjic, V., Zwattendorfer, B., & Zefferer, T. (2013). Assessment of Redactable Signature Schemes for Trusted and Reliable Public Sector Data. In W. Castelnovo & E. Ferrari (Eds.) (pp. 508–516). Academic Conferences Limited.
0,97,468,007	Bulazel, A., DiFranzo, D., Erickson, J. S., & Hendler, J. A. (2016). The Importance of Authoritative URI Design Schemes for Open Government Data. International Journal of Public Administration in the Digital Age, 3(2), 1–18. https://doi.org/10.4018/IJPADA.2016040101
0,97,239,776	Shukair, G., Loutas, N., Peristeras, V., & Sklarß, S. (2013). Towards semantically interoperable metadata repositories: The Asset Description Metadata Schema. Computers in Industry, 64(1), 10–18. https://doi.org/10.1016/j.compind.2012.09.003
0,96,672,986	Loutas, N., Goedertier, S., Frade, J. R., Colas, C., Keyzer, M. de, Di Giacomo, D., Dekkers, M., & Peristeras, V. (2013). Realising a federation of repositories of reusable metadata. In 2013 International Conference on Dublin Core and Metadata Applications: Linking to the Future, DC 2013, Lisbon, Portugal.
0,96,341,444	Kirstein, F., Dittwald, B., Dutkowski, S., Glikman, Y., Schimmler, S., & Hauswirth, M. (2019). Linked Data in the European Data Portal: A Comprehensive Platform for Applying DCAT-AP. In I. Lindgren, M. Janssen, H. Lee, A. Polini, M. P. Rodríguez Bolívar, H. J. Scholl, & E. Tambouris (Eds.), Lecture Notes in Computer Science: Vol. 11,685 (pp. 192–204). Springer International Publishing. https://doi.org/10.1007/978-3-030-27325-5_15
0,96,313,011	Shukair, G., Loutas, N., & Peristeras, V. (2007). Integrating linked metadata repositories into the web of data. In 19th International Conference on Advanced Information Systems Engineering, CAiSE 2007, Trondheim, Norway.
0,96,269,674	Stranacher, K., Krnjic, V., Zwattendorfer, B., & Zefferer, T. (2013). Evaluation and Assessment of Editable Signatures for Trusted and Reliable Public Sector Data. Electronic Journal of *E*-Government, 11(2), 360–372.
0,95,808,843	Musyaffa, F. A., Orlandi, F., Jabeen, H., & Vidal, M. E. (2018). Classifying Data Heterogeneity Within Budget and Spending Open Data. In ICEGOV '18 (pp. 81–90). ACM. https://doi.org/10.1145/3209415.3209482
0,95,078,038	Tambouris, E., Manouselis, N., & Costopoulou, C. (2007). Metadata for digital collections of e-government resources. The Electronic Library, 25(2), 176–192. https://doi.org/10.1108/02640470710741313
0,95,004,534	Missier, P., Lalk, G., Verykios, V., Grillo, F., Lorusso, T., & Angeletti, P. (2003). Improving data quality in practice: A case study in the Italian Public Administration. Distributed and Parallel Databases, 13(2), 135–160. 000179704900001

Topic 3 –Dynamic Capabilities.

Top-ranked references for topic III

Gamma	Reference
0,96,031,385	Wanckel, C. (2022). An ounce of prevention is worth a pound of cure – building capacities for the use of big data algorithm systems (BDAS) in early crisis detection. Government Information Quarterly, 39(4), [1–13] 101705. https://doi.org/10.1016/j.giq.2022.101705
0,95,121,103	Rukanova, B., Tan, Y.-H., Slegt, M., Molenhuis, M., van Rijnsoever, B., Migeotte, J., Labare, M. L. M., Plecko, K., Caglayan, B., Shorten, G., van der Meij, O., & Post, S. (2021). Identifying the value of data analytics in the context of government supervision: Insights from the customs domain. Government Information Quarterly, 38(1), [1−20] 101496. https://doi.org/10.1016/j.giq.2020.101496
0,94,679,832	Gong, Y., & Janssen, M. (2021). Roles and capabilities of enterprise architecture in big data analytics technology adoption and implementation. Journal of Theoretical and Applied Electronic Commerce Research, 16(1), 37–51. https://doi.org/10.4067/S0718-18762021000100104
0,93,328,479	Huber, F., Wainwright, T., & Rentocchini, F. (2020). Open data for open innovation: Managing absorptive capacity in SMEs. R & D Management, 50(1), 31–46. https://doi.org/10.1111/radm.12347

(*continued on next page*)

(*continued*)

Gamma	Reference
0,92,933,785	Campana, K., Mills, J. E., & Martin, M. H. (2019). The Role of Community Data in Helping Public Libraries Reach and Serve Underserved Communities. In Lecture Notes in Computer Science (pp. 247–253). Springer International Publishing.
0,91,506,019	Sayogo, D. S., & Pardo, T. A. (2011). Understanding the capabilities and critical success factors in collaborative data sharing network: the case of dataONE. In dg.o '11 (pp. 74–83). ACM. http://doi.acm.org/10.1145/2037556.2037568
0,91,176,698	Bihl, T., Gutierrez, R., Bauer, K., Boehmke, B., & Saie, C. (2020). Topological Data Analysis for Enhancing Embedded Analytics for Enterprise Cyber Log Analysis and Forensics. In (pp. 1937–1946). University of Hawai'i at Manoa. https://doi.org/10.24251/HICSS.2020.238
0,89,851,348	Zeleti, F. A., & Ojo, A. (2019). Agile Mechanisms for Open Data Process Innovation in Public Sector Organisations: Towards Theory Building. In S. Ben Dhaou, L. Carter, & M. Gregory (Eds.) (pp. 164–174). ACM. https://doi.org/10.1145/3326365.3326387
0,89,758,792	Allard Scott, W., Wiegand Emily, R., Schlecht, C., Datta, A. R., George Robert, M., & Weigensberg, E. (2018). State Agencies' Use of Administrative Data for Improved Practice: Needs, Challenges, and Opportunities. Public Administration Review, 78(2), 240–250. https://doi.org/10.1111/puar.12883
0,89,617,687	Choi, Y., Gil-Garcia, J., Burke, G. B., Costello, J., Werthmuller, D., & Aranay, O. (2021). Towards data-driven decision-making in government: Identifying opportunities and challenges for data use and analytics. In (pp. 2183–2192). University of Hawai'i at Manoa. https://doi.org/10.24251/HICSS.2021.268

Topic 4 – Open Government Data.

