

Winter 12-31-2017

# A Comprehensive Review of Data Governance Literature

Olivia Benfeldt Nielsen  
Aalborg University, obn@dps.aau.dk

Follow this and additional works at: <http://aisel.aisnet.org/iris2017>

---

## Recommended Citation

Benfeldt Nielsen, Olivia, "A Comprehensive Review of Data Governance Literature" (2017). *Selected Papers of the IRIS, Issue Nr 8 (2017)*. 3.  
<http://aisel.aisnet.org/iris2017/3>

This material is brought to you by the Scandinavian (IRIS) at AIS Electronic Library (AISEL). It has been accepted for inclusion in Selected Papers of the IRIS, Issue Nr 8 (2017) by an authorized administrator of AIS Electronic Library (AISEL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# A Comprehensive Review of Data Governance Literature

Olivia Benfeldt Nielsen

Aalborg University, Aalborg, Denmark  
obn@dps.aau.dk

**Abstract.** Organizations have found that seemingly tedious data problems are fundamentally business problems, and cannot be solved by the IT group alone. Public organizations routinely store large volumes of data about its citizens and while analysis of this data can improve decision-making and better address individual needs, this fails due to a lack of data governance. Data governance has received growing attention from both practitioners and academics as a promising approach to solving organizational data issues. This paper presents a review of data governance literature, classifying authors, research disciplines, methods and related theoretical fields, providing researchers with an overview of this emerging field. The paper is concluded by suggesting four areas for future development of the data governance field in the context of the public sector.

**Keywords:** Data Governance, Literature Review, E-government, Public sector

## 1 Introduction

Although data has long been heralded as “the new oil” [1], organizations still struggle to generate business value. Especially public organizations, which routinely store large volumes of data about its citizens [2], are keen to pursue new opportunities. Yet they are often restrained by seemingly tedious data problems [3]. Issues of quality, availability or accuracy complicate data efforts, but solving these in isolation constitute short-term solutions [4]. Harvesting value from data requires an organization-wide approach and as such cannot be solved by the IT group alone [5]. Here, data governance has been examined by both practitioners and academics as a promising approach to solving these organizational data issues [6], [7]. The objective of this paper is to present an overview of the current state of the data governance field and based on this, identify potential for future research on the governance of data in the public sector.

Many scholars follow Weill and Ross’ [8] definition of IT governance and define data governance as specifying a framework for decision rights and accountabilities to encourage desirable behavior in the use of data [4], [6], [7], [9]. As such, researchers have proposed initial frameworks for designing data governance [7], [10], analysed influencing factors [9], [10], observed the application of data governance in small-medium enterprises [11], [12] and defined data governance principles and activities [4], [13]. While the data governance literature offers valuable contributions, these

approaches all focus on isolated aspects and no systematic review of the data governance literature exists.

To close this gap, this paper presents a comprehensive review of data governance research. While some literature reviews exist, these are focused either on defining the underlying principles of data governance [4] or mapping the data governance activities related to selected decision domains [13]. As data governance is an emerging academic area, understanding the current state of the field is imperative for advancing its knowledge base. A literature review “[...] facilitates theory development, closes areas where a plethora of research exists, and uncovers areas where research is needed” [14], where the specific contribution of cumulative reviews lies in its ability to evaluate available literature on a particular subject matter and inform researchers about a new area for future research [15]. This review considers 62 peer-reviewed journal publications and conference proceedings which study the mechanisms involved in governing data as an asset. The paper draws on Schlichter & Kraemmergaard’s [16] methodological framework for conducting comprehensive literature reviews and Templier & Paré’s [15] approach to cumulative literature reviews.

The purpose is twofold. The first objective is to gain an overview of the current state of the data governance field by classifying authors, research disciplines, methods, units of analysis and related theoretical areas of concern. The second objective is then, based on the overview, to identify potent areas within the governance of data in the public sector that could benefit from further development. To accomplish this, the paper is organized as follows. In section 2, key concepts of the review are established and previous literature reviews introduced. In section 3, the methods for carrying out the review are described. In section 4, findings from the analysis of the selected papers are presented. In section 5, the findings are discussed, concluding the paper.

