ALREADY GROWN-UP OR STILL IN PUBERTY? A BIBLIOMETRIC REVIEW OF 16 YEARS OF DATA GOVERNANCE RESEARCH

Marvin Jagals [*], Erik Karger [**], Frederik Ahlemann [**]

* *Corresponding author*, Chair of Information Systems and Strategic IT Management, University of Duisburg-Essen, Essen, Germany
Contact details: Chair of Information Systems and Strategic IT Management, University of Duisburg-Essen, Universitätsstraße 9, 45141 Essen, Germany
** Chair of Information Systems and Strategic IT Management, University of Duisburg-Essen, Essen, Germany

How to cite this paper: Jagals, M., Karger, E., & Ahlemann, F. (2021). Already grown-up or still in puberty? A bibliometric review of 16 years of data governance research. *Corporate Ownership & Control, 19*(1), 105–120.
https://doi.org/10.22495/cocv19i1art9

ISSN Online: 1810-3057
ISSN Print: 1727-9232

Received: 24.08.2021
Accepted: 14.10.2021

JEL Classification: G31, G32, K22
DOI: 10.22495/cocv19i1art9

Abstract

The amount of data and the speed at which it increases grows rapidly. Companies and public institutions try to manage this increasing flood of data effectively and in a manner that adds value. Besides, the companies and public institutions also join corporate networks or platforms to increase their value by sharing their data. The evolution of traditional business intelligence into business analytics, including real-time analysis, increases the high demand for qualitative data. Data governance tries to create a framework to manage these issues. This interdisciplinary research field has now been in existence for nearly two decades. With this contribution, we attempt to provide the research field with a blueprint. This paper aims to explore the past to understand the present and shape the future of data governance. We give an overview of how the research field changed from 2005 to 2020, commenting on its development and pointing out future research paths based on our findings. We, therefore, conducted a bibliometric analysis to describe the research field's bibliometric and intellectual structure. The findings show that for years the research field concentrated on a few topics, which currently undergoes change and has led to an opening up of the research field. Finally, the results are discussed and future research strands are highlighted.

Keywords: Data Governance, Data Management, Data Quality, Governance, Bibliometric Review, Quantitative Analysis, Big Data Governance

Authors' individual contribution: Conceptualization — M.J.; Methodology — E.K.; Validation — M.J.; Formal Analysis — M.J.; Investigation — E.K.; Resources — M.J.; Data Curation — E.K.; Writing — Original Draft — M.J.; Writing — Review & Editing — E.K.; Visualization — M.J.; Supervision — F.A.; Project Administration — M.J.

Declaration of conflicting interests: The Authors declare that there is no conflict of interest.

Acknowledgements: We acknowledge support by the Open Access Publication Fund of the University of Duisburg-Essen.

1. INTRODUCTION

Governance literature has traditionally used an agency theory approach, concentrating solely on identifying conflicts of interest (agency problems) between management and shareholders, where one group delegated work to another. In this regard, the division of decision rights and accountabilities concerning assets is the crucial anchor of the whole governance mechanism (Akerlof, 1978; Eisenhardt, 1989; Jensen & Meckling, 1976).

Furthermore, the concept of data as an asset dates back to the 1980s. This was when methodologies and expertise from the administration of physical products were applied to the management of immaterial assets for the first time, with a strong focus on data administration (Legner, Pentek, & Otto, 2020). The purpose was to stimulate executives to recognize their organization's information as a strategic asset (Gillenson, 1985; Horne, 1995). However, in a space of ten years, numerous authors realized that a governance framework is urgently needed to manage these upcoming challenges of treating data as an asset. These first multiple occurrences of contributions that dealt with topics concerning the new notion of data governance were timely related to the simultaneously increasing amount of data (Cheong & Chang, 2007; Griffin, 2005; Wende, 2007).

Today, data is regarded as an enabler of a firm's business models and value propositions. Strategic data management focusing on data-driven innovation is a must-have property of inventive organizations (Schüritz, Seebacher, Satzger, & Schwarz, 2017). In this context, the rise of self-styled data ecosystems was understandable. Within these ecosystems, which present novel collaboration forms between various actors, organizations share their data assets, for example, to reduce costs or increase their revenues (De Prieëlle, De Reuver, & Rezaei, 2020). This sight on governance from an inter-organizational perspective therefore begins to gain more research attention in scientific communities, such as business, information systems, and computer engineering research (De Prieëlle et al., 2020; Schreieck, Wiesche, & Krcmar, 2016). Especially in the last few years, this kind of publication on data governance has risen very sharply.

Almost two decades have passed between the first emergence of data governance papers and today's focus on data as a value proposition enabler or an inter-organizational sharing good. Within this timeframe, data governance research (DGR) has clearly developed exorbitantly fast in its breadth and depth. A blueprint is therefore needed to not only draw a conclusion concerning but also obtain a general overview of the past research to lay the foundations for future endeavors. The research field faces key decisions and will take on a central role because the framing of data management will become even more crucial in the context of upcoming domains like big data, Industry 4.0, and artificial intelligence (Amadori, Altendeitering, & Otto, 2020; Winter & Davidson, 2019). Obviously, the domains where data governance has been applied have evolved — but so has the concept itself. With this publication, we aim to provide a holistic view of DGR since its beginning. To grasp what is known and what needs to be learned, it is critical to comprehend the evolution, the level of scientific knowledge, and the structure of the research field.

To reach this goal, the following research questions are created:

RQ1: What are the leading and most influential papers in terms of popularity and prestige in the field of data governance within business, information systems, and computer engineering research?

RQ2: What are the existing clusters of collaborating authors within DGR?

RQ3: Where does the field currently stand, and how can the field progress and mature?

For this purpose, we use a bibliometric analysis of all data governance contributions. Bibliometric analysis is a research field of information and library sciences that analyzes bibliographic material quantitatively. One of the main advantages of bibliometric analysis is that it provides general overviews identifying the leading trends of the bibliographical material considered (Broadus, 1987).

Within the following sections, we give an overview of DGR to locate our study. In the third section, we provide details about the research process, and then we present our findings in the fourth section, which are discussed and placed in the overall context of data governance within the discussion. Lastly, in the fifth section, we identify possible research avenues for more specific investigations and summarize the main findings within our concluding remarks.

2. RELATED WORK

The goal of corporate governance research is to address challenges in cases where principals and agents have conflicting desires, conflicting goals, or conflicting risk attitudes (Akerlof, 1978; Eisenhardt, 1989; Jensen & Meckling, 1976). Dividing decision rights, determining how parties can be held accountable, and examining how rewards can be used to address divergent interests can all be viewed through the prism of agency theory (Fama & Jensen, 1983).

The rights concerning an individual asset are referred to as decision rights. The optimum allocation of decision rights is important for businesses, markets, and cultures to achieve positive outcomes (Grossman & Hart, 1986; Simon, 1951). Fama and Jensen (1983) differentiate between decision management rights, which make it feasible to create decision proposals and conduct decisions, and decision control rights, which involve deciding whether to execute decisions and tackling how decisions are observed. The latter subdomain is related to accountability. The most basic definition is to be accountable for one's actions (Mulgan, 2000). However, this is just one aspect of establishing transparency. Officers responsible for their actions and the consequences they cause must deal with the actions and the consequences (Burritt & Welch, 1997).

The concepts of accountability and decision rights were the foundation for the development of information technology (IT) governance (Weill & Ross, 2004) and data governance (Weber, Otto, & Österle, 2009), as both research streams are essentially based on these two concepts. IT governance has primarily evolved from the initial concept of corporate governance and is a more mature area than the even more recent data governance concept (De Haes & Van Grembergen, 2008). Weill and Ross (2004) understand IT governance as an instrument to support the resulting requirements.

In the understanding of data governance as an outgrowth of the original corporate governance, this paper follows the interpretation of Otto (2011c). He construed data governance as companywide policies that describe decision-making rights and

duties linked with organizational objectives to stimulate desired behaviors surrounding treating data as a business asset. Researchers and practitioners generally agree that data governance is primarily concerned with assigning roles, decision-making authority, and duties (Khatri & Brown, 2010; Weber et al., 2009), emphasizing that these interpretations are subject to a broad understanding of data governance. Research also focuses on definite areas of data governance, such as data quality, data security, and data processes (Tallon, Ramirez, & Short, 2013, p. 142). Based on this assumption, a data governance framework should discuss accountability: appointing individuals to data management positions and giving them the power to enforce, consolidate, and oversee all enterprise-wide data-related activities by binding their success to benefits or rewards (Griffin, 2005). Organizational instead of technical issues are, however, more critical to the success of data governance.