Top-ranked references for topic IV

Gamma	Reference
0,99,067,577	Marczak, P., & Sieber, R. (2018). Linking legislative openness to open data in Canada. The Canadian Geographer, 62(2), 106–119. https://doi.org/10.1111/cag.12408
0,96,667,727	Khurshid, M. M., Zakaria, N. H., Arfeen, M. I., Rashid, A., Nasir, S. U., & Shehzad, H. M. F. (2022). Factors influencing citizens' intention to use open government data—A case study of Pakistan. Big Data and Cognitive Computing, 6(1), 1–22. https://doi.org/10.3390/bdcc6010031
0,96,139,064	Krishnana, S., & Lymmb, J. (2016). Determinants of virtual social networks diffusion: Insights from cross-country data. Computers in Human Behavior, 54, 691–700. https://doi.org/10.1016/j.chb.2015.07.055
0,95,832,561	Seo, J., Kim, B., & Yeong Kwon, H. (2021). Open data policies analysis disputes mediation cases in Korea: Based on OUR Data Index and ODB. In J. Lee, G. Vilale Pereira, & S. Hwang (Eds.) (pp. 153–167). Association for Computing Machinery. https://doi.org/10.1145/3463677.3463735
0,95,720,328	Saxena, S. (2018). National open data frames across Japan, The Netherlands and Saudi Arabia: Role of culture. Foresight, 20(1), 123–134. https://doi.org/10.1108/FS-07-2017-0038
0,95,462,158	Zuiderwijk, A., Pirannejad, A., & Susha, I. (2021). Comparing open data benchmarks: Which metrics and methodologies determine countries' positions in the ranking lists? Telematics and Informatics, 62(n/a), [1−23] 101634. https://doi.org/10.1016/j.tele.2021.101634
0,95,453,509	Purwanto, A., Zuiderwijk, A., & Janssen, M. (2020). Citizens' Trust in Open Government Data: A Quantitative Study about the Effects of Data Quality, System Quality and Service Quality. In S.-J. Eom & J. Lee (Eds.), dg.o '20 (pp. 310–318). Association for Computing Machinery. https://doi.org/10.1145/3396956.3396958
0,94,854,578	Fitriani, W. R., Hidayanto, A. N., Sandhyaduhita, P., Purwandari, B., & Kosandi, M. (2019). Determinants of Continuance Intention to Use Open Data Website: An Insight from Indonesia. Pacific Asia Journal of the Association for Information Systems, 11(2), 96–120. https://doi.org/10.17705/1pais.11205
0,94,597,617	Fitriani, W. R., Hidayanto, A. N., Sandhyaduhita, P. I., & Purwandari, B. (2017). Determinants of Intention to Use Open Data Website: An Insight from Indonesia. In 21st Pacific Asia Conference on Information Systems (PACIS 2017), Langkawi Island, Malaysia.
0,93,704,634	Mutambik, I., Almuqrin, A., Lee, J., Gauthier, J., & Homadi, A. (2022). Open government data in gulf cooperation council countries: An analysis of progress. Sustainability, 14(12), [1–18] 7200. https://doi.org/10.3390/su14127200

Topic 5 – Smart Cities.

Top-ranked references for topic V

Gamma	Reference
0,97,161,019	Sudhakar Yadav, N., Gogula, S., Krishna Sharma, G., Mallikarjuna Rao, C., & Lalita Parameswari, D. V. (2022). IoT and big data analytics-based intelligent decision-making systems. In S. Rajendran, M. Sabharwal, G. Ghinea, R. Kumar Dhanaraj, & B. Balusamy (Eds.), IoT and big data analytics for smart cities [1–19]. Chapman and Hall/CRC.
0,9,702,331	Du, J., Kuang, B., & Yang, Y. (2019). A Data-Driven Framework for Smart Urban Domestic Wastewater: A Sustainability Perspective. Advances in Civil Engineering, 2019, [1–16]. https://doi.org/10.1155/2019/6530626
0,96,402,704	Liu, Z., & Ota, K. (2018). Preface to Data storages in wireless sensor networks to deal with disataster management. In Z. Liu & K. Ota (Eds.), Advances in public policy and administration (APPA) book series. Smart technologies for emergency response and disaster management (pp. xii–xiv). Information Science Reference.
0,96,307,953	Gheisari, M., & Esnaashari, M. (2018). Data storages in wireless sensor networks to deal with disaster management. In Z. Liu & K. Ota (Eds.), Advances in public policy and administration (APPA) book series. Smart technologies for emergency response and disaster management (pp. 196–222). Information Science Reference.
0,95,989,501	Hussain, M. M., Beg, M. M. S., Alam, M. S., & Laskar, S. H. (2020). Big data analytics platforms for electric vehicle integration in transport oriented smart cities: Computing platforms for platforms for electric vehicle integration in smart cities. In Information Resources Management Association (Ed.), Cyber warfare and terrorism: Concepts, methodologies, tools, and applications (pp. 833–854). IGI Global. www.igi-global.com/book/cyber-warfare-terrorism/236595
0,95,804,416	Parameswaran, T., Padmanabha Reddy, Y. C. A., Nagaveni, V., & Sathiyaraj, R. (2022). Era of computational big data analytics and IoT techniques in smart city applications. In S. Rajendran, M. Sabharwal, G. Ghinea, R. Kumar Dhanaraj, & B. Balusamy (Eds.), IoT and big data analytics for smart cities [1−22]. Chapman and Hall/CRC.
0,95,736,304	Akerkar, R. (2018). Processing big data for emergency management. In Z. Liu & K. Ota (Eds.), Advances in public policy and administration (APPA) book series. Smart technologies for emergency response and disaster management (pp. 144–166). Information Science Reference.
0,94,570,362	Martinez-Gil, J., Pichler, M., Turkanović, M., Beranič, T., Gradišnik, M., Lentini, G., Lué, A., Vitale, A. C., Doukhan, G., & Belet, C. (2020). Framework for Data Analysis in the Context of the Smart Villages. In A. Kő, E. Francesconi, G. Kotsis, A. M. Tjoa, & I. Khalil (Eds.), Lecture Notes in Computer Science (pp. 31–45). Springer International Publishing. https://doi.org/10.1007/978-3-030-58957-8_3
0,94,552,772	Rajendran, S., Sabharwal, M., Ghinea, G., Kumar Dhanaraj, R., & Balusamy, B. (2022). IoT and big data analytics for smart cities. Chapman and Hall/CRC.
0,94,504,906	Batty, M. (2020). Defining Smart Cities: High and Low Frequency Cities, Big Data and Urban Theory. In K. S. Willis & Routledge (Eds.), The Routledge companion to smart cities (pp. 51–60). Routledge.

Topic 6 – Privacy & Security.

Top-ranked references for topic VI

Gamma	Reference
0,97,167,648	Schulz, W., Wittner, F., Bavendiek, K., & Schupp, S. (2020). Modelling and Verification in GDPR's Data Protection Impact Assessment: A Case Study on the AccuWeather/Reveal Mobile Case. In R. Leenes, S. Gutwirth, & P. de Hert (Eds.), Computers, privacy and data protection: Volume 12. Data protection and privacy: data protection and democracy. Hart Publishing, an imprint of Bloomsbury Publishing.
0,97,120,989	Lokhorst, G., & van Eechhoud, M. (2020). Public Registers Caught between Open Government and Data Protection – Personal Data, Principles of Proportionality and the Public Interest. In R. Leenes, S. Gutwirth, & P. de Hert (Eds.), Computers, privacy and data protection: Volume 12. Data protection and privacy: data protection and democracy. Hart Publishing, an imprint of Bloomsbury Publishing.
0,97,099,762	Hallinan, D., Gutwirth, S., & Hert, P. de. (2020). Data protection and privacy: data protection and democracy. Computers, privacy and data protection. Hart Publishing, an imprint of Bloomsbury Publishing.
0,9,680,535	Alazab, M., Hong, S.-H., & Ng, J. (2021). Louder bark with no bite: Privacy protection through the regulation of mandatory data breach notification in Australia. Future Generation Computer Systems-the International Journal of Escience, 116(n/a), 22–29. https://doi.org/10.1016/j.future.2020.10.017
0,96,241,704	Karasek-Wojciechowicz, I. (2021). Reconciliation of anti-money laundering instruments and European data protection requirements in permissionless blockchain spaces. Journal of Cybersecurity, 7(1), [1–28] tyab004. https://doi.org/10.1093/cybsec/tyab004
0,95,725,961	Trenwith, P. M., & Venter, H. S. (2019). Freadypass: A digital forensic ready passport to control access to data across jurisdictional boundaries. Australian Journal of Forensic Sciences, 51(5), 583–595. https://doi.org/10.1080/00450618.2018.1444090
0,95,121,516	McLeod, J. (2020). Public trust in online records: The case of the UK care.data programme. In H. Stančić (Ed.), Routledge guides to practice in libraries, archives and information science. Trust and records in an open digital environment (First). Routledge Taylor & Francis Ltd.
0,95,032,036	He, Z. (2022). When data protection norms meet digital health technology: China's regulatory approaches to health data protection. Computer Law & Security Review, 47(N/A), [1–14] 105758. https://doi.org/10.1016/j.clsr.2022.105758
0,94,799,787	Lowry, J., & Sexton, A. (2020). Open Data and Privacy. In H. Stančić (Ed.), Routledge guides to practice in libraries, archives and information science. Trust and records in an open digital environment (First). Routledge Taylor & Francis Ltd.
0,9,442,368	Chertoff, M. (2018). Exploding data: reclaiming our cybersecurity in the digital age (First edition). Atlantic Monthly Press.