## **2 Data Governance**

As a common, widely accepted definition of data governance has yet to be established, it is imperative for this review to clarify some aspects of the concept. Where data may be defined as “raw material”, information is data in context [17]. In the practitioners’ community the two are often used interchangeably, and no distinction is made between data governance and information governance [18]. This may be appropriate for communicating with practice, but does not fit the methodology of a systematic, cumulative review, where determining clear boundaries for key concepts is central [19]. Thus, this paper maintains the first distinction and focus solely on data governance.

Both practitioners and researchers frequently discuss data governance in the context of data quality [9], [10], [20] and place great emphasis on improving data quality as the main goal of data governance [20], [21]. While data quality is important, effective data governance must be driven by and aligned with business goals [5], [17]. To account for this, some scholars then adapt Weill & Ross’ [8] definition of IT governance, indicating data-related decisions and behaviors must be

aligned with organizational performance goals. Here, data governance refers to the allocation of decision-making rights and related responsibilities to encourage desirable behavior in the use of data [7], [13], [20].

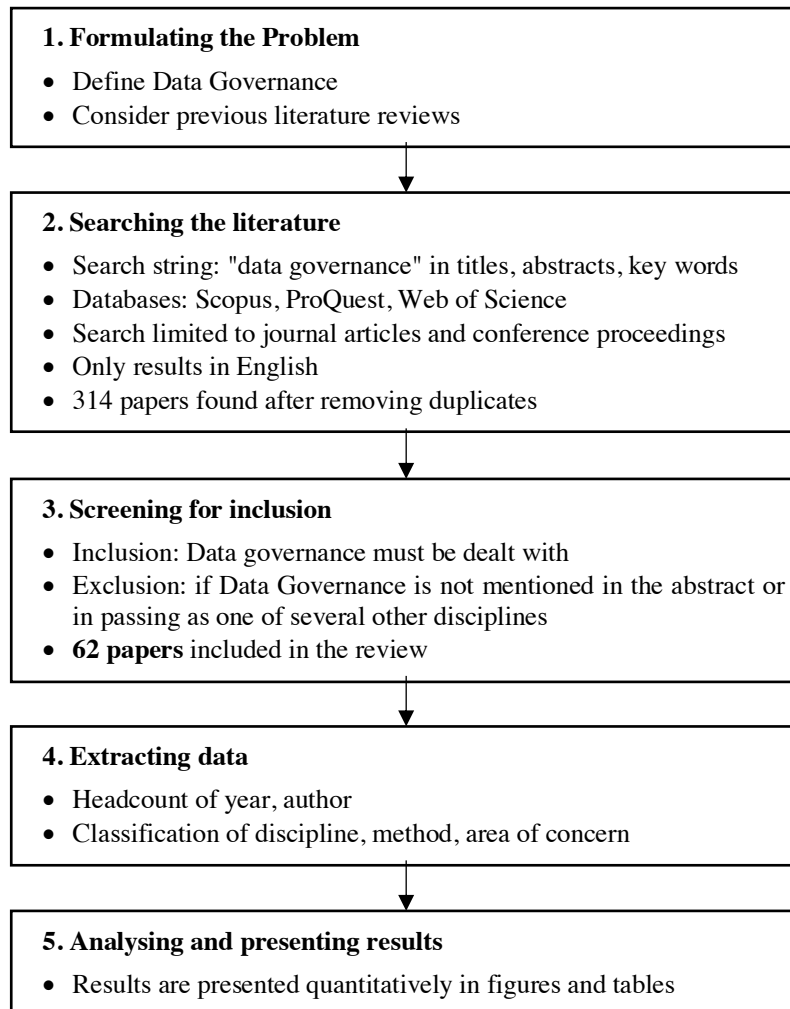
Pierce et al. [22] indicate several definitions of data governance exist among organizations, but provide no further elaboration of these. Instead they adopt a definition of data governance as “the collective set of decision-making processes for the use and value-maximization of an organization’s data assets” [22], adding the notion that data is an enterprise asset, the value of which organizations must work to increase. Otto [20] then defines data governance as “a companywide framework for assigning decision-related rights and duties in order to be able to adequately handle data as a company asset”. This paper adopts an understanding of data governance as companywide processes that specify decision-making rights and responsibilities aligned with organizational goals to encourage desirable behavior in the treatment of data as an organizational asset.