Besides, scholars have recently expanded the scope of data governance research to include external interactions in network-related formations (Lis & Otto, 2021; De Prieëlle et al., 2020; van den Broek & van Veenstra, 2015). However, presently there appears to be no consensus on data governance in inter-organizational environments. Furthermore, not only has the research field expanded in organizational scope and at the same time diluted, but it has also expanded in data scope (traditional versus big data) and in domain scope of data governance (structural mechanisms versus procedural and relational mechanisms) (Abraham, Schneider, & vom Brocke, 2019). This also emphasizes the need for a quantitative bibliometric meta-analysis of the research field to provide a detailed overview of the research field's overall development.

3. RESEARCH METHOD

3.1. Goals and database selection

To answer the research questions of this paper, a bibliometric analysis was conducted. The bibliometric analysis is a research method that has increasingly gained relevance since its first introduction by Pritchard (1969). Scholars can use the method to understand a research field's temporal developments and structures (Tandon, Kaur, Mäntymäki, & Dhir, 2021). In recent research, bibliometric studies have been conducted for a large number of different topics and areas. Examples include corporate governance research (Eulerich, Haustein, Zipfel, & van Uum, 2013), family businesses (Bekiaris & Papanastasiou, 2021), the development of business model research (Budler, Župič, & Trkman, 2021), and interactive digital marketing (Krishen, Dwivedi, Bindu, & Kumar, 2021).

Bibliometrics is nowadays an essential tool for measuring many different characteristics and metrics of scientific output. Examples include the production of researchers (Ellegaard & Wallin, 2015), the collaboration between different institutions (Skute, Zalewska-Kurek, Hatak, & de Weerd-Nederhof, 2019), and the most relevant sources and outlets for a specific academic field. In our study, the bibliometric analysis is carried out to investigate the most influential papers, outlets, and authors. Furthermore, we aimed to investigate clusters of the author's keywords and the intellectual structure of the research field by investigating co-citation networks and relevant key terms in different periods.

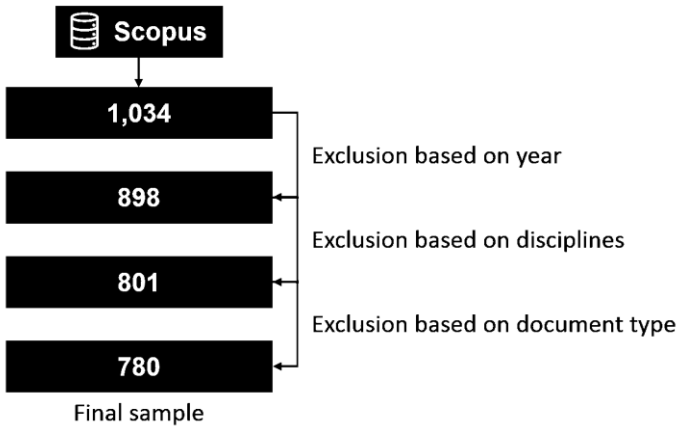
The first step is the collection of metadata that is further used for the bibliometric analysis. Information that can serve as the foundation for bibliometric insights is, among others, citations, outlets and sources, keywords, or institutions (Moral-Muñoz, Herrera-Viedma, Santisteban-Espejo, & Cobo, 2020). Although there are many databases for scientific publications, Scopus and the Web of Science (WoS) are the largest and most relevant ones (Forliano, De Bernardi, & Yahiaoui, 2021). We chose Scopus as the database for our initial search of the literature. Scopus covers more scientific journals than WoS (Paul & Criado, 2020). Furthermore, the search of Scopus takes place in patent databases (Kulkarni, Aziz, Shams, & Busse, 2009), which might also help cover publications dealing with data governance from a more technical point of view. The authors of this paper have also chosen several other bibliometric studies over the past years (Caviggioli & Ughetto, 2019; Donthu, Kumar, & Pattnaik, 2020; Fahimnia, Sarkis, & Davarzani, 2015; Tandon et al., 2021).

3.2. Data collection and extraction

To collect data from Scopus, we used "Data Governance" as the search term. The search was executed on July 15, 2021, and resulted in an initial sample of 1,034 identified documents. First, since we focused on publications until 2020, we excluded articles published in 2021. This led to the exclusion of 136 publications, with 898 documents remaining. We followed the argumentation of Forliano et al. (2021) and Merigó, Mas-Tur, Roig-Tierno, and Ribeiro-Soriano (2015) and excluded more recent publications to improve the comparability of the bibliometric analysis. When including articles that have been published too recently, the comparability of the bibliometric analysis may be impaired since there was not sufficient time for the recent publications to be read and cited (Merigó et al., 2015).

As a next step, we excluded articles of disciplines that are not related to our research question. We excluded disciplines with a strong focus on natural or physical sciences. This led to excluding all articles related to specific topics like dermatology, psychology, or chemical engineering. Our core research lines are therefore business, information systems, and computer engineering research. All directly related disciplines were also left in the sample. After filtering based on disciplines, 801 publications remained. As the last step, we filtered irrelevant document types. This led to a final sample of 780 publications that provided the base for our bibliometric analysis. We exported the final sample as a BibTex and a CSV (comma-separated values) format for further analysis. The whole process of literature exclusion is depicted in Figure 1.

Figure 1. Overview of the literature collection and exclusion criteria



3.3. Data analysis

There is a wide range of different software, tools, and libraries for conducting the bibliometric study (for a current overview, see Moral-Muñoz et al., 2020). To conduct our bibliometric analysis, we used two visualization tools and one R package. R is an open-source software ecosystem that includes statistical techniques, mathematical capability, and graphical features, making it an effective tool for bibliometric analysis. R is a programming language that runs on Windows and Linux and has a graphical user interface (RStudio) that makes it easy to use for both novice and advanced users (Derviş, 2019). First, we used the visualization tool VOSviewer, which was developed by the Centre for Science and Technology Studies (CWTS) at Leiden University (The Netherlands) (Moral-Muñoz et al., 2020). VOSviewer allows the creation of a broad range of bibliometric visualizations, considering authors, journals, or bibliographic coupling (Moral-Muñoz et al., 2020; van Eck & Waltman, 2010). VOSviewer has been used in several bibliometric studies that have recently been published (see, e.g., Khanra, Dhir, Parida, and Kohtamäki, 2021; Fahimnia et al., 2015; Tandon et al., 2021; Danvila-del-Valle, Estévez-Mendoza, and Lara, 2019). Furthermore, we used Bibliometrix, which is an open-source R package developed by Aria and Cuccurullo (2017). Bibliometrix allows graphs and supports features that are not common in other libraries (Moral-Muñoz et al., 2020). We furthermore complemented Bibliometrix with Biblioshiny. Biblioshiny was programmed with R, and is a web-based graphical interface (Moral-Muñoz et al., 2020). We found supplementing VOSviewer with Bibliometrix and Biblioshiny meaningful because, considering the number of supported analysis options, "Bibliometrix and its user interface Biblioshiny stand out since they incorporate a great variety of different analyses" (Moral-Muñoz et al., 2020, p. 16). Although VOSviewer is highly suitable for analyzing and visualizing keyword co-occurrences, Biblioshiny stands out in the statistical analysis of bibliometric data. While Biblioshiny was used to conduct the performance analysis, we used VOSviewer to create the keyword co-occurrence visualizations.

4. FINDINGS

4.1. Overview and annual production

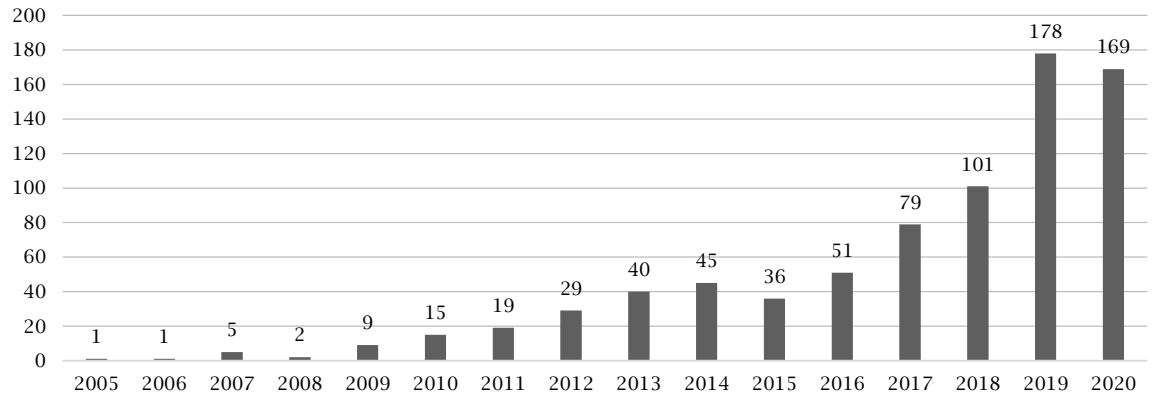
In the following, we present a number of initial metrics and statistics with regard to this article's final sample. A summary is given in Table 1. In total, our initial sample included 780 publications published in 491 different sources including conference proceedings, journals, and anthologies. In the 1990s, the first publications appeared that initially dealt with the topic of data governance. In 2005, however, DGR actually started, since publications in this area have continuously increased since then. This is in line with the results of our literature collection, since the first publication Scopus identified was published in 2005 (Trope & Power, 2005). On average, the documents of our initial sample have been cited seven times. In total, 24,427 references were cited and 1,726 keywords were used throughout the documents. Two thousand one hundred and twenty-five different authors have contributed to the research in data governance, with 2,485 author appearances in the documents. Of the 780 documents, 173 were single-authored, which is equal to 22.12%. The last row of Table 1 shows the collaboration index (CI). This index attempts to measure how strong and frequent cooperation and collaborations between researchers are. The CI is calculated by dividing the total number of authors of multi-authored articles by the total number of multi-authored articles (Elango & Rajendran, 2012; Koseoglu, 2016). Of our 780 publications, 173 were single-authored. This results in 607 multi-authored documents. Dividing the number of authors of multi-authored documents (1,976) by the 607 multi-authored documents results in a CI of 3.26.