Topic 7 – Analytics.

Top-ranked references for topic VII

Gamma	Reference
0,99,092,641	Xuemiri, L., Jian, X., Qing, Z., Hongjun, L., Jeffrey Xu, Y., Xiaofang, Z., & Yidong, Y. (2006). Approximate Processing of Massive Continuous Quantile Queries over High-Speed Data Streams. IEEE Transactions on Knowledge & Data Engineering, 18(5), 683–698. http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=20744041&site=ehost-live
0,96,827,605	Yaqub, U., Sharma, N., Pabreja, R., Chun, S. A., Atluri, V., & Vaidya, J. (2020). Location-based Sentiment Analyses and Visualization of Twitter Election Data. Digital Government: Research and Practice, 1(2), 14 [1–19]. https://doi.org/10.1145/3339909
0,96,594,224	Yaqub, U., Sharma, N., Pabreja, R., Chun, S. A., Atluri, V., & Vaidya, J. (2018). Analysis and Visualization of Subjectivity and Polarity of Twitter Location Data. In C. C. Hinnant, A. Zuiderwijk, B. Lal, Y. K. Dwivedi, N. P. Rana, D. Frost, & S. Chirara (Eds.), Understanding 'development' from the perspective of *E*-government, digital divide and ICT4D literature: A research agenda [1−10]. Association for Computing Machinery. https://doi.org/10.1145/3209281.3209313
0,96,249,043	Wena, C.-H., Hsu, P.-Y., Wang, C.-Y., Wuc, T.-L., & Hsu, M.-J. (2012). E-government Information Application: Identifying Smuggling Vessels with Data mining Technology. Electronic Journal of Electronic Government, 10(1), 47–58.
0,95,322,973	Li, X., Guo, X., Kim, S. K., & Lee, H. (2022). Analysis of social media utilization based on big data-focusing on the Chinese government Weibo. KSII Transactions on Internet and Information Systems, 16(8), 2571–2586. https://doi.org/10.3837/tiis.2022.08.006
0,94,994,524	Clark, S. D., Morris, M. A., & Lomax, N. (2018). Estimating the outcome of UKs referendum on EU membership using e-petition data and machine learning algorithms. Journal of Information Technology & Politics, 15(4), 344–357. https://doi.org/10.1080/19331681.2018.1491926
0,94,592,852	Lo, W. H., Yan Lam, B. S., & Fung Cheung, M. M. (2021). The dynamics of political elections: A big data analysis of intermedia framing between social media and news media. Social Science Computer Review, 39(4), 627–647. https://doi.org/10.1177/0894439319876593
0,93,797,759	Rao, K. S., & Chandran, K. R. (2013). Mining of customer walking path sequence from RFID supermarket data. Electronic Government: An International Journal, 10 (1), 34–55. http://www.inderscience.com/info/inarticle.php?artid=51278
0,93,430,226	Banerjee, S., Swearingen, T., Shillair, R., Bauer, J. M., Holt, T., & Ross, A. (2022). Using machine learning to examine cyberattack motivations on web defacement data. Social Science Computer Review, 40(4), 914–932. https://doi.org/10.1177/0894439321994234
0,92,411,395	Abdullayeva, F. J. (2022). Distributed denial of service attack detection in e-government cloud via data clustering. Array, 15(N/A), [1−12]. https://doi.org/10.1016/j.array.2022.100229

Topic 8 – Geospatial Data.

Top-ranked references for topic VIII

Gamma	Reference
0,98,795,417	Chomkhamsri, K., Mungcharoen, T., & Yuvaniyama, C. (2017). 10-year experience with the Thai national LCI database: Case study of "refinery products". International Journal of Life Cycle Assessment, 22(11), 1760–1770. https://doi.org/10.1007/s11367-016-1160-3
0,97,771,375	Choudhury, P. R., & Behera, M. K. (2017). Using Administrative Data for Monitoring and Improving Land Policy and Governance in India. In (pp. 127–135). ACM. https://doi.org/10.1145/3047273.3047296
0,96,816,599	Assis-Rodrigues, F. de, Cassiano-Lopes, R. d. C., Grossi de Carvalho, Angela Maria, & Goncalves-Santana, R. C. (2016). Informational behavior and public information policies: theoretical considerations arising from the case of DATASUS in Brazil. Investigacion Bibliotecologica, 30(69 Supplement), 17–39.
0,96,128,825	Esperanca, C., & Samet, H. (2000). Experience with SAND-Tcl: A Scripting Tool for Spatial Databases. In (pp. 1–24). Digital Government Research Center. http://www.digitalgovernment.org/library/library/dgo2000/
0,94,909,501	Lo Duca, A., & Marchetti, A. (2019). Open data for tourism: The case of Tourpedia. Journal of Hospitality and Tourism Technology, 10(3), 382–398. https://doi.org/10.1108/JHTT-07-2017-0042

(*continued on next page*)

(*continued*)

Gamma	Reference
0,94,744,535	McGee, S., & Graves, R. K. (2008). The International Heartbeat Beacon for interoperability & synchronicity of event, alert data. In F. Fiedrich & B. de van Walle (Eds.) (pp. 128–133). Information Systems for Crisis Response and Management, ISCRAM.
0,93,094,698	Aydinoglu, A. C., & Bovkir, R. (2017). Generic land registry and cadastre data model supporting interoperability based on international standards for Turkey. Land Use Policy, 68, 59–71. https://doi.org/10.1016/j.landusepol.2017.07.029
0,92,898,684	Chua, H. H. (2003). TranStats: The Intermodal Transportation Database. Reference Reviews, 17(7), 27. http://proquest.umi.com/pqdweb?did=949893901&Fmt=7&clientId=8991&RQT=309&VName=PQD
0,91,419,929	Huffman, T., Olesen, M., Green, M., Leckie, D., Liu, J., & Shang, J. (2018). Agricultural data analytics for environmental monitoring in Canada. In F. A. Batarseh & R. Yang (Eds.), Federal data science: transforming government and agricultural policy using artificial intelligence (pp. 55–79). Academic Press is an imprint of Elsevier.
0,90,961,558	Abbas, S., & Ojo, A. (2014). Applying Design Patterns in URI Strategies - Naming in Linked Geospatial Data Infrastructure. In [1–10]. IEEE.