Existing literature reviews have addressed concepts of data governance activities and data governance principles. In their review of 31 papers, Alhassan et al. [13] code the literature to categorize a series of data governance activities. They find a high volume of data governance activities are associated with ‘defining’ areas of governance, but lacking when it comes to ‘implementing’ and ‘monitoring’, and suggest future research focus on examining the latter. Brous et al. [4] consider 35 journal articles, conference proceedings and books to identify data governance principles. They identify four principles, organization, alignment, compliance and common understanding, and suggest they can be used by researchers to focus on important data governance issues, and by practitioners to develop effective data governance strategy.

These papers both aim to advance the knowledge base of the field of data governance through their reviews. Yet they do so only on narrowly defined aspects of the concept. As such, no systematic overview of data governance literature exists, leaving a gap about the state of the research field and little direction for academic professionals and practitioners alike about where future research should concentrate its efforts. It is to close this gap this review addresses the research questions: what is the current state of the data governance field and what areas could benefit from further development?

### **3 Research Method**

To address the questions, this review uses a structured methodology adopted from Templier & Paré’s [15] procedure for conducting a cumulative literature review. The cumulative review compiles evidence to map bodies of literature, draw overall conclusions and work to provide readers with a comprehensive description of the current state of knowledge in the area.



**Figure 1.** Structure for research method adapted from Templier & Pare

As such, the research method is divided into five major steps, represented in Figure 1. As this paper follows the steps specified for cumulative reviews, the process is not meant to be iterative, but rather chronological.

The first step includes clearly defining the concepts to be studied and establishing the boundaries of the review. Where some literature reviews often start wide and narrow down throughout the process, cumulative reviews use predefined concepts. As such, data governance was defined in section 2, and boundaries of the concept were established.

The second step involves searching the literature using a documented search procedure. The search was conducted in the spring of 2017, using the search string “data governance” and included titles, abstracts and author keywords. Only academic journal publications and conference proceedings were included in the search, as these for the most part constitute peer-reviewed, up-to-date information that book chapters do not. Only results in English were considered. Duplicates were removed, leaving a total number of 314 papers for further examination.

The third step includes specifying criteria for excluding or including studies based on the review’s goals. As a broad and comprehensive search was conducted in step 2, it yielded many papers that were not particularly relevant to the overall research questions. To ensure only appropriate papers that help answer the research questions were selected criteria for inclusion and exclusion were established. Papers were only selected if they followed a definition of data governance according to the conceptualization defined in section 2, and excluded if data governance was only mentioned in passing, as one of many other disciplines. 62 papers were included in the final pool of papers.

The fourth step involves identifying the type of data to be extracted from the pool of studies and using a structured approach to do so. To extract data, this paper drew on a framework for conducting comprehensive literature reviews that was developed by Schlichter and Kraemmergaard [16] to analyze a vast amount of academic publications. The framework specifies two strategies for extracting data, using headcounts and classifications. Headcounts were used to address how many papers were published each year, and which authors contributed the most. This was done by simply counting the number of articles published per year and counting how many articles each author had published.

Classifications were used to address what research disciplines, methods, units of analysis and other theoretical or conceptual areas of concern were used to study data governance. To classify papers according to research disciplines Schlichter and Kraemmergaard’s categories on information systems, computer science, organisation and management were used and health informatics, e-government, education and other were added after reading the abstracts of the 62 papers (see Table 1). For classifying papers according to research methods, Schlichter and Kraemmergaard’s categories on case study, theoretical, survey, experiment combined, and design science were used (see Table 2). Additionally, some thematic analysis of the abstracts was undertaken to classify papers according to what other theoretical areas of concern are used to study data governance, as well as units of analysis. An open coding technique was used, where categories were collected as the papers’ abstracts were read.