Table 1. Overview the final sample's general metrics

Metric	*Value*
Timespan	2005–2020
The total number of documents	780
Sources	491
Average citations per document	7
Average citations per year per document	1,307
The total amount of references	24,427
Author's keywords	1,726
Authors	2,125
Author appearances	2,485
Authors of single-authored documents	149
Authors of multi-authored documents	1,976
Single-authored documents	173
Multi-authored documents	607
Documents per author	0.367
Authors per document	2.72
Collaboration index	3.26

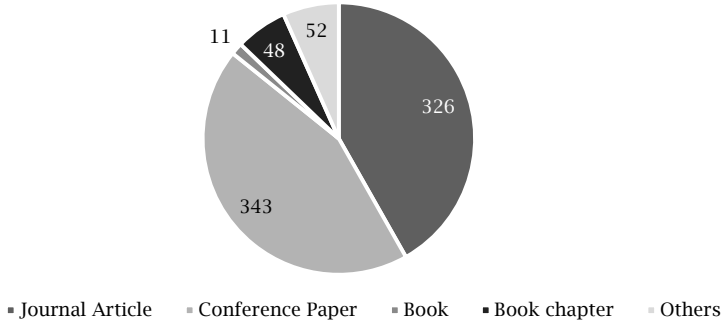
Publications have increased over the past 16 years. In 2005, one data governance-related publication was published. This increased to 178 publications in 2019 and 169 publications in 2020. The trend continues upward (Figure 2). It is striking that in the first years, the increase and development of articles were very slow. After only one publication in 2005 and another one in 2006, there was a small increase in 2007. As the importance of data management has generally increased, so have articles in the research field of data governance also gained more relevance. As a result, the largest increases per year have recently occurred in 2017, 2018, and 2019. The main cause is the tightening of far-reaching data protection regulations, for example, the General Data Protection Regulation (GDPR) within the European Union in 2018 (Voss, 2019; Voss & Bouthinon-Dumas, 2020). In large parts of corporate practice and academia, data governance is still partly or even exclusively associated with information security or data protection (Al-Ruithe, Mthunzi, & Benkhelifa, 2016; Begg & Caira, 2009; Power & Trope, 2006). Another reason for the steep increase is the attention and importance of the topic of big data and big data analytics, which was also responsible for part of the significant increase in contributions within the last years (Al-Badi, Tarhini, & Khan, 2018; Betzu, Coinu, & Demuro, 2019; Kim & Cho, 2018). The last major domain responsible for the high increase in publications deals with data governance mechanisms at the inter-organizational level. This research direction is also often referred to as data ecosystem research (Lis & Otto, 2021). This line of research has now existed for a few years. This becomes clear when one considers the years of publication. All publications involving the appropriate treatment of inter-organizational issues were, with a few exceptions (Felici, Koulouris, & Pearson, 2013), only published as of 2017.

Figure 2. Number of publications per year



During the early years, the research field of data governance was not yet mature enough to be represented in journals with a high impact factor. Most papers appeared in non-scientific outlets or only in proceedings of conferences. This was mainly because primarily practitioners or a wide variety of consultancies dealt with data governance. Specific journals, calls for papers in special issues, or specific tracks at conferences did not exist. Over the last ten years, however, this has changed. Journal articles, along with conference papers, form the core of the outlets (Figure 3).

Figure 3. Distribution of contributions among different types of publications



4.2. Trending topics and content

In analyzing the keywords used during the first seven years of our study period, we identified that data quality was the core driver of the research field. One can even argue that data quality and data quality management were often the sole objectives of data governance; in certain cases, data quality management was also equated with data governance (Clement, Guetari, & Laboisse, 2010; Even & Shankaranarayanan, 2007). Especially in the 2000s, many contributions placed data governance's goal and success parameters primarily on data quality. This is also shown by master data management having been present in science since the beginning of the 2000s and being understood as a tool for data quality management. Data quality management can be considered as a generalized result of master data management, which was previously based purely on master data (White, Newman, Logan, & Radcliffe, 2006). Between 2005 and 2012, other basic terms were used a lot (Figure 4). Corporate governance is regarded in certain data governance definitions as the basis of data governance and takes the management perspective of this research field (see Section 2). Comparative effectiveness research also took more of a business perspective on data governance and was used frequently during this period (Holve, Segal, Lopez, Rein, & Johnson, 2012).

Figure 4. Wordcloud data governance 2005–2012

Figure 5. Wordcloud data governance 2013–2016

In the following four years, the research field changed thematically for the first time (Figure 5). The cornerstone data quality has remained a core driver, but the rest has disappeared or has lost attractiveness. Topics like cloud computing, business intelligence, and accountability have moved into the spotlight (Al-Ruithe & Benkhelifa, 2017; Avery & Cheek, 2015; Kim & Cho, 2018). The digital transformation picked up speed, the amount of data increased significantly during this period (Vial, 2019). Therefore, not only research fields such as cloud computing but also the increased interest in data accountability have risen strikingly. As the biggest driver, big data has entered the research field, making various contributions to the data governance body of knowledge.

Another four years in the research field have once again demonstrated fundamentally changed research directions (Figure 6). Thus, it remains to be said that the great revolution in the research field took place between 2012 and 2016. Thereafter it only evolved in terms of research depth. Data privacy, privacy, and GDPR, however, occur increasingly, confirming that the stronger focus on information security and data protection is reinforced by introducing the GDPR in the European Union (Voss & Bouthinon-Dumas, 2020).

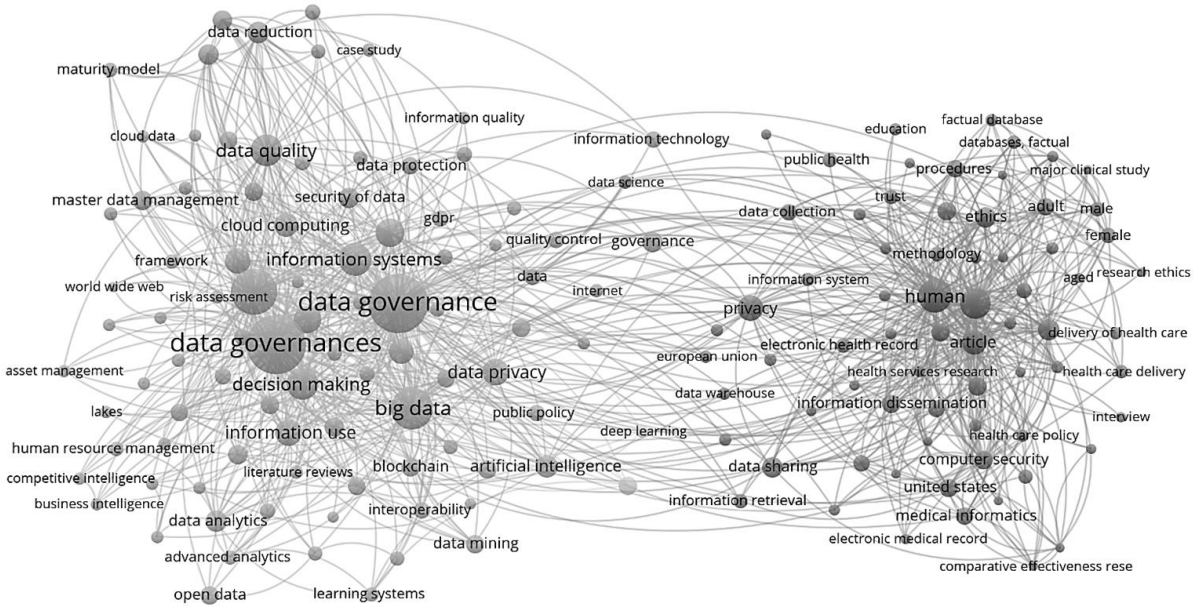
Figure 6. Wordcloud data governance 2017–2020

Figure 7. Wordcloud 2005–2020

Taking a final look at the overall period, it is remarkable that the last few years in the research field have had and will have the greatest influence on the further development of data governance (Figure 7). Topics like cloud computing, big data, and privacy still dominate DGR and will do so for years. Data management and data quality were constantly topical and part of active research throughout the entire period under review.