Appendix B. Appendix II

A topic model with 8 topics, 1441 documents and a 6989 word dictionary.
Topic 1 Top Words:
Highest Prob: citizen, innov, polici, particip, servic, ecosystem, transpar.
FREX: corrupt, ecosystem, innov, particip, busi, econom, privat.
Lift: dubai, khda, ts, ifr, oda, explic, odbm.
Score: corrupt, innov, ecosystem, transpar, democrat, polit, privat.
Topic 2 Top Words:
Highest Prob: portal, dataset, evalu, publish, base, design, metadata.
FREX: metadata, schema, semant, portal, ontolog, format, standard.
Lift: berner, lee, odg, misconcept, schema, codelist, middlewar.
Score: portal, metadata, semant, schema, interoper, signatur, ontolog.
Topic 3 Top Words:
Highest Prob: analyt, technolog, organ, decis, polici, process, local.
FREX: analyt, capabl, organiz, od, decis, organ, local.
Lift: binomi, saharan, ssa, underperform, verg, viability.th, casual.
Score: analyt, organiz, od, 311, capabl, lsdi, da.
Topic 4 Top Words:
Highest Prob: ogd, countri, factor, initi, user, qualiti, analysi.
FREX: ogd, factor, perceiv, influenc, countri, licens, barrier.
Lift: asian, gogd, donne, l'ouvertur, ouvert, km, a.
Score: ogd, portal, countri, des, perceiv, pakistan, les.
Topic 5 Top Words:
Highest Prob: citi, smart, manag, urban, technolog, system, servic.
FREX: smart, urban, citi, disast, iot, energi, grid.
Lift: ie, grid, foremost, imperfect, hurrican, mdm, height.
Score: smart, citi, urban, disast, iot, energi, fire.
Topic 6 Top Words:
Highest Prob: privaci, protect, secur, person, digit, regul, law.
FREX: protect, gdpr, breach, encrypt, privaci, surveil, regul.
Lift: blood, infect, biometr, dataveil, illeg, ipf, pii.
Score: protect, privaci, gdpr, blockchain, person, secur, regul.
Topic 7 Top Words:
Highest Prob: social, analysi, media, method, user, base, result.
FREX: cluster, twitter, bold, sentiment, mine, media, candid.
Lift: dbscan, pca, bolld, siot, lightweight, armw, clustw.
Score: mine, twitter, media, textless, cluster, cloud, queri.
Topic 8 Top Words:
Highest Prob: databas, spatial, link, statist, nation, infrastructur, project.
FREX: sdi, lci, marin, land, thai, losd, spatial.
Lift: underspecifi, commenc, columbia, lkb, sensus, sim, argo.
Score: lci, spatial, thai, land, databas, tourpedia, sand.

Highlights•

The paper reviews data collaboration in the Digital Government Reference Library and introduces a data collaboration framework.•

The computational literature review uses Structural Topic Modelling.•

Digital government research focuses on innovation and participation.•

There is a decline in research discussions on standardisation, privacy, security, and trust.•

Literature emphasises open data, leading to the neglect of critical restricted data.

References

Abbas, S., & Ojo, A. (2014). *Applying design patterns in URI strategies - Naming in linked geospatial data infrastructure*. IEEE.

Abdullayeva, Fargana J. (2022). Distributed denial of service attack detection in E-government cloud via data clustering. *Array, 15*, 100229. https://doi.org/10.1016/j.array.2022.100229

Agbozo, E., Alhassan, D., & Spassov, K. (2018). Personal data and privacy barriers to E-government adoption, implementation and development in sub-Saharan Africa. In A. Chugunov, Y. Misnikov, E. Roshchin, & D. Trutnev (Eds.), *Communications in Computer and Information Science* (pp. 82–91). Springer International Publishing. https://doi.org/10.1007/978-3-030-13283-5.

AISBL. (2024). Gaia-X Committees & Working Groups. https://gaia-x.eu/wp-content/uploads/2024/08/Gaia-X-Commities-and-WG-Structure-2024.pdf.

Akerkar, R. (2018). Processing big data for emergency management. In Z. Liu, & K. Ota (Eds.), *Advances in public policy and administration (APPA) book series. Smart technologies for emergency response and disaster management* (pp. 144–166). Information Science Reference.

Alazab, M., Hong, S.-H., & Ng, J. (2021). Louder bark with no bite: Privacy protection through the regulation of mandatory data breach notification in Australia. *Future Generation Computer Systems-the International Journal of Escience, 116*, 22–29. https://doi.org/10.1016/j.future.2020.10.017

Alhawawsha, M., & Panchenko, T. (2020). Open data platform architecture and its advantages for an open e-government. In Z. Hu, S. Petoukhov, I. Dychka, & M. He (Eds.), *Advances in intelligent systems and computing* (pp. 631–639). Springer International Publishing. https://doi.org/10.1007/978-3-030-55506-1.

Allard Scott, W., Wiegand Emily, R., Schlecht, C., Datta, A. R., Goerge Robert, M., & Weigensberg, E. (2018). State agencies' use of administrative data for improved practice: Needs, challenges, and opportunities. *Public Administration Review, 78*(2), 240–250. https://doi.org/10.1111/puar.12883

Aranda, A. M., Sele, K., Etchanchu, H., Guyt, J. Y., & Vaara, E. (2021). From big data to rich theory: Integrating critical discourse analysis with structural topic modeling. *European Management Review, 18*(3), 197–214. https://doi.org/10.1111/emre.12452

de Assis-Rodrigues, F., Cassiano-Lopes, R. D. C., de Carvalho, G., Maria, A., & Goncalves-Santana, R. C. (2016). Informational behavior and public information policies: Theoretical considerations arising from the case of DATASUS in Brazil. *Investigacion Bibliotecologica, 30*(69 Supplement), 17–39.

Aydinoglu, A. C., & Bovkir, R. (2017). Generic land registry and cadastre data model supporting interoperability based on international standards for Turkey. *Land Use Policy, 68*, 59–71. https://doi.org/10.1016/j.landusepol.2017.07.029

Bagheri, S., Brandt, T., Sheombar, H., & van Oosterhout, M. (2021). *Value creation through urban data platforms: A conceptual framework* (pp. 2464–2473). University of Hawai'i at Manoa. https://doi.org/10.24251/HICSS.2021.302

Banerjee, S., Swearingen, T., Shillair, R., Bauer, J. M., Holt, T., & A.Ross. (2021). Using machine learning to examine cyberattack motivations on web defacement data. *social science computer review, 40*(4), 914–932. https://doi.org/10.1177/0894439321994234

Bargh, M. S., Choenni, S., & Meijer, R. (2016). On design and deployment of two privacy-preserving procedures for judicial-data dissemination. *Government Information Quarterly, 33*(3), 481–493. https://doi.org/10.1016/j.giq.2016.06.002

Barthelemy, F., Cochez, M., Dimitriadis, I., Karim, N., Loutas, N., Magnisalis, I., … Wyns, B. (2022). Towards a standard-based open data ecosystem: Analysis of DCAT-AP use at national and European level. *Electronic Government, an International Journal, 18*(2), 137–180. https://doi.org/10.1504/EG.2022.121856

Batty, M. (2020). Defining smart cities: High and low frequency cities, big data and urban theory. In K. S. Willis, & Routledge (Eds.), *The Routledge companion to smart cities* (pp. 51–60). Routledge.