**Table 1.** Classification of Research Disciplines

<b>Discipline</b>	<b>Description</b>
Information systems	The use or management aspects of information technology
Computer Science	Technical aspects of information technology
E-Health	Use of information technology in healthcare
Organisation	Organisational and management issues that do not have an explicit

and management	technical view
E-government	Use and management of information technology to provide public services to citizens
Education	Higher education institutions and learning
Other	Not focusing on any of the above and only present once in the pool of papers

**Table 2.** Classification of Research Methods.

<b>Category</b>	<b>Description</b>
Case study	Papers reporting on studies involved with a single site or a few sites over a certain period of time
Theoretical	Papers analyzing or synthesizing existing theory, typically with the aim of developing new contributions
Design science	Papers that construct systems or tools
Survey	Papers gathering data from questionnaires
Experiment	Papers that use field experiments to test hypotheses
Action research	Papers where methods of data collection were also used to change a process in practice
Combined	Papers relying on more than one method

The last step of the method concerns summarizing the included studies and presenting the findings of the review. To give an overview of the field, the results of this review were presented through the use of tables and figures, rather than summaries of the individual texts.

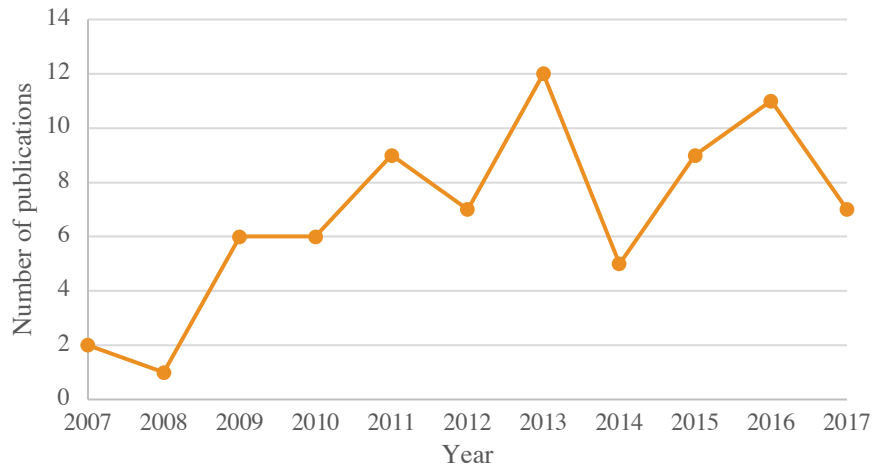
## 4 Findings

In this section, the findings from the literature review are presented. First, the questions related to authors, and year will be answered, namely: how many articles have been published each year and which authors have contributed the most? Then the questions of what research disciplines contribute, what methods are used and how methods are distributed across disciplines will be addressed. Finally, the question of which theoretical areas of concern and units of analysis are drawn upon will be addressed through a presentation of the identified concepts.

### 4.1 Publications, authors and time frame

A total number of 62 papers are included in the pool of papers, published between 2007 and 2017, spanning ten years (see Figure 2). The number of publications rises steadily from 2008 to 2013, where it drops from 12 papers annually to 5 in 2014.

After that, the curve recovers reaching 11 published papers in 2016. 7 papers are published in 2017, but can be attributed to the fact this review was conducted in the early spring of 2017.



**Figure 2.** Publications per year

Overall, the 62 articles were published in 55 different journals. In total, 155 authors contributed to the papers on data governance, either as single or co-author. 138 authors only contributed one article, whereas 13 authors contributed two articles and four authors contributed three articles. For authors, who have authored three articles, information on affiliation and country also show a varied distribution with not two authors working from the same institution or country (see Table 4). Furthermore, some of the authors are not affiliated with universities, but rather consulting agencies or private research institutions.