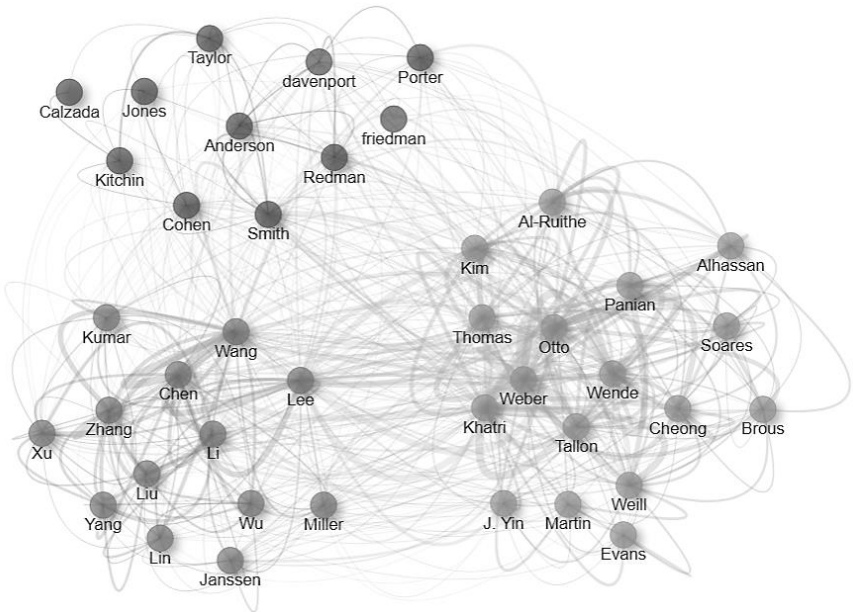
After identifying the most relevant keywords for different periods and for the whole sample, we sought further insights into the exact content and linkages between the keywords. For this purpose, we measured the co-occurrence of the author's keywords. Callon, Courtial, and Laville (1991) were the first to propose this approach to study interactions between academic and technological research. Keyword co-occurrence can be used to identify future research avenues (Tandon et al., 2021), to illustrate and visualize the content of publications (Leung, Sun, & Bai, 2017; Vallaster, Kraus, Merigó Lindahl, & Nielsen, 2019), and to illustrate the logic connection of keywords within recent research. We analyzed the co-occurrence of the keywords used to identify frequent thematic links within the 16-year DGR (Figure 8). The co-occurrence around the core term first forms decision-making, big data, data privacy, and cloud computing, which confirms our results from the analysis of the specific word clouds. Another linking node is the human being itself, which deals with data governance. This also corresponds to the close affiliation with information systems research, which deals with organizations, people, and IT (Banker & Kauffman, 2004). Several domains could be identified, such as ethics and health. The domain of trust can also be found here, receiving increasing attention in data sharing (Cao et al., 2016; James et al., 2014). The third node focuses on data quality as already identified in the word clouds. Data quality has already been established as an independent research field and has a co-occurrence, such as master data management, information quality, or data reduction.

Figure 8. Keyword co-occurrence



Furthermore, we did a co-citation analysis. As suggested by Leung et al. (2017) and Chang, Huang, and Lin (2015), complementing keyword co-occurrence with a co-citation analysis may help obtain a better understanding of the development and intellectual structure of a research field (Tandon et al., 2021). Co-citation analysis is a bibliometric method that was first proposed by Small (1973) and "allows quantifying the cocitation relationship between documents" (Shiau, Dwivedi, & Yang, 2017, p. 391). By creating a co-citation network, three blocks of authors could be identified, mostly assigned to different research domains: computer scientists, economists, and representatives from the field of information systems (Figure 9).

Figure 9. Co-citation networks



Of course, this cannot be generalized. However, it is remarkable that the network around Otto (on the right-hand side), i.e., the representatives of the decision rights camp, are often quoted together with other representatives of this camp. This network mainly investigated topics including (but not limited to) data stewardship, data ecosystems, accountabilities, data governance roles, data quality management, and organizational data governance aspects in general. The author network on the left-hand side, which is also the largest network, is relatively strongly connected to the technical-oriented data governance network, published in more technical outlets (mostly outlets receiving contributions from the Institute of Electrical and Electronics Engineers (IEEE)). Authors who belong to

this network basically dealt mostly with the topics of data protection and security, cloud data management, big data, the Internet of Things, device clustering, and the trustworthiness of data systems. The network at the top left-hand side focused on business research. This network is not strongly connected to the other networks and appears to have less relevance in the mutual development of data governance-related issues. However, this network deals with related research domains like business intelligence, data intelligence centers, and legal and compliance research. The closely linked networks are strongly connected, confirming that authors from computer engineering and information systems (and related disciplines) work closely together to further develop the whole research field dealing with humans, organizations, and IT to consider all facets of DGR.

4.3. Performance analysis

In this last subsection, we present the results of our performance analysis. A performance analysis aims to analyze and measure the number of publications or citations of the dataset's documents (Forliano et al., 2021). These can be sorted, among others, by journals and outlets, authors, or the author's affiliations (Massaro, Dumay, & Guthrie, 2016; Thelwall, 2008).

First, we present an overview of the most relevant sources for publications dealing with data governance. Due to a deeper analysis, we could confirm that most publications appear in conference proceedings or conference-related outlets (Table 2). Interestingly, practice-related journals, such as *IBM Data Management Magazine,* are also among the top outlets for data governance. This confirms the practical relevance of the research field, which continues to this day. Inductive contributions that extend existing theories or establish new theories are non-existent. What is striking in this context is the broad thematic spread of the outlets, which ultimately published many data governance papers. These range from technical outlets (*IEEE Access*) to strongly topic-related journals (*Sustainability*) to proceedings of a specific data-related conference (Proceedings of the 22nd MIT International Conference on Information Quality ICIQ).

Table 2. Most ranked outlets of data governance papers

Rank	*Source*	*No.*
1	Lecture notes in computer science	36
2	*ACM International Conference Proceedings Series*	18
3	*IBM Data Management Magazine*	16
4	*Journal of the American Medical Informatics Association*	14
5	CEUR Workshop Proceedings	12
6	Lecture notes in business information processing	11
7	*Advances in Intelligent Systems and Computing*	10
8	*Communications in Computer and Information Science*	10
9	*Procedia Computer Science*	7
10	*Sustainability*	7
11	*International Journal of Population Data Science*	6
12	*Yearbook of Medical Informatics*	6
13	*Big Data & Society*	5
14	*IEEE Access*	5
15	*IFIP Advances in Information and Communication Technology*	5
16	*International Journal of Information Management*	5
17	*Journal of Decision Systems*	5
18	*Journal of Medical Internet Research*	5
19	Proceedings of the 22nd MIT International Conference on Information Quality ICIQ	5
20	*BMC Medical Informatics and Decision Making*	4
21	*BMJ Open*	4
22	*Computer Fraud & Security*	4
23	*Internet Policy Review*	4
24	*Journal of Direct Data and Digital Marketing Practice*	4
25	*Journal of Enterprise Information Management*	4

Another relevant analysis of this study is listing the most cited articles of the entire study period (Table 3). The goal of citation analysis is to quantify the acceptance of publications to evaluate their acceptance (Ding & Cronin, 2011; Xue, Wang, & Yang, 2018). Citation analysis can therefore help measure the impact and relevance of articles and highlight which publications are the most relevant in a specific academic field. Almost all of the articles that have been repeatedly cited in the introductory chapters of specific data governance papers in recent years can also be found here. For instance, the data governance definition of Weber et al. (2009) or the fields of action within data governance established by Khatri and Brown (2010) are cited in most data governance articles, which are now understood as absolutely fundamental articles. Apart from Weber/Wende, Otto, Khatri, and Brown, and related authors, no familiar researcher from the DGR community can be located in this table. This finding verifies that DGR does not yet have the impact factor that generalist disciplines do. For example, the article by Tallis et al. (2010) addresses the effective management of ecosystems. This exemplary paper is cited almost as frequently as data governance-specific articles by Otto. Furthermore, all subdomains of DGR that we also found in the word clouds, such as big data, corporate governance, or business analytics, are represented.