Bihl, T., Gutierrez, R., Bauer, K., Boehmke, B., & Saie, C. (2020). *Topological data analysis for enhancing embedded analytics for enterprise cyber log analysis and forensics* (pp. 1937–1946). University of Hawai'i at Manoa. https://doi.org/10.24251/HICSS.2020.238

Bonina, C., & Eaton, B. (2020). Cultivating open government data platform ecosystems through governance: Lessons from Buenos Aires, Mexico City and Montevideo. *Government Information Quarterly, 37*(3), Article 101479 [1-15] https://doi.org/10.1016/j.giq.2020.101479.

Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research, 6*(1), 97–113. https://doi.org/10.1177/1468794106058877

Budnitsky, S. (2022). A relational approach to digital sovereignty: E-Estonia between Russia and the West. *International Journal of Communication, 16*, 1–22.

Buhin Pandur, M., Dobša, J., & Kronegger, L. (2020). *Topic modelling in social sciences: Case study of web of science* (file:///C:/Users/sju/Downloads/CECIIS-2020_Topic_modelling_Buhin_Dobsa_Kronegger.pdf).

Bulazel, A., DiFranzo, D., Erickson, J. S., & Hendler, J. A. (2016). The importance of authoritative URI design schemes for open government data. *International Journal of Public Administration in the Digital Age, 3*(2), 1–18. https://doi.org/10.4018/IJPADA.2016040101

Calzada, I., & Almirall, E. (2020). Data ecosystems for protecting European citizens' digital rights. *Transforming Government- People Process and Policy, 14*(2), 133–147. https://doi.org/10.1108/TG-03-2020-0047

Campana, K., Mills, J. E., & Martin, M. H. (2019). The role of community data in helping public libraries reach and serve underserved communities. In *Lecture, notes in computer science* (pp. 247–253). Springer International Publishing.

Chen, X., Xie, H., Li, Z., Cheng, G., Leng, M., & Wang, F. L. (2023). Information fusion and artificial intelligence for smart healthcare: A bibliometric study. *Information Processing & Management, 60*(1), Article 103113. https://doi.org/10.1016/j.ipm.2022.103113

Chertoff, M. (2018). *Exploding data: Reclaiming our cybersecurity in the digital age* (1st ed.). Atlantic Monthly Press.

Choi, Y., Gil-Garcia, J., Burke, G. B., Costello, J., Werthmuller, D., & Aranay, O. (2021). *Towards data-driven decision-making in government: Identifying opportunities and challenges for data use and analytics* (pp. 2183–2192). University of Hawai'i at Manoa. https://doi.org/10.24251/HICSS.2021.268

Chomkhamsri, K., Mungcharoen, T., & Yuvaniyama, C. (2017). 10-year experience with the Thai national LCI database: Case study of "refinery products". *International Journal of Life Cycle Assessment, 22*(11), 1760–1770. https://doi.org/10.1007/s11367-016-1160-3

Choudhury, P. R., & Behera, M. K. (2017). *Using administrative data for monitoring and improving land policy and governance in India* (pp. 127–135). ACM. https://doi.org/10.1145/3047273.3047296

Chua, H. H. (2003). TranStats: The intermodal transportation database. *Reference Reviews, 17*(7), 27. http://proquest.umi.com/pqdweb?did=949893901&Fmt=7&clientId=8991&RQT=309&VName=PQD.

Clark, S. D., Morris, M. A., & Lomax, N. (2018). Estimating the outcome of UKs referendum on EU membership using e-petition data and machine learning algorithms. *Journal of Information Technology & Politics, 15*(4), 344–357. https://doi.org/10.1080/19331681.2018.1491926

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly, 35*, 128–152.

Cordasco, G., de Donato, R., Malandrino, D., Palmieri, G., Petta, A., Pirozzi, D., … Vicidomini, L. (2017). *Engaging citizens with a social platform for open data* (pp. 242–249). ACM. https://doi.org/10.1145/3085228.3085302

Couture, S., & Toupin, S. (2019). What does the notion of "sovereignty" mean when referring to the digital? *New Media & Society, 21*(10), 2305–2322. https://doi.org/10.1177/1461444819865984

Dawes, S. S., Gharawi, M. A., & Burke, B. (2012). Transnational public sector knowledge networks: Knowledge and information sharing in a multi-dimensional context. *Government Information Quarterly, 29*(1), S112–S120. https://doi.org/10.1016/j.giq.2011.08.002

Della Porta, D., & Keating, M. (Eds.). (2008). *Approaches and methodologies in the social sciences: A pluralist perspective*. Cambridge University Press.

Du, J., Kuang, B., & Yang, Y. (2019). A data-driven framework for smart urban domestic wastewater: A sustainability perspective. *Advances in Civil Engineering, 2019*. https://doi.org/10.1155/2019/6530626 [1-16].

Esperanca, C., & Samet, H. (2000). *Experience with SAND-Tcl: A scripting tool for spatial databases* (pp. 1–24). Digital Government Research Center. http://www.digitalgovernment.org/library/library/dgo2000/.

European Commission. (2020). A European strategy for data: EUR-Lex - 52020DC0066. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066.

European Commission. (2022). Staff working document on data spaces. https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces.

Finfgeld-Connett, D. (2014). Use of content analysis to conduct knowledge-building and theory-generating qualitative systematic reviews. *Qualitative Research, 14*(3), 341–352. https://doi.org/10.1177/1468794113481790

Fitriani, W. R., Hidayanto, A. N., Sandhyaduhita, P., Purwandari, B., & Kosandi, M. (2019). Determinants of continuance intention to use open data website: An insight from Indonesia. *Pacific Asia Journal of the Association for Information Systems, 11*(2), 96–120. https://doi.org/10.17705/1pais.11205

Franklin, M., Halevy, A., & Maier, D. (2005). From databases to dataspaces. *ACM SIGMOD Record, 34*(4), 27–33. https://doi.org/10.1145/1107499.1107502

Gelhaar, J., Groß, T., & Otto, B. (2017). A taxonomy for data ecosystems. In *, 2017. Proceedings of the 50th Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2021.739.

Gheisari, M., & Esnaashari, M. (2018). Data storages in wireless sensor networks to deal with disaster management. In Z. Liu, & K. Ota (Eds.), *Advances in public policy and administration (APPA) book seriesSmart technologies for emergency response and disaster management* (pp. 196–222). Information Science Reference.