**Table 3.** Number of publications and authors

Number of publications	Number of authors
3	4
2	13
1	138
<b>Total</b>	<b>155</b>

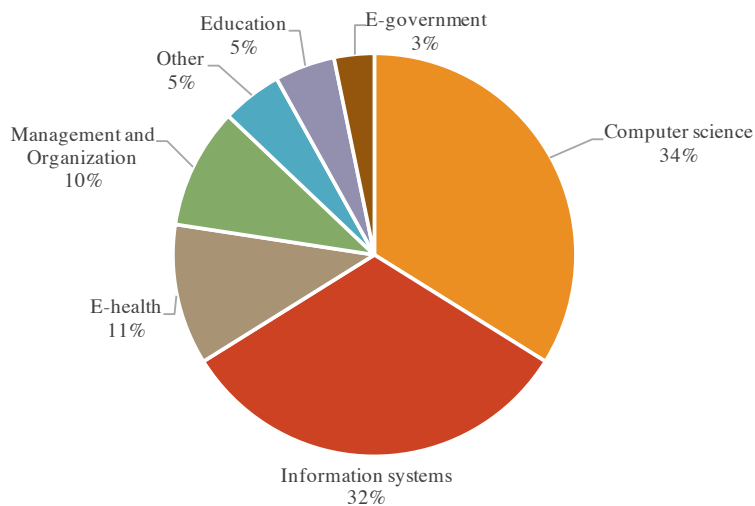


**Table 4.** Authors and affiliations

Author	Affiliation	Country
Otto, Boris	Managing Director, Fraunhofer ISST,	Germany
Brous, Paul	Global Lead Data Architect, Unit4	Netherlands
Panian, Zeljko	Professor, Graduate School of Economics and Business	Croatia
Majid Al-Ruithe	Lecturer, Staffordshire University	UK

## 4.2 Research Disciplines

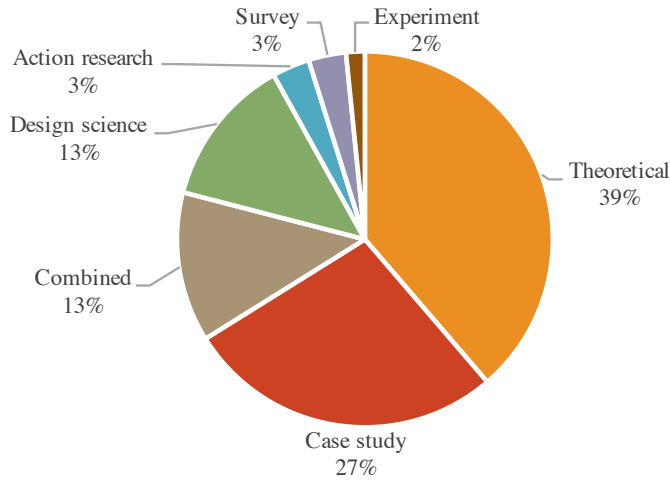
Of the papers, a majority were published either in computer science with 34% of the articles, or information systems with 32% (see Figure 3). This constitutes 2/3 of the total pool of papers, indicating these two are the most prolific disciplines, when it comes to publishing on data governance. Additionally, E-health contributes 11% and Management and Organization 10%, contributing 20% of the overall papers. Education and Other contribute 5% each, with E-government only making out 3%.



**Figure 3.** Distributions of papers among research disciplines

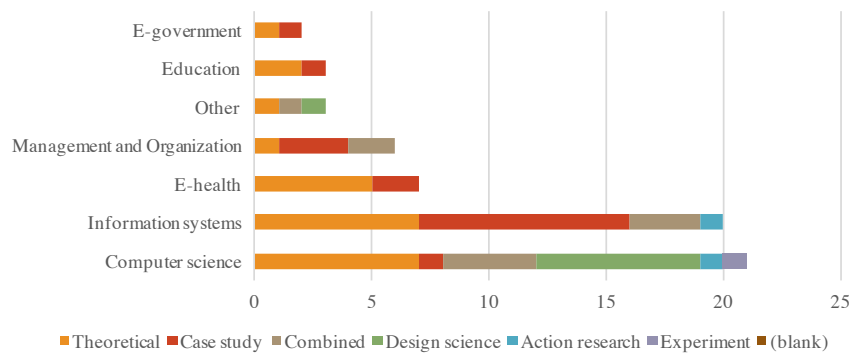
## 4.3 Research Methods

As for the research methods identified, theoretical was the most prevalent at 39% followed by the case study with 27% (see Figure 4). 13% of the articles used a combination of methods, most often theoretical combined with another method, only 13% used design science, with 3% using survey and action research and 2% experiments.