Table 3. Most-cited articles 2005–2020

Rank	*Authors*	*Title*	*Citations*
01	Khatri and Brown (2010)	"Designing data governance"	252
02	Tallis et al. (2010)	"The many faces of ecosystem-based management: Making the process work today in real places"	195
03	Weber et al. (2009)	"One size does not fit all — A contingency approach to data governance"	151
04	Rein and Memmert (2016)	"Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science"	149
05	Chute, Beck, Fisk, and Mohr (2010)	"The enterprise data trust at Mayo Clinic: A semantically integrated warehouse of biomedical data"	128
06	Zwitter (2014)	"Big data ethics"	123
07	Tallon (2013)	"Corporate governance of big data: Perspectives on value, risk, and cost"	118
08	Brewster, Roussaki, Kalatzis, Doolin, and Ellis (2017)	"IoT in agriculture: Designing a Europe-wide large-scale pilot"	102
09	Fernandes et al. (2013)	"Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records"	95
10	Phillips-Wren, Iyer, Kulkarni, and Ariyachandra (2015)	"Business analytics in the context of big data: A roadmap for research"	85
11	Spencer et al. (2016)	"Patient perspectives on sharing anonymized personal health data using a digital system for dynamic consent and research feedback: A qualitative study"	77
12	Alreshidi, Mourshed, and Rezgui (2017)	"Factors for effective BIM governance"	69
13	Gökalp, Şener, and Eren (2017)	"Development of an assessment model for Industry 4.0: Industry 4.0-MM"	68
14	Rosenbaum (2010)	"Data governance and stewardship: Designing data stewardship entities and advancing data access"	66
15	Wende (2007)	"A model for data governance — Organising accountabilities for data quality management"	60
16	Hripcsak et al. (2014)	"Health data use, stewardship, and governance: Ongoing gaps and challenges: A report from AMIA's 2012 health policy meeting"	55
17	Otto (2011b)	"Organizing data governance: Findings from the telecommunications industry and consequences for large service providers"	55
18	O'Leary (2014)	"Embedding AI and crowdsourcing in the big data lake"	53
19	Arnaboldi, Busco, and Cuganesan (2017)	"Accounting, accountability, social media and big data: Revolution or hype?"	50
20	Vayena and Blasimme (2017)	"Biomedical big data: New models of control over access, use and governance"	44

We also investigated the authors with the most local impact within the field of data governance (Table 4). We sorted the list based on the number of total citations. To obtain a more detailed view of the author's impact, we included the h, g, and m-index. Due to its easy interpretability, the h-index is generally accepted and one of the most prominent indicators for scientific impact (Forliano et al., 2021). The h-index was first proposed by Hirsch (2005) and "indicates the number of publications that have received at least h citations" (Forliano et al., 2021, p. 3). According to Vanclay (2007, p. 1550), the "robustness to perturbations in the tails of the publication-citations distribution, and the ease of verifying" make the h-index a suitable index for measuring scientific impact.

Table 4. Most-cited authors

Rank	*Author*	*TC*	*Publications*	*H-index*	*M-index*
01	B. Otto	314	9	7	0.467
02	C. V. Brown	252	1	1	0.083
03	V. Khatri	252	1	1	0.083
04	D. L. Fluharty	195	1	1	0.083
05	B. S. Halpern	195	1	1	0.083
06	S. E. Lester	195	1	1	0.083
07	P. S. Levin	195	1	1	0.083
08	K. L. McLeod	195	1	1	0.083
09	M. Ruckelshaus	195	1	1	0.083
10	H. Tallis	195	1	1	0.083
11	C. G. Chute	183	2	2	0.167
12	H. Sterle	151	1	1	0.077
13	K. Weber	151	1	1	0.077
14	D. Memmert	149	1	1	0.167
15	R. Rein	149	1	1	0.167
16	S. A. Beck	128	1	1	0.083
17	T. B. Fisk	128	1	1	0.083
18	D. N. Mohr	128	1	1	0.083
19	A. Zwitter	123	1	1	0.125
20	P. P. Tallon	118	1	1	0.111

Although the h-index aims to measure the author's impact objectively, Kelly and Jennions (2006) have noted concerns regarding the suitability of the h-index to compare authors from different disciplines or seniority levels. To help overcome this possible flaw, we followed Forliano et al.'s (2021) approach and included the m-index. The m-index aims to weigh a scholar's productivity with the length of his or her career. The scholar's productivity is defined as the prior calculated h-index divided by the difference between an author's first and last publication (Hirsch, 2007). Furthermore, it is also noticeable that important community authors (Weber, Al-Ruithe) do not appear in the table.

The places are occupied by few very well-known papers and distort the first impression of this investigation's results. This is also emphasized by Table 4, which shows that generalist articles receive more attention than articles often cited in the DGR community. The supplement for these findings is provided in Table 5. Examining the study's absolute publication strength in this context, it becomes apparent that well-known researchers from the data governance community are returning. The most cited authors form the core of the research to date and will also, we assume, define the research picture of DGR in the coming years. This is supported by top researchers in the data governance community, such as Otto and Al-Ruithe, having already contributed several articles on data governance and future topics like cloud computing and data ecosystems.

Table 5. Most-productive authors 2016–2020

Rank	*Author*	*Publications*	*Rank*	*Author*	*Publications*
01	M. Al-Ruithe	10	11	J. Attard	4
02	E. Benkhelifa	10	12	R. Brennan	4
03	B. Otto	9	13	M. Daly	4
04	P. Brous	7	14	K. Hameed	4
05	M. Janssen	6	15	F. Haneem	4
06	S. U. Lee	6	16	S. Heys	4
07	M. Felici	5	17	N. Kama	4
08	R. Jeffery	5	18	X. Liu	4
09	L. Zhu	5	19	S. Person	4
10	I. Alhassan	4	20	C. Raymundo	4

5. CONCLUSION

This paper aimed to conduct a quantitative, bibliometric analysis of the research field of data governance. We sought to analyze the last 16 years of the research field, as the first emergence and then permanent evolution of DGR contributions appeared in 2005. We present crucial contributions as well as authors, keywords used and their context, and linkages among relevant authors. We formed topic clusters and commented on the thematic change of the research field during the last years.

As shown in the course of our investigation, data governance is understood as a further development of IT governance, which has outgrown the frames of the business research-oriented corporate governance approach. This led to the birth of the data governance research field. The increasingly popular topics of data quality and data management, in general, gave the first upturn to the research field. The aging of the research field then inevitably led to the diversification of the topics covered. Areas like big data, cloud computing, and data privacy, especially related to the GDPR or other regulations (Voss, 2019; Voss & Bouthinon-Dumas, 2020), attracted increasingly more attention and provided a significant boost to the field. However, this diversification also caused a blurring of the previous development of data governance, which was specifically focused on data quality.

To deal with this article's title, "Already grown-up or still in puberty?": No, data governance is still in puberty. On the one hand, this is so since the research field has not yet defined itself. The data governance term is still not precisely delineated from similar terms. Data governance is sometimes treated synonymously with data management or information governance. Repeated contributions attempt to distinguish various terms in a meaningful manner (Alhassan, Sammon, & Daly, 2016; Merkus, Helms, & Kusters, 2019), but the application of these separations remains absent. Future contributions have to ensure that the research field of data governance is clearly distinguished from other research domains.

Moreover, data governance is still viewed and defined in many different manners. The management-oriented camp represented by, for example, Otto (Otto, 2011a, 2012, 2013), has pursued the further development of the handling of data assets based on decision rights for years. The technical-oriented camp represented by, for example, Lee (Lee, Zhu, Jeffery, & Bui, 2018, 2019), focuses on more technical implementation proposals of data governance, defined by configuration approaches in technical environments. Besides, researchers, such as Al-Ruithe (Al-Ruithe et al., 2016; Al-Ruithe & Benkhelifa, 2017, 2018), also cross both camps, which completes the heterogeneous picture. Therefore, it is quite noticeable that this field of research had not completed its development, and its application was then investigated in specific domains. The core concept of data governance has never finally evolved but has been developed by different authors in the application. Many attempts to interpret data governance therefore exist today. The authors have made an effort to present this clearly in this paper.

For the next few years, however, the most directional and crucial research stream deals with inter-organizational data governance, also referred to as data ecosystem research (Lis & Otto, 2021; Oliveira, Barros Lima, & Lóscio, 2019). The entry of companies into networks now occupies researchers, with governance approaches for inter-organizational formations to assist corporate practice and government institutions in entering such ecosystems in a manner that is data value-oriented and compliant with data protection. The initial contributions in recent years (Calzada & Almirall,

2020; Lee et al., 2019; Lis & Otto, 2020; De Prieëlle et al., 2020) provide an excellent foundation for further developing this research stream. By increasing the breadth of this stream, the DGR can succeed in gaining even more relevance.

The exponential increase in the amount of data that needs to be managed in a value-oriented manner, the emergence of relevant digital transformation technologies such as blockchain, and the ever-increasing collaboration of companies in inter-organizational networks continue to drive the research field into a decentralized and non-uniform development. The DGR field developed itself. Future contributors should therefore try to create stronger synergies and pull them together. On a positive note, DGR has gained so much relevance that it is now part of current special issues in journals or forms its tracks at relevant research conferences. This provides the opportunity to network more closely to steer the entire research field in managed directions.