Gil-García, J. R., & Sayogo, D. S. (2016). Government inter-organizational information sharing initiatives: Understanding the main determinants of success. *Government Information Quarterly, 33*(3), 572–582. https://doi.org/10.1016/j.giq.2016.01.006

Goedegebuure, A., Kumara, I., Driessen, S., van den Heuvel, W.-J., Monsieur, G., Tamburri, D. A., & Di Nucci, D. (2025). Data mesh: A systematic gray literature review. *ACM Computing Surveys, 57*(1), 1–36. https://doi.org/10.1145/3687301

Gong, Y., & Janssen, M. (2021). Roles and capabilities of enterprise architecture in big data analytics technology adoption and implementation. *Journal of Theoretical and Applied Electronic Commerce Research, 16*(1), 37–51. https://doi.org/10.4067/S0718-18762021000100104

Graham, F. S., Gooden, S. T., & Martin, K. J. (2016). Navigating the transparency-privacy paradox in public sector data sharing. *The American Review of Public Administration, 46*(5), 569–591. https://doi.org/10.1177/0275074014561116

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis, 21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Guenduez, A. A., & Mettler, T. (2022). Strategically constructed narratives on artificial intelligence: What stories are told in governmental artificial intelligence policies? *Government Information Quarterly*. , Article 101719. https://doi.org/10.1016/j.giq.2022.101719

Habibzadeh, H., Nussbaum, B. H., Anjomshoa, F., Kantarci, B., & Soyata, T. (2019). A survey on cybersecurity, data privacy, and policy issues in cyber-physical system deployments in smart cities. *Sustainable Cities and Society, 50*, 1–20. https://doi.org/10.1016/j.scs.2019.101660

Halevy, A., Franklin, M., & Maier, D. (2006). Principles of dataspace systems. In G. Gottlob, & J. van den Bussche (Eds.), *Proceedings of the twenty-fifth ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems* (pp. 1–9). ACM. https://doi.org/10.1145/1142351.1142352.

Hallinan, D., Gutwirth, S., & de Hert, P. (2020). *Data protection and privacy: Data protection and democracy. Computers, privacy and data protection*. Hart Publishing, an imprint of Bloomsbury Publishing. doi:Volume.

Hawig, D., Zhou, C., Fuhrhop, S., Fialho, A. S., & Ramachandran, N. (2019). Designing a distributed ledger technology system for interoperable and general data protection regulation-compliant health data exchange: A use case in blood glucose data. *JMIR Public Health and Surveillance, 21*(6), 1–13. https://doi.org/10.2196/13665

He, Z. (2022). When data protection norms meet digital health technology: China's regulatory approaches to health data protection. *Computer Law and Security Review, 47*(N/A), 1–14, 105758 https://doi.org/10.1016/j.clsr.2022.105758.

Huber, F., Wainwright, T., & Rentocchini, F. (2020). Open data for open innovation: Managing absorptive capacity in SMEs. *R and D Management, 50*(1), 31–46. https://doi.org/10.1111/radm.12347

Huffman, T., Olesen, M., Green, M., Leckie, D., Liu, J., & Shang, J. (2018). Agricultural data analytics for environmental monitoring in Canada. In F. A. Batarseh, & R. Yang (Eds.), *Federal data science: Transforming government and agricultural policy using artificial intelligence* (pp. 55–79). Academic Press is an imprint of Elsevier.

Hummel, P., Braun, M., Tretter, M., & Dabrock, P. (2021). Data sovereignty: A review. *Big Data & Society, 8*(1). https://doi.org/10.1177/2053951720982012, 205395172098201.

Hussain, M. M., Beg, M. M. S., Alam, M. S., & Laskar, S. H. (2020). Big data analytics platforms for electric vehicle integration in transport oriented smart cities: Computing platforms for platforms for electric vehicle integration in smart cities. In Information Resources Management Association (Ed.), *Cyber warfare and terrorism: Concepts, methodologies, tools, and applications* (pp. 833–854). IGI Global. www.igi-global.com/book/cyber-warfare-terrorism/236595.

Iftikhar Hussain Shah, S., Peristeras, V., & Magnisalis, I. (2021). Government big data ecosystem: Definitions, types of data, actors, and roles and the impact in public administrations. *ACM Journal of Data and Information Quality, 13*(2), 1–25, 8 https://doi.org/10.1145/3425709.

Karasek-Wojciechowicz, I. (2021). Reconciliation of anti-money laundering instruments and European data protection requirements in permissionless blockchain spaces. *Journal of Cybersecurity, 7*(1), 1–28. tyab004 https://doi.org/10.1093/cybsec/tyab004.

Kassen, M. (2019). Open data and e-government - related or competing ecosystems: A paradox of open government and promise of civic engagement in Estonia. *Information Technology for Development, 25*(3), 552–578. https://doi.org/10.1080/02681102.2017.1412289

Kassen, M. (2020). Open data and its peers: Understanding promising harbingers from Nordic Europe. *Aslib Journal of Information Management, 72*(5), 765–785. https://doi.org/10.1108/AJIM-12-2019-0364

Khurshid, M. M., Zakaria, N. H., Arfeen, M. I., Rashid, A., Nasir, S. U., & Shehzad, H. M. F. (2022). Factors influencing citizens' intention to use open government data—A case study of Pakistan. *Big Data and Cognitive Computing, 6*(1), 1–22. https://doi.org/10.3390/bdcc6010031

Kirstein, F., Dittwald, B., Dutkowski, S., Glikman, Y., Schimmler, S., & Hauswirth, M. (2019a). Linked data in the European data portal: A comprehensive platform for applying DCAT-AP. In I. Lindgren, M. Janssen, H. Lee, A. Polini, M. P. Rodríguez Bolívar, H. J. Scholl, & E. Tambouris (Eds.)*, 11685. Lecture notes in computer science* (pp. 192–204). Springer International Publishing. https://doi.org/10.1007/978-3-030-27325-5_15.

Kirstein, F., Dittwald, B., Dutkowski, S., Glikman, Y., Schimmler, S., & Hauswirth, M. (2019b). Linked data in the European data portal: A comprehensive platform for applying DCAT-AP. In I. Lindgren, M. Janssen, H. Lee, A. Polini, M. P. Rodríguez Bolívar, H. J. Scholl, & E. Tambouris (Eds.), Lecture notes in computer science (pp. 192–204). Springer International Publishing. doi:https://doi.org/10.1007/978-3-030-27325-5\textunderscore.

Krishnana, S., & Lymmb, J. (2016). Determinants of virtual social networks diffusion: Insights from cross-country data. *Computers in Human Behavior, 54*, 691–700. https://doi.org/10.1016/j.chb.2015.07.055

Li, X., Guo, X., Kim, S. K., & Lee, H. (2022). Analysis of social media utilization based on big data-focusing on the Chinese government Weibo. *KSII Transactions on Internet and Information Systems, 16*(8), 2571–2586. https://doi.org/10.3837/tiis.2022.08.006

Linåker, J., & Runeson, P. (2021). How to enable collaboration in open government data ecosystems: A public platform provider's perspective. *JeDEM - EJournal of EDemocracy and Open Government, 13*(1), 1–30. https://doi.org/10.29379/jedem.v13i1.634

Lo Duca, A., & Marchetti, A. (2019). Open data for tourism: The case of Tourpedia. *Journal of Hospitality and Tourism Technology, 10*(3), 382–398. https://doi.org/10.1108/JHTT-07-2017-0042

Lo, O., Buchanan, W. J., Sayeed, S., Papadopoulos, P., Pitropakis, N., & Chrysoulas, C. (2022). GLASS: A citizen-centric distributed data-sharing model within an e-governance architecture. *Sensors (Basel), 22*(6), Article 2291. https://doi.org/10.3390/s22062291

Lo, W. H., Yan Lam, B. S., & Fung Cheung, M. M. (2021). The dynamics of political elections: A big data analysis of intermedia framing between social media and news media. *Social Science Computer Review, 39*(4), 627–647. https://doi.org/10.1177/0894439319876593

Lokhorst, G., & van Eechhoud, M. (2020). Public registers caught between open government and data protection – Personal data, principles of proportionality and the public interest. In R. Leenes, S. Gutwirth, & P. de Hert (Eds.), *Computers, privacy and data protection: Volume 12. Data protection and privacy: data protection and democracy*. Hart Publishing, an imprint of Bloomsbury Publishing.