**Figure 4.** Research methods

The distribution of methods across the different research disciplines indicate most of the theoretical publications are distributed within computer science or information systems, constituting nearly half of the total publications in this research discipline (see Figure 5). Furthermore, the majority of case studies are employed in information systems, comprising one third of the discipline’s employed methods and nearly half the papers using the case study method. Computer science is the most diverse discipline, incorporating at least one study employing each method, except for the survey. Combined methods are used mainly in Computer science and Information systems.



**Figure 5.** Distribution of methods across disciplines

Both E-government, Education and E-health are only comprised of case studies and theoretical papers, while design science is used mainly in Computer science, except for one study in the other category.

**Table 5.** Frequency of methods

Method	Frequency	
	<i>n</i>	%
<b>Theoretical</b>	24	39%
<b>Case study</b>	17	27%
<b>Combined</b>	8	13%
<b>Design science</b>	8	13%
<b>Action research</b>	2	3%
<b>Survey</b>	2	3%
<b>Experiment</b>	1	2%
<b>Total</b>	62	100%

#### 4.4 Related Areas of Concern

During the reading of the abstracts, 10 different theoretical areas of concern were mentioned in relation to data governance (see Table 6). These may be categorized following Gregor's [23] different theory types. Three out of the 10 areas, business process management, data quality management and master data management may be considered *Theory for Design and Action*, which according to Gregor is about "how to do something" [23], and account for five articles in total. Contingency theory, organizational design and resource based view relate to overall management of an organization. This may be considered *Theory for Explaining and Predicting*, which is "understanding of underlying causes and prediction, as well as description of theoretical constructs and the relationships among them"[23], accounting for four of the articles. Panopticism, theory of paradoxes and system of systems constitute grander theories mostly concerned with world views. They may be considered *Theory for Explaining*, which Gregor deem less concerned with developing testable hypotheses and more about "understanding phenomena" [23], accounting for three articles.

IT governance is by far the most prominent theory used, supporting the notion that the data governance field is in large part build on the same foundation.

**Table 6.** Identified areas for concern

Theory	Description	Frequency
<b>IT Governance</b>	aligning IT with performance goals and assigning accountability for those actions and their outcomes.	8

<b>Data Quality Management</b>	the disciplines related to managing data as an enterprise asset in order to improve quality and increase value	3
<b>Resource Based View</b>	the idea that competitive advantage lies primarily in the application of a bundle of valuable tangible or intangible resources at the firm's disposal	2
<b>Master Data Management</b>	disciplines related to enabling an enterprise to link all of its critical data to one file, called a master file, that provides a common point of reference	1
<b>Business Process Management</b>	the development and implementation of business solutions	1
<b>Contingency Theory</b>	the idea that that there is no best way to organize a corporation and the optimal course of action is contingent upon internal and external factors	1
<b>Organizational Design</b>	the frameworks through which organizations aim to realize their core qualities and goals	1
<b>Panopticism</b>	the idea behaviour can be modified or controlled if subjects are unsure when or if they are under surveillance	1
<b>System of Systems</b>	collection of task-oriented or dedicated systems that pool their resources together to create a new, more complex system which offers more functionality and performance	1
<b>Theory of Paradoxes</b>	the study of problems or dilemmas that cannot be solved and thus must be managed as paradoxes	1

#### 4.5 Unit of Analysis

During the reading of the abstracts, a concept-centric analysis was undertaken and a pattern emerged. 85% of the papers examined data governance at the organizational level, which makes sense considering the definition of data governance often includes a conceptualization of enterprise-wide frameworks for accountability. It is interesting to note however that data governance is also examined as a way to manage data between discrete organizations. Two papers even consider how data governance can facilitate data sharing across the private and public sector.