The authors' subjectivity in assigning keywords may skew the results, which is one of our study's limitations. Certain authors may utilize keywords incorrectly, resulting in a keyword that does not accurately reflect the substance of their research. Furthermore, the general habit of placing keywords in the field is another limitation and impacts how the results are presented. Moreover, the meaning of keywords in articles might change from one context to the next, and it can evolve, even if it is impossible to determine how close the new keywords are to the originals. Another drawback of our study is that our analytical technique is retrospective rather than prospective. The investigation focuses on the most heavily referenced publications and utilizes a threshold, among other things. As a result, the quantity of citations is proportional to the publication's age. Based on widely cited sources, our samples could generate a bias in favor of earlier research, which is also a limitation of this study.

Finally, we suppose DGR to define itself more strongly thematically, on the one hand, and to define a unique selling point for itself while simultaneously exploiting the potential of burgeoning research fields, such as big data, cloud computing, or data ecosystems, for itself. In that case, we assume that in a few years, DGR will succeed in leaving puberty behind — and then data governance can be called a research field that has come of age.

REFERENCES

1. Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management, 49,* 424–438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008
2. Akerlof, G. A. (1978). The market for "lemons": Quality uncertainty and the market mechanism. In P. Diamond, & M. Rothschild (Eds.), *Uncertainty in economics* (pp. 235–251). https://doi.org/10.1016/B978-0-12-214850-7.50022-X
3. Al-Badi, A., Tarhini, A., & Khan, A. I. (2018). Exploring big data governance frameworks. *Procedia Computer Science, 141,* 271–277. https://doi.org/10.1016/j.procs.2018.10.181
4. Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: An analysis of the literature. *Journal of Decision Systems, 25*(Sup. 1), 64–75. https://doi.org/10.1080/12460125.2016.1187397
5. Alreshidi, E., Mourshed, M., & Rezgui, Y. (2017). Factors for effective BIM governance. *Journal of Building Engineering, 10,* 89–101. https://doi.org/10.1016/j.jobe.2017.02.006
6. Al-Ruithe, M., & Benkhelifa, E. (2017). A conceptual framework for cloud data governance-driven decision making. In *2017 International Conference on the Frontiers and Advances in Data Science, FADS 2017* (pp. 1–16). https://doi.org/10.1109/FADS.2017.8253182
7. Al-Ruithe, M., & Benkhelifa, E. (2018). Cloud data governance in-light of the Saudi vision 2030 for digital transformation. *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*. https://doi.org/10.1109/AICCSA.2017.217
8. Al-Ruithe, M., Mthunzi, S., & Benkhelifa, E. (2016). Data governance for security in IoT & cloud converged environments. *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*. https://doi.org/10.1109/AICCSA.2016.7945737
9. Amadori, A., Altendeitering, M., & Otto, B. (2020). Challenges of data management in Industry 4.0: A single case study of the material retrieval process. In W. Abramowicz, & G. Klein (Eds.), *International Conference on Business Information Systems* (pp. 379–390). https://doi.org/10.1007/978-3-030-53337-3_28
10. Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007
11. Arnaboldi, M., Busco, C., & Cuganesan, S. (2017). Accounting, accountability, social media and big data: Revolution or hype? *Accounting, Auditing and Accountability Journal, 30*(4), 762–776. https://doi.org/10.1108/AAAJ-03-2017-2880
12. Avery, A. A., & Cheek, K. (2015). Analytics governance: Towards a definition and framework. Paper presented at the *2015 Americas Conference on Information Systems, AMCIS 2015*. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84963620207&partnerID=40&md5=db98566be42906a747c4f1f91a59e07e
13. Banker, R. D., & Kauffman, R. J. (2004). The evolution of research on information systems: A fiftieth-year survey of the literature in "Management Science". *Management Science, 50*(3), 281–298. https://doi.org/10.1287/mnsc.1040.0206
14. Begg, C., & Caira, T. (2009). Data governance in practice: The SME quandary reflections on the reality of data governance in the small to medium enterprise (SME) sector. Paper presented at the *4th International Conference on Information Warfare and Security, ICIW 2009*. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84893111638&partnerID=40&md5=614725064adeb72cae2c1fbec184acb3
15. Bekiaris, M., & Papanastasiou, P. (2021). A bibliometric analysis of family business: Insights from interdisciplinary studies. In S. Hundal, A. Kostyuk, & D. Govorun (Eds.), *Corporate governance: A search for emerging trends in the pandemic times* (pp. 59–63). https://doi.org/10.22495/cgsetpt10

16. Betzu, M., Coinu, G., & Demuro, G. (2019). Gobernanza de los macrodatos y democracia representative [Big data governance and representative democracy]. *Revista de Derecho Politico, 1*(106), 253–264. https://doi.org/10.5944/rdp.106.2019.26156
17. Brewster, C., Roussaki, I., Kalatzis, N., Doolin, K., & Ellis, K. (2017). Iot in agriculture: Designing a Europe-wide large-scale pilot. *IEEE Communications Magazine, 55*(9), 26–33. https://doi.org/10.1109/MCOM.2017.1600528
18. Broadus, R. N. (1987). Toward a definition of "bibliometrics". *Scientometrics, 12*(5–6), 373–379. https://doi.org/10.1007/BF02016680
19. Budler, M., Župič, I., & Trkman, P. (2021). The development of business model research: A bibliometric review. *Journal of Business Research, 135,* 480–495. https://doi.org/10.1016/j.jbusres.2021.06.045
20. Burritt, R. L., & Welch, S. (1997). Accountability for environmental performance of the Australian Commonwealth public sector. *Accounting, Auditing & Accountability Journal, 10*(4), 532–561. https://doi.org/10.1108/09513579710367494
21. Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics, 22*(1), 155–205. https://doi.org/10.1007/BF02019280
22. 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
23. Cao, Q. H., Khan, I., Farahbakhsh, R., Madhusudan, G., Lee, G. M., & Crespi, N. (2016). A trust model for data sharing in smart cities. In *IEEE International Conference on Communications (ICC)* (pp. 1–7). https://doi.org/10.1109/ICC.2016.7510834
24. Caviggioli, F., & Ughetto, E. (2019). A bibliometric analysis of the research dealing with the impact of additive manufacturing on industry, business and society. *International Journal of Production Economics, 208,* 254–268. https://doi.org/10.1016/j.ijpe.2018.11.022
25. Chang, Y.-W., Huang, M.-H., & Lin, C.-W. (2015). Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses. *Scientometrics, 105*(3), 2071–2087. https://doi.org/10.1007/s11192-015-1762-8
26. Cheong, L. K., & Chang, V. (2007). The need for data governance: A case study. In *ACIS 2007 Proceedings of the 18th Australasian Conference on Information Systems.* Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84869136537&partnerID=40&md5=8d9b3dafa8d3bcdc15f8506b30948c81
27. Chute, C. G., Beck, S. A., Fisk, T. B., & Mohr, D. N. (2010). The enterprise data trust at Mayo Clinic: A semantically integrated warehouse of biomedical data. *Journal of the American Medical Informatics Association, 17*(2), 131–135. https://doi.org/10.1136/jamia.2009.002691
28. Clement, D., Guetari, S. B. H., & Laboisse, B. (2010). Data quality as a key success factor for migration projects. *Proceedings of the 2010 International Conference on Information Quality, ICIQ 2010.* Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84871570429&partnerID=40&md5=84f72d4c7f2b9aa66dce196303e27a8d
29. Danvila-del-Valle, I., Estévez-Mendoza, C., & Lara, F. J. (2019). Human resources training: A bibliometric analysis. *Journal of Business Research, 101,* 627–636. https://doi.org/10.1016/j.jbusres.2019.02.026
30. De Prieëlle, F., De Reuver, M., & Rezaei, J. (2020). The role of ecosystem data governance in adoption of data platforms by Internet-of-Things data providers: Case of Dutch horticulture industry. *IEEE Transactions on Engineering Management*. https://doi.org/10.1109/TEM.2020.2966024
31. De Haes, S., & Van Grembergen, W. (2008). Practices in IT governance and business/IT alignment. *Information Systems Control Journal, 2*. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.609.950&rep=rep1&type=pdf
32. Derviş, H. (2019). Bibliometric analysis using Bibliometrix an R Package. *Journal of Scientometric Research, 8*(3), 156–160. https://doi.org/10.5530/jscires.8.3.32
33. Ding, Y., & Cronin, B. (2011). Popular and/or prestigious? Measures of scholarly esteem. *Information Processing & Management, 47*(1), 80–96. https://doi.org/10.1016/j.ipm.2010.01.002
34. Donthu, N., Kumar, S., & Pattnaik, D. (2020). Forty-five years of Journal of Business Research: A bibliometric analysis. *Journal of Business Research, 109,* 1–14. https://doi.org/10.1016/j.jbusres.2019.10.039
35. Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review, 14*(1), 57–74. https://doi.org/10.5465/amr.1989.4279003
36. Elango, B., & Rajendran, P. (2012). Authorship trends and collaboration pattern in the marine sciences literature: A scientometric study. *International Journal of Information Dissemination and Technology, 2*(3), 166–169. Retrieved from https://www.researchgate.net/publication/232763775_Authorship_Trends_and_Collaboration_Pattern_in_the_Marine_Sciences_Literature_A_Scientometric_Study
37. Ellegaard, O., & Wallin, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics, 105*(3), 1809–1831. https://doi.org/10.1007/s11192-015-1645-z
38. Eulerich, M., Haustein, S., Zipfel, S., & van Uum, C. (2013). The publication landscape of German corporate governance research: A bibliometric analysis. *Corporate Ownership and Control, 10*(2–4), 661–673. https://doi.org/10.22495/cocv10i2c4art1
39. Even, A., & Shankaranarayanan, G. (2007). Utility-driven assessment of data quality. *ACM SIGMIS Database: The Database for Advances in Information Systems, 38*(2), 75–93. https://doi.org/10.1145/1240616.1240623
40. Fahimnia, B., Sarkis, J., & Davarzani, H. (2015). Green supply chain management: A review and bibliometric analysis. *International Journal of Production Economics, 162,* 101–114. https://doi.org/10.1016/j.ijpe.2015.01.003
41. Fama, E. F., & Jensen, M. C. (1983). Separation of ownership and control. *The Journal of Law & Economics, 26*(2), 301–325. https://doi.org/10.1086/467037
42. Felici, M., Koulouris, T., & Pearson, S. (2013). Accountability for data governance in cloud ecosystems. *Proceedings of the International Conference on Cloud Computing Technology and Science* (pp. 327–332). https://doi.org/10.1109/CloudCom.2013.157
43. Fernandes, A. C., Cloete, D., Broadbent, M. T. M., Hayes, R. D., Chang, C.-K., Jackson, R. G.,…Callard, F. (2013). Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records. *BMC Medical Informatics and Decision Making, 13*(1), 71. https://doi.org/10.1186/1472-6947-13-71