Lowry, J., & Sexton, A. (2020). Open data and privacy. In H. Stančić (Ed.), *Routledge guides to practice in libraries, archives and information science. Trust and records in an open digital environment (First)*. Routledge Taylor & Francis Ltd.

Machado, I. A., Costa, C., & Santos, M. Y. (2022). Data mesh: Concepts and principles of a paradigm shift in data architectures. *Procedia Computer Science, 196*, 263–271. https://doi.org/10.1016/j.procs.2021.12.013

Maerz, S. F., & Puschmann, C. (2020). Text as data for conflict research: A literature survey. In E. Deutschmann, J. Lorenz, L. G. Nardin, D. Natalini, & A. F. X. Wilhelm (Eds.), *Computational social sciences. Computational conflict research* (pp. 43–65). Springer International Publishing. https://doi.org/10.1007/978-3-030-29333-8_3.

Marczak, P., & Sieber, R. (2018). Linking legislative openness to open data in Canada. *The Canadian Geographer, 62*(2), 106–119. https://doi.org/10.1111/cag.12408

Martinez-Gil, J., Pichler, M., Turkanović, M., Beranič, T., Gradišnik, M., Lentini, G., … Belet, C. (2020). Framework for data analysis in the context of the smart villages. In A. Kő, E. Francesconi, G. Kotsis, A. M. Tjoa, & I. Khalil (Eds.), *Lecture notes in computer science* (pp. 31–45). Springer International Publishing. https://doi.org/10.1007/978-3-030-58957-8_3.

McGee, S., & Graves, R. K. (2008). *The international heartbeat Beacon for interoperability & synchronicity of event, alert data*. In F. Fiedrich & B. de van Walle (Eds.) (pp. 128–133). Information Systems for Crisis Response and Management, ISCRAM.

McLeod, J. (2020). Public trust in online records: The case of the UK care. Data programme. In H. Stančić (Ed.), *Routledge guides to practice in libraries, archives and information science. Trust and records in an open digital environment (first)*. Routledge Taylor & Francis Ltd.

Missier, P., Lalk, G., Verykios, V., Grillo, F., Lorusso, T., & Angeletti, P. (2003). Improving data quality in practice: A case study in the Italian public administration. *Distributed and Parallel Databases, 13*(2), 135–160, 000179704900001.

Musyaffa, F. A., Orlandi, F., Jabeen, H., & Vidal, M. E. (2018). Classifying data heterogeneity within budget and spending open data. In *ICEGOV '18* (pp. 81–90). ACM. https://doi.org/10.1145/3209415.3209482.

Mutambik, I., Almuqrin, A., Lee, J., [John], Gauthier, J., & Homadi, A. (2022). Open government data in gulf cooperation council countries: An analysis of progress. *Sustainability, 14*(12), Article 7200. https://doi.org/10.3390/su14127200

Oliveira, S., Barros Lima, M. I., Farias Lóscio, G.d. F., & Farias Lóscio, B. (2019). Investigations into data ecosystems: A systematic mapping study. *Knowledge and Information Systems, 61*(2), 589–630. https://doi.org/10.1007/s10115-018-1323-6

Otto, B. (2022). The evolution of data spaces. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing data spaces*. Springer International Publishing.

Parameswaran, T., Padmanabha Reddy, Y. C. A., Nagaveni, V., & Sathiyaraj, R. (2022). Era of computational big data analytics and IoT techniques in smart city applications. In S. Rajendran, M. Sabharwal, G. Ghinea, R. Kumar Dhanaraj, & B. Balusamy (Eds.), *IoT and big data analytics for smart cities* (pp. 1–22). Chapman and Hall/CRC.

Purwanto, A., Zuiderwijk, A., & Janssen, M. (2020). Citizens' Trust in open government data: A quantitative study about the effects of data quality, system quality and service quality. In S.-J. Eom, & J. Lee (Eds.), *dg.o '20* (pp. 310–318). Association for Computing Machinery. https://doi.org/10.1145/3396956.3396958.

Rajendran, S., Sabharwal, M., Ghinea, G., Kumar Dhanaraj, R., & Balusamy, B. (2022). *IoT and big data analytics for smart cities*. Chapman and Hall/CRC.

Rao, K. S., & Chandran, K. R. (2013). Mining of customer walking path sequence from RFID supermarket data. *Electronic Government, An International Journal, 10*(1), 34–55. http://www.inderscience.com/info/inarticle.php?artid=51278.

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association, 111*(515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm : An R package for structural topic models. *Journal of Statistical Software, 91*(2). https://doi.org/10.18637/jss.v091.i02

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., … Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science, 58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Ruijer, E. (2021). Designing and implementing data collaboratives: A governance perspective. *Government Information Quarterly, 38*(4), Article 101612. https://doi.org/10.1016/j.giq.2021.101612

Rukanova, B., Tan, Y.-H., Slegt, M., Molenhuis, M., van Rijnsoever, B., Migeotte, J., … Post, S. (2021). Identifying the value of data analytics in the context of government supervision: Insights from the customs domain. *Government Information Quarterly, 38*(1), 1–20, 101496 https://doi.org/10.1016/j.giq.2020.101496.

Saxena, S. (2018). National open data frames across Japan, the Netherlands and Saudi Arabia: Role of culture. *Foresight, 20*(1), 123–134. https://doi.org/10.1108/FS-07-2017-0038

Sayogo, D. S., & Pardo, T. A. (2011). Understanding the capabilities and critical success factors in collaborative data sharing network: The case of data ONE. In *dg.O '11* (pp. 74–83). ACM. https://doi.org/10.1145/2037556.2037568.

Scholl, H. J. (2021). The digital government reference library (DGRL) and its potential formative impact on digital government research (DGR). *Government Information Quarterly, 38*(4), Article 101613. https://doi.org/10.1016/j.giq.2021.101613

Scholl, H. J. (2022). The Digital Government Reference Library (DGRL): Version 18.5. http://faculty.washington.edu/jscholl/dgrl/.

Schulz, W., Wittner, F., Bavendiek, K., & Schupp, S. (2020). Modelling and verification in GDPR'S data protection impact assessment: A case study on the AccuWeather/reveal Mobile case. In R. Leenes, S. Gutwirth, & P. de Hert (Eds.), *Computers, privacy and data protection: Volume 12. Data protection and privacy: data protection and democracy*. Hart Publishing, an imprint of Bloomsbury Publishing.

Schwartz-Shea, P., & Yanow, D. (2012). *Interpretive research design: Concepts and processes. Routledge series on interpretive methods*. Routledge. https://doi.org/10.4324/9780203854907

Seo, J., Kim, B., & Yeong Kwon, H. (2021). In J. Lee, G. Vilale Pereira, & S. Hwang (Eds.), *Open data policies analysis disputes mediation cases in Korea: Based on OUR data index and ODB* (pp. 153–167). Association for Computing Machinery. https://doi.org/10.1145/3463677.3463735.