**Table 7.** Unit of analysis

<b>Unit of analysis</b>	<b>Frequency</b>
Organization	53
Inter-organizational	6

Cross-sectoral	2
Regional	1
Total	62

## 5 Discussion

As such, the state of the data governance field can be tentatively evaluated from the findings in section 4. The distribution of relatively few publications amongst a plethora of authors suggest the field has not matured enough for researchers to publish several articles within the domain. Additionally, the top publishing authors count both practitioners and academics, which suggests it is a field attracting the attention of both communities.

The data governance field is dominated largely by theoretical methods and case studies, with few studies using more practice-oriented methods, like design science, action research or experiments. Considering the presence of practitioners in the research field and the potential benefits to be derived from implementing data governance in practice, more studies of this kind should be encouraged.

In the same vein, most of the publications are within the fields of information systems and computer science, which could suggest the field is still largely driven by the IT oriented researchers. This could also have implications for how the discourses surrounding data governance are shaped. Few studies are conducted within E-government and Organization and Management, which suggests data governance is still anchored in IT, and has yet to become a general management discipline. A heavy presence of theoretical papers in largely technical fields could slow or obscure the development of data governance as a management practice, because it may continuously be framed as an IT responsibility. Additionally, a majority of papers also draw on IT governance to conceptualize data governance, which only work to strengthen the idea of data as something inherently connected to IT.

In addition, many of the related theoretical areas are predominantly prescriptive, rendering data governance largely intangible. Presented mainly through frameworks or charts, in the various management and design disciplines, data governance is dealt with on a formal basis, with little attention paid to the context it will be implemented in. Some studies do adapt more explanatory approaches, such as the theories of panopticism and paradoxes, but there is little no attention paid to the processes of change and implementation of these formal frameworks, which could be relevant for practice. This is further supported by the majority of studies considering data governance at the organizational level, rather than the intra-organizational level between different member groups or functions.

Lastly, the relatively low number of studies within E-government and the lack of studies in public administration suggest there is potential for further research here. While two publications actually focus on the dynamics between public and private sector organizations in the use of data as an organizational resource, these are in the

minority. From these considerations, some areas for future development of the field of data governance in the context of public organizations may be identified (see Table 8)

**Table 8.** Suggestions for further development

<b>Suggestion</b>	<b>Purpose</b>
<b>Case studies that examine data governance through E-government and Management and Organization perspectives</b>	To move away from seeing data as IT's responsibility and toward seeing data as an enterprise asset that moves across the organization and must be governed as such
<b>Experiments of data governance in E-government contexts</b>	To examine if and how data governance can facilitate improvements or innovation in public administration
<b>Action research that consider data governance at the intra-organizational level</b>	To engage practice and examine how to implement some of the many formal governance frameworks at a group or functional level within an organization
<b>Case studies that examine data governance discourse</b>	To understand how data governance is framed and how this affects its implementation or potential efficiency

From the state of the current field, these four areas for further development can be used by researchers, who wish to engage data governance in the context of public organizations. Focusing on these suggestions will provide rich opportunity for scholars to engage with practice, where data governance has potential.

## 6 Conclusion

This paper aimed to provide an overview of the state of the data governance field by considering a systematic treatment of the literature. 62 papers were considered using headcounts and classifications, and four areas of concern for future development of data governance in the context of the public sector were listed. Limitations include that only the abstracts were read to evaluate the papers and that no deeper analysis of the individual papers was undertaken. Scholars and practitioners may use these findings to position their future work on the field of data governance. As use of data has the potential to facilitate better decision-making and improve service delivery in the public sector, the findings of this paper provide a valuable resource for practice-oriented scholars who wish to develop the field.