44. Forliano, C., De Bernardi, P., & Yahiaoui, D. (2021). Entrepreneurial universities: A bibliometric analysis within the business and management domains. *Technological Forecasting and Social Change, 165,* 120522. https://doi.org/10.1016/j.techfore.2020.120522
45. Gillenson, M. L. (1985). Trends in data administration. *MIS Quarterly, 9*(4), 317–325. https://doi.org/10.2307/249232
46. Gökalp, E., Şener, U., & Eren, P. E. (2017). Development of an assessment model for Industry 4.0: Industry 4.0-MM. In A. Mas, A. Mesquida, R. O'Connor, T. Rout, & A. Dorling (Eds.), *Software process improvement and capability determination* (SPICE 2017: Communications in Computer and Information Science, Vol. 770, pp. 128–142). https://doi.org/10.1007/978-3-319-67383-7_10
47. Griffin, J. (2005). Data governance: A strategy for success. *Information Management, 15*(6), 49.
48. Grossman, S. J., & Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy, 94*(4), 691–719. https://doi.org/10.1086/261404
49. Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*(46), 16569–16572. https://doi.org/10.1073/pnas.0507655102
50. Hirsch, J. E. (2007). Does the H index have predictive power? *Proceedings of the National Academy of Sciences of the United States of America, 104*(49), 19193–19198. https://doi.org/10.1073/pnas.0707962104
51. Holve, E., Segal, C., Lopez, M. H., Rein, A., & Johnson, B. H. (2012). The electronic data methods (EDM) forum for comparative effectiveness research (CER). *Medical Care, 50*(Suppl. 1), S7–S10. https://doi.org/10.1097/MLR.0b013e318257a66b
52. Horne, N. W. (1995). Information as an asset — The board agenda. *Computer Audit Update, 1995*(9), 5–11. https://doi.org/10.1016/0960-2593(95)90246-5
53. Hripcsak, G., Bloomrosen, M., FlatelyBrennan, P., Chute, C. G., Cimino, J., Detmer, D. E.,…Wilcox, A. B. (2014). Health data use, stewardship, and governance: Ongoing gaps and challenges: A report from AMIA's 2012 health policy meeting. *Journal of the American Medical Informatics Association, 21*(2), 204–211. https://doi.org/10.1136/amiajnl-2013-002117
54. James, R., Tsosie, R., Sahota, P., Parker, M., Dillard, D., Sylvester, I.,…Olsen, P. (2014). Exploring pathways to trust: A tribal perspective on data sharing. *Genetics in Medicine, 16*(11), 820–826. https://doi.org/10.1038/gim.2014.47
55. Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3*(4), 305–360. https://doi.org/10.1016/0304-405X(76)90026-X
56. Kelly, C. D., & Jennions, M. D. (2006). The h index and career assessment by numbers. *Trends in Ecology & Evolution, 21*(4), 167–170. https://doi.org/10.1016/j.tree.2006.01.005
57. Khanra, S., Dhir, A., Parida, V., & Kohtamäki, M. (2021). Servitization research: A review and bibliometric analysis of past achievements and future promises. *Journal of Business Research, 131,* 151–166. https://doi.org/10.1016/j.jbusres.2021.03.056
58. Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM, 53*(1), 148–152. https://doi.org/10.1145/1629175.1629210
59. Kim, H. Y., & Cho, J.-S. (2018). Data governance framework for big data implementation with NPS Case Analysis in Korea. *Journal of Business and Retail Management Research, 12*(3). Advance online publication. https://doi.org/10.24052/JBRMR/V12IS03/ART-04
60. Koseoglu, M. A. (2016). Mapping the institutional collaboration network of strategic management research: 1980–2014. *Scientometrics, 109*(1), 203–226. https://doi.org/10.1007/s11192-016-1894-5
61. Krishen, A. S., Dwivedi, Y. K., Bindu, N., & Kumar, K. S. (2021). A broad overview of interactive digital marketing: A bibliometric network analysis. *Journal of Business Research, 131,* 183–195. https://doi.org/10.1016/j.jbusres.2021.03.061
62. Kulkarni, A. V., Aziz, B., Shams, I., & Busse, J. W. (2009). Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *JAMA, 302*(10), 1092–1096. https://doi.org/10.1001/jama.2009.1307
63. Lee, S. U., Zhu, L., Jeffery, R., & Bui T.X. (2018). Designing data governance in platform ecosystems. *Proceedings of the Annual Hawaii International Conference on System Sciences.* Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85097715317&partnerID=40&md5=5edf9d8a48a685c7a41ae2049c517f1b
64. Lee, S. U., Zhu, L., Jeffery, R., & Bui T.X. (2019). Data governance decisions for platform ecosystems. *Proceedings of the Annual Hawaii International Conference on System Sciences.* Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85108284003&partnerID=40&md5=70f34da3295da90434d6272ea11a2a75
65. Legner, C., Pentek, T., & Otto, B. (2020). Accumulating design knowledge with reference models: Insights from 12 years' research into data management. *Journal of the Association for Information Systems, 21*(3), 735–770. https://doi.org/10.17705/1jais.00618
66. Leung, X. Y., Sun, J., & Bai, B. (2017). Bibliometrics of social media research: A co-citation and co-word analysis. *International Journal of Hospitality Management, 66,* 35–45. https://doi.org/10.1016/j.ijhm.2017.06.012
67. Lis, D., & Otto, B. (2020). Data governance in data ecosystems — Insights from organizations. Paper presented at the *26th Americas Conference on Information Systems, AMCIS 2020.* Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85097716540&partnerID=40&md5=6d4a810cfdd82c48394b1c88730a2637
68. Lis, D., & Otto, B. (2021). Towards a taxonomy of ecosystem data governance. In *Proceedings of the 54th Hawaii International Conference on System Sciences.* https://doi.org/10.24251/HICSS.2021.733
69. Massaro, M., Dumay, J., & Guthrie, J. (2016). On the shoulders of giants: Undertaking a structured literature review in accounting. *Accounting, Auditing & Accountability Journal, 29*(5), 767–801. https://doi.org/10.1108/AAAJ-01-2015-1939
70. Merigó, J. M., Mas-Tur, A., Roig-Tierno, N., & Ribeiro-Soriano, D. (2015). A bibliometric overview of the Journal of Business Research between 1973 and 2014. *Journal of Business Research, 68*(12), 2645–2653. https://doi.org/10.1016/j.jbusres.2015.04.006
71. Merkus, J., Helms, R., & Kusters, R. (2019). Data governance and information governance: Set of definitions in relation to data and information as part of DIKW. In *Proceedings of the 21st International Conference on Enterprise Information Systems — Volume 2: ICEIS.* https://doi.org/10.5220/0007411901430154

72. Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *El Profesional De La Información, 29*(1). https://doi.org/10.3145/epi.2020.ene.03
73. Mulgan, R. (2000). 'Accountability': An ever-expanding concept? *Public Administration, 78*(3), 555–573. https://doi.org/10.1111/1467-9299.00218
74. O'Leary, D. E. (2014). Embedding AI and crowdsourcing in the big data lake. *IEEE Intelligent Systems, 29*(5), 70–73. https://doi.org/10.1109/MIS.2014.82
75. Oliveira, M. I. S., Barros Lima, G. D. F., & Lóscio, B. F. (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
76. Otto, B. (2011a). A morphology of the organisation of data governance. Paper presented at the *19th European Conference on Information Systems, ECIS 2011*. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84870632185&partnerID=40&md5=0d5dac8d1388f4469bdf84a9296c2595
77. Otto, B. (2011b). Organizing data governance: Findings from the telecommunications industry and consequences for large service providers. *Communications of the Association for Information Systems, 29*(1), 45–66. https://doi.org/10.17705/1cais.02903
78. Otto, B. (2011c). Data governance. *Business & Information Systems Engineering, 3*(4), 241–244. https://doi.org/10.1007/s12599-011-0162-8
79. Otto, B. (2012). Managing the business benefits of product data management: The case of Festo. *Journal of Enterprise Information Management, 25*(3), 272–297. https://doi.org/10.1108/17410391211224426
80. Otto, B. (2013). On the evolution of data governance in firms: The case of Johnson & Johnson consumer products North America. In S. Sadiq (Eds.), *Handbook of data quality: Research and practice* (pp. 93–118). https://doi.org/10.1007/978-3-642-36257-6_5
81. Paul, J., & Criado, A. R. (2020). The art of writing literature review: What do we know and what do we need to know? *International Business Review, 29*(4), 101717. https://doi.org/10.1016/j.ibusrev.2020.101717
82. Phillips-Wren, G., Iyer, L. S., Kulkarni, U., & Ariyachandra, T. (2015). Business analytics in the context of big data: A roadmap for research. *Communications of the Association for Information Systems, 37,* 448–472. https://doi.org/10.17705/1cais.03723
83. Power, E. M., & Trope, R. L. (2006). The 2006 survey of legal developments in data management, privacy, and information security: The continuing evolution of data governance. *Business Lawyer, 62*(1), 251–294. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-33847697280&partnerID=40&md5=41d1f67a8c1d27ab44e2f4995d93ca3f
84. Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation, 25*(4), 348–349.
85. Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science. *SpringerPlus, 5*(1), 1410. https://doi.org/10.1186/s40064-016-3108-2
86. Rosenbaum, S. (2010). Data governance and stewardship: Designing data stewardship entities and advancing data access. *Health Services Research, 45*(5p2), 1442–1455. https://doi.org/10.1111/j.1475-6773.2010.01140.x
87. Schreieck, M., Wiesche, M., & Krcmar, H. (2016). Design and governance of platform ecosystems — Key concepts and issues for future research. Retrieved from https://www.researchgate.net/publication/303924671_Design_and_Governance_of_Platform_Ecosystems_-_Key_Concepts_and_Issues_for_Future_Research
88. Schüritz, R. M., Seebacher, S., Satzger, G., & Schwarz, L. (Eds.). (2017). *Datatization as the next frontier of servitization — Understanding the challenges for transforming organizations.* Retrieved from https://www.researchgate.net/publication/319998754_Datatization_as_the_Next_Frontier_of_Servitization_-_Understanding_the_Challenges_for_Transforming_Organizations
89. Shiau, W.-L., Dwivedi, Y. K., & Yang, H. S. (2017). Co-citation and cluster analyses of extant literature on social networks. *International Journal of Information Management, 37*(5), 390–399. https://doi.org/10.1016/j.ijinfomgt.2017.04.007
90. Simon, H. A. (1951). A formal theory of the employment relationship. *Econometrica: Journal of the Econometric Society, 19*(3), 293–305. https://doi.org/10.2307/1906815
91. Skute, I., Zalewska-Kurek, K., Hatak, I., & de Weerd-Nederhof, P. (2019). Mapping the field: A bibliometric analysis of the literature on university — Industry collaborations. *The Journal of Technology Transfer, 44*(3), 916–947. https://doi.org/10.1007/s10961-017-9637-1
92. Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science and Technology, 24*(4), 265–269. https://doi.org/10.1002/asi.4630240406
93. Spencer, K., Sanders, C., Whitley, E. A., Lund, D., Kaye, J., & Dixon, W. G. (2016). Patient perspectives on sharing anonymized personal health data using a digital system for dynamic consent and research feedback: A qualitative study. *Journal of Medical Internet Research, 18*(4), e66. https://doi.org/10.2196/jmir.5011
94. Tallis, H., Levin, P. S., Ruckelshaus, M., Lester, S. E., McLeod, K. L., Fluharty, D. L., & Halpern, B. S. (2010). The many faces of ecosystem-based management: Making the process work today in real places. *Marine Policy, 34*(2), 340–348. https://doi.org/10.1016/j.marpol.2009.08.003
95. Tallon, P. P. (2013). Corporate governance of big data: Perspectives on value, risk, and cost. *Computer, 46*(6), 32–38. https://doi.org/10.1109/MC.2013.155
96. Tallon, P. P., Ramirez, R. V., & Short, J. E. (2013). The information artifact in IT governance toward a theory of information governance. *Journal of Management Information Systems, 30*(3), 141–178. https://doi.org/10.2753/MIS0742-1222300306
97. Tandon, A., Kaur, P., Mäntymäki, M., & Dhir, A. (2021). Blockchain applications in management: A bibliometric analysis and literature review. *Technological Forecasting and Social Change, 166,* 120649. https://doi.org/10.1016/j.techfore.2021.120649
98. Thelwall, M. (2008). Social networks, gender, and friending: An analysis of MySpace member profiles. *Journal of the American Society for Information Science and Technology, 59*(8), 1321–1330. https://doi.org/10.1002/asi.20835
99. Trope, R. L., & Power, E. M. (2005). Lessons in data governance: A survey of legal developments in data management, privacy and security. *Business Lawyer, 61*(1), 471–516. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-33646748917&partnerID=40&md5=18ad3703a66d1dd03bc43845983311df

100. Vallaster, C., Kraus, S., Merigó Lindahl, J. M., & Nielsen, A. (2019). Ethics and entrepreneurship: A bibliometric study and literature review. *Journal of Business Research, 99,* 226–237. https://doi.org/10.1016/j.jbusres.2019.02.050
101. van den Broek, T., & van Veenstra, A. F. (2015). Modes of governance in inter-organizational data collaborations. In *ECIS Completed Research Papers*. Retrieved from https://aisel.aisnet.org/ecis2015_cr/188
102. van Eck, N. J., & Waltman, L. (2010). Software survey: Vosviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3
103. Vanclay, J. K. (2007). On the robustness of the h-index. *Journal of the American Society for Information Science and Technology, 58*(10), 1547–1550. https://doi.org/10.1002/asi.20616
104. Vayena, E., & Blasimme, A. (2017). Biomedical big data: New models of control over access, use and governance. *Journal of Bioethical Inquiry, 14*(4), 501–513. https://doi.org/10.1007/s11673-017-9809-6
105. Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems, 28*(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003
106. Voss, W. G. (2019). Cross-border data flows, the GDPR, and data governance. *Washington International Law Journal, 29*(3), 485. https://digitalcommons.law.uw.edu/wilj/vol29/iss3/7
107. Voss, W. G., & Bouthinon-Dumas, H. (2020). EU general data protection regulation sanctions in theory and in practice. *Santa Clara High Technology Law Journal, 37*(1), 1–96. Retrieved from https://digitalcommons.law.scu.edu/chtlj/vol37/iss1/2
108. Weber, K., Otto, B., & Österle, H. (2009). One size does not fit all — A contingency approach to data governance. *Journal of Data and Information Quality, 1*(1), 1–27. https://doi.org/10.1145/1515693.1515696
109. Weill, P., & Ross, J. W. (2004). *IT governance on one page*. https://doi.org/10.2139/ssrn.664612
110. Wende, K. (2007). A model for data governance — Organising accountabilities for data quality management. *ACIS 2007 Proceedings of the 18th Australasian Conference on Information Systems*. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84869133918&partnerID=40&md5=30239e30b734e35add25b0a7b0c72c68
111. White, A., Newman, D., Logan, D., & Radcliffe, J. (2006). Mastering master data management. *Gartner*. Retrieved from https://www.gartner.com/en/documents/488207/mastering-master-data-management
112. Winter, J. S., & Davidson, E. (2019). Big data governance of personal health information and challenges to contextual integrity. *Information Society, 35*(1), 36–51. https://doi.org/10.1080/01972243.2018.1542648
113. Xue, X., Wang, L., & Yang, R. J. (2018). Exploring the science of resilience: Critical review and bibliometric analysis. *Natural Hazards, 90*(1), 477–510. https://doi.org/10.1007/s11069-017-3040-y
114. Zwitter, A. (2014). Big data ethics. *Big Data and Society, 1*(2). Advance online publication. https://doi.org/10.1177/2053951714559253