Sharma, A., Rana, N. P., & Nunkoo, R. (2021). Fifty years of information management research: A conceptual structure analysis using structural topic modeling. *International Journal of Information Management, 58*, Article 102316. https://doi.org/10.1016/j.ijinfomgt.2021.102316

Slobodova, O. (2020). Motivations of civil society actors to reuse open data. Case study of discrete actors in Lyon. *Bitácora Urbano Territorial, 30*(3), 55–70. https://doi.org/10.15446/bitacora.v30n3.80740

Soundarabai, P. B., Raj, P., & Mahmood, Z. (2021). Citizen data in distributed computing environments: Privacy and protection mechanisms. In Z. Mahmood (Ed.), *Web 2.0 and cloud technologies for implementing connected government* (pp. 183–203). IGI Global.

Stebbins, R. A. (2001). *Exploratory research in the social sciences. Qualitative research methods. 48*. SAGE.

Stranacher, K., Krnjic, V., Zwattendorfer, B., Zefferer, T., Castelnovo, W., & Ferrari, E. (2013). *Assessment of Redactable signature schemes for trusted and reliable public sector data* (pp. 508–516). Academic Conferences Limited.

Sudhakar Yadav, N., Gogula, S., Krishna Sharma, G., Mallikarjuna Rao, C., & Lalita Parameswari, D. V. (2022). IoT and big data analytics-based intelligent decision-making systems. In S. Rajendran, M. Sabharwal, G. Ghinea, R. Kumar Dhanaraj, & B. Balusamy (Eds.), *IoT and big data analytics for smart cities* (pp. 1–19). Chapman and Hall/CRC.

Susha, I., & Gil-Garcia, J. R. (2019). A collaborative governance approach to partnerships addressing public problems with private data. In *Proceedings of the 52nd Hawaii International Conference on system sciences (HICSS-52), Wailea, Hawaii, US*. http://scholarspace.manoa.hawaii.edu/handle/10125/59726https://scholarspace.manoa.hawaii.edu/handle/10125/59726.

Susha, I., Janssen, M., & Verhulst, S. (2017). Data Collaboratives as a new frontier of cross-sector partnerships in the age of open data: Taxonomy development. In *Proceedings of the 50th Hawaii International Conference on System Sciences (2017)*. Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2017.325.

Susha, I., Pardo, T. A., Janssen, M., Adler, N., Verhulst, S. G., & Harbour, T. (2018). A research roadmap to advance data collaboratives practice as a novel research direction. *International Journal of Electronic Government Research (IJEGR), 14*(3), 1–11. https://doi.org/10.4018/IJEGR.2018070101

Susha, I., Rukanova, B. D., Gil-García, J. R., Hua Tan, Y., & Gasco, M. (2019). Identifying mechanisms for achieving voluntary data sharing in cross-sector partnerships for public good. In Y.-C. Chen, F. Salem, & A. Zuiderwijk (Eds.), *dg.o 2019* (pp. 227–236). ACM. https://doi.org/10.1145/3325112.3325265.

Susha, I., van den Broek, T., van Veenstra, A.-F., & Linåker, J. (2022). An ecosystem perspective on developing data collaboratives for addressing societal issues: The role of conveners. *Government Information Quarterly*. https://doi.org/10.1016/j.giq.2022.101763, 101763.

Takahashi, A. R. W., & Araujo, L. (2019). Case study research: Opening up research opportunities. *RAUSP Management Journal, 55*(1), 100–111. https://doi.org/10.1108/RAUSP-05-2019-0109

Trenwith, P. M., & Venter, H. S. (2019). Freadypass: A digital forensic ready passport to control access to data across jurisdictional boundaries. *Australian Journal of Forensic Sciences, 51*(5), 583–595. https://doi.org/10.1080/00450618.2018.1444090

Verhulst, S. G., & Sangokoya, D. (2015). Data collaboratives: Exchanging data to improve people's lives. Medium. https://sverhulst.medium.com/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a.

Wanckel, C. (2022). An ounce of prevention is worth a pound of cure – Building capacities for the use of big data algorithm systems (BDAS) in early crisis detection. *Government Information Quarterly, 39*(4), 1–13, 101705 https://doi.org/10.1016/j.giq.2022.101705.

Wareham, J. D., Fox, P. B., & Cano Giner, J. L. (2013). Technology ecosystem governance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2201688. Advance online publication.

Wena, C.-H., Hsu, P.-Y., Wang, C.-Y., Wuc, T.-L., & Hsu, M.-J. (2012). E-government information application: Identifying smuggling vessels with data mining technology. *Electronic Journal of Electronic Government, 10*(1), 47–58.

Xuemiri, L., Jian, X., Qing, Z., Hongjun, L., Jeffrey, X. Y., Xiaofang, Z., & Yidong, Y. (2006). Approximate processing of massive continuous quantile queries over high-speed data streams. *IEEE Transactions on Knowledge and Data Engineering, 18*(5), 683–698. http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=20744041&site=ehost-live.

Yang, T.-M., & Maxwell, T. A. (2011). Information-sharing in public organizations: A literature review of interpersonal, intra-organizational and inter-organizational success factors. *Government Information Quarterly, 28*(2), 164–175. https://doi.org/10.1016/j.giq.2010.06.008

Yaqub, U., Sharma, N., Pabreja, R., Chun, S. A., Atluri, V., & Vaidya, J. (2018). Analysis and visualization of subjectivity and polarity of twitter location data. In C. C. Hinnant, A. Zuiderwijk, B. Lal, Y. K. Dwivedi, N. P. Rana, D. Frost, & S. Chirara (Eds.), *Understanding "development" from the perspective of E-government, digital divide and ICT4D literature: A research agenda* (pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/3209281.3209313.

Yaqub, U., Sharma, N., Pabreja, R., Chun, S. A., Atluri, V., & Vaidya, J. (2020). Location-based sentiment analyses and visualization of twitter election data. *Digital Government: Research and Practice, 1*(2), 14 [1-19] https://doi.org/10.1145/3339909.

Zahra, S. A., & George, G. (2002). Absorptive capacity: A review, reconceptualization, and extension. *The Academy of Management Review, 27*, 185–203.

Zeleti, F. A., & Ojo, A. (2019). *Agile mechanisms for open data process innovation in public sector organizations: Towards theory building*. S. Ben Dhaou, L. Carter, & M. Gregory (Eds.) (pp. 164–174). ACM. https://doi.org/10.1145/3326365.3326387.

Zhang, M. M., Wang, X., & Hu, Y. (2023). Strategic framing matters but varies: A structural topic modeling approach to analyzing China's foreign propaganda about the 2019 Hong Kong protests on twitter. *Social Science Computer Review, 41*(1), 265–285. https://doi.org/10.1177/08944393211042575

Zuiderwijk, A., Pirannejad, A., & Susha, I. (2021). Comparing open data benchmarks: Which metrics and methodologies determine countries' positions in the ranking lists? *Telematics and Informatics, 62*, 1–23, 101634 https://doi.org/10.1016/j.tele.2021.101634.

Juliane Schmeling holds a Master's degree in Public Governance and began her role as a research associate at the renowned Fraunhofer Institute for Open Communication Systems in 2012. Within the institute, she drives forward research initiatives, especially in the areas of public sector innovation, digital transformation, and the development of data-informed government infrastructures.

Sami al Darkruni has been employed as a student assistant at the Fraunhofer Institute for Open Communication Systems since 2022. In his role, he is actively involved in conducting research in the field of digital government, delving into areas such as digital governance at the intersection of technology and public administration.

Ines Mergel is a full professor of public administration at the department of politics and public administration, University of Konstanz, Germany. She conducts research on the digital transformation of public bureaucracies and its impact on positive changes of public sector organisations.