## References

1. Ransbotham, S., Kiron, D., Prentice, P. K.: Beyond the Hype: The Hard Work Behind Analytics Success, In: MIT Sloan Management Review, vol. 57, no. 3, pp. 1–16, (2016)

2. Brous, P., Janssen, M., Vilminko-Heikkinen, R.: Activities: A Systematic Review of Data Governance Principles, In: IFIP International Federation for Information Processing 2016, pp. 115–125, (2016)
3. Thompson, N., Ravindran, R., Nicosia, S.: Government data does not mean data governance: Lessons learned from a public sector application audit, In: Government Information Quarterly, vol. 32, no. 3, pp. 316–322, (2015)
4. Brous, P., Janssen, M., Vilminko-Heikkinen, R.: Coordinating decision-making in data management activities: A systematic review of data governance principles, vol. 9820. pp. 115–125, (2016)
5. Lee, Y. W., Madnick, S. E., Wang, R. Y., Wang, F. L., Zhang, H.: A cubic framework for the chief data officer: Succeeding in a world of big data, In: MIS Quarterly Executive, vol. 13, no. 1, pp. 1–13, (2014)
6. Otto, B.: Data governance, In: Business and Information Systems Engineering, vol. 3, no. 4, pp. 241–244, (2011)
7. Khatri, V., Brown, C. V.: Designing data governance, In: Communications of the ACM, vol. 53, no. 1, p. 148, (2010)
8. Weill, P., Ross, J. W.: IT Governance on One Page, In: Harvard Business Press, p. 18, (2004)
9. Weber, K., Otto, B., Österle, H.: One Size Does Not Fit All - A Contingency Approach to Data Governance, In: ACM Journal of Data and Information Quality, vol. 1, no. 1, pp. 1–27, (2009)
10. Wende, K., Otto, B.: A contingency approach to data governance, In: Proceedings, 12th International Conference on Information Quality, Cambridge, USA, pp. 1–14, (2007)
11. Begg, C., Caira, T.: Data Governance in Practice: The SME Quandary Reflections on the Reality of Data Governance in the Small to Medium Enterprise (SME) Sector, In: 5th European Conference on Information Management and Evaluation (ECIME), pp. 75–83, (2011)
12. Begg, C., Caira, T.: Exploring the SME Quandary : Data Governance in Practise in the Small to Medium-Sized Enterprise Sector, In: The Electronic Journal Information Systems Evaluation, vol. 15, no. 1, pp. 3–13, (2012)
13. Alhassan, I., Sammon, D., Daly, M.: Data governance activities: an analysis of the literature, In: Journal of Decision Systems, vol. 25, no. sup1, pp. 64–75, (16 2016)
14. Webster, J., Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review., In: MIS Quarterly, vol. 26, no. 2, pp. xiii–xxiii, (2002)
15. Templier, M., Paré, G.: A Framework for Guiding and Evaluating Literature Reviews., In: Communications of the Association for Information Systems, vol. 37, (2015)
16. Schlichter, B. R., Kraemmergaard, P.: A comprehensive literature review of the ERP research field over a decade, vol. 23, no. 4, (2010)
17. Boisot, M., Canals, A.: Data, information and knowledge: Have we got it right?, In: Journal of Evolutionary Economics, vol. 14, no. 1, pp. 43–67, (2004)
18. Ladley, J.: Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program. Newnes, (2012)
19. Arksey, H., O'Malley, L.: Scoping studies: towards a methodological framework, In: International Journal of Social Research Methodology, vol. 8, no. 1, pp. 19–32, (2005)
20. Otto, B.: Organizing Data Governance : Findings from the Telecommunications Industry and Consequences for Large Service Providers, In: Communications of the Association for Information Systems, vol. 29, no. 1, pp. 45–66, (2011)
21. Soares, S.: The IBM Data Governance Unified Process. (2010)
22. Pierce, E., Dismute, W. S., Yonke, C. L.: The State of Information and Data Governance - Understanding How Organizations Govern Their Information and Data Assets, (2008)
23. Gregor, S.: The Nature of Theory in Information Systems, In: MIS Quarterly, vol. 30, no. 3, pp. 611–642, (2012)