DOI: 10.47743/eufire-2024-1-25

RELEVANCE OF MASTER DATA MANAGEMENT AS PART OF DATA GOVERNANCE AND A CRITICAL FACTOR FOR CORPORATE SUCCESS: A SCIENTOMETRIC ANALYSIS

MATTHIAS SCHMUCK
Alexandru Ioan Cuza University of Iasi
Iasi, Romania
rsrch@bid-intelligence.com

Abstract
Master data management and data governance are critical in today's business world, as organizations increasingly rely on data to make informed decisions and gain competitive advantage. This article examines the relevance of master data management as an integral part of data governance for business success using scientometric methods. Through the systematic evaluation of specialist literature and research work, a comprehensive understanding of the importance of master data management and data governance is gained. The analysis identifies key concepts, trends and research gaps in the field of master data management and shows how effective management of master data strengthens data governance and thus has a significant impact on corporate success. These findings provide valuable impulses for companies to optimize their data strategies and gain a sustainable competitive advantage. The results also form the basis for further research based on design science research, supported by qualitative literature analysis and expert interviews in a real business environment, to construct a master data management artifact that is influenced by data governance experiences.
Keywords: *data management; data governance; quantitative analysis; scientometrics.*
JEL Classification*:* C80, O10, O25, O33.

1. INTRODUCTION

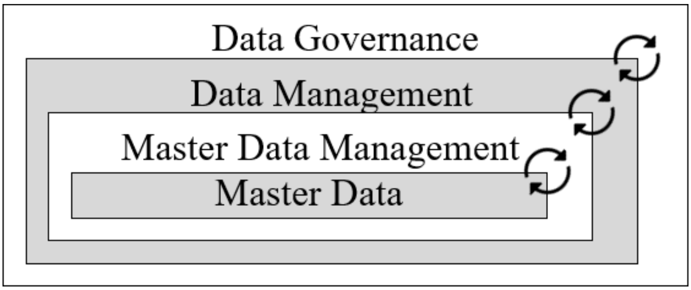
In an era characterized by an exponential increase in data (data is the new oil; The Economist, 2017), companies are faced with the challenge of managing this data effectively to gain valuable insights and increase their business success (BearingPoint, 2016; PWC, 2018). In this context, master data management (MDM) is becoming increasingly important as part of data governance (DG), as it forms the basis for efficient data management and use (Grosser, 2019; Mittelstand-Digital, 2022). Master data, as fundamental information about business units, products, customers and suppliers, is of crucial importance for a company's business processes and decision-making.

The integration of MDM into a comprehensive DG strategy is considered a critical factor for business success (Allen and Cervo, 2015). This integration not only ensures the quality, consistency and availability of data, but also enables companies to meet regulatory requirements, minimize risks and identify opportunities. In essence, it is about understanding the interplay between MDM and DG and their impact on business success.

For this study, the researcher conducted a scientometric analysis to investigate the current research and development in the field of MDM as part of DG. This type of study is, to the researcher's knowledge, the only one of its kind to date in this field of research and he was guided by the following questions: (RQ1) What is the current state of MDM research? (RQ2) What are the key scientometric factors? (RQ3) Which topics influence the research field? (RQ4) What are the connections to DG? Through the systematic evaluation of specialist literature and research papers, this study contributes to gaining insights into existing concepts and trends and to gaining important insights into the relevance and influence of MDM on corporate success.

2. THEORETICAL BACKGROUND

At this point, the central concepts will be introduced in a focused scope using a theoretical framework in Figure 1.



Source: contribution by the researcher

Figure 1. Theoretical framework

MASTER DATA (MD): MD are fundamental and relatively stable entities over time (e.g. customers, products, suppliers) that are of central importance for a company's business processes and decisions (Schmidt, 2010; Scheuch *et al.,* 2012). MD is characterized by its persistence and extensive use in various areas of the company. They serve as reference data for transactions, analyses, reports and other business activities. The quality and consistency of MD are crucial for the efficiency and effectiveness of business processes and decisions.

MASTER DATA MANAGEMENT (MDM): MD planning, management and control, also known as MDM as special data management (DM), includes processes, policies, technologies and standards that ensure MD is accurate, up-to-date, complete and consistent (Otto, 2009; Loshin, 2009; Scheuch *et al.,* 2012). MDM enables organizations to improve the quality of their MD, reduce maintenance costs and provide a solid information base for decision-making.

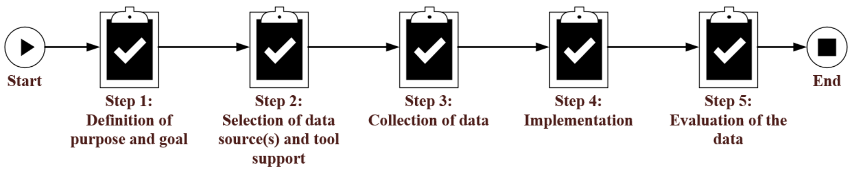
DATA GOVERNANCE (DG): DG is an organization-wide approach that formally orchestrates processes, people and technology in terms of a human-task-technology system, ensuring that data is managed in a way that meets business needs, ensures data integrity and quality, and ensures compliance with legal regulations and internal policies (Khatri and Brown, 2010; Otto, 2011; Lee *et al.,* 2018; Abraham *et al.,* 2019; Jagals *et al.,* 2021). The data types that are the control object of the DG include data in general and master data in particular.

SCIENTOMETRICS: The term scientometrics comes from the Russian word "naukometriya" and was first introduced in 1969 by Nalimov and Mulcjenko (1969) and used by Brindha and Murugesapandian (2016). For the present study, the researcher follows the definition of scientometrics as "The quantitative methods of the research on the development of science as an informational process" (Nalimov and Mulcjenko, 1971, p. 2). Scientometrics methods includes citation mapping, visualization, bibliographic linking, co-authorship network or co-word mapping. Researchers receive tool support with VOSviewer, Bibliometrix or CiteSpace.

3. METHODOLOGY AND MATERIAL

This study is based on existing research findings on the topic in question. The study is therefore qualitative in nature (literature analysis with the use of scientometric methods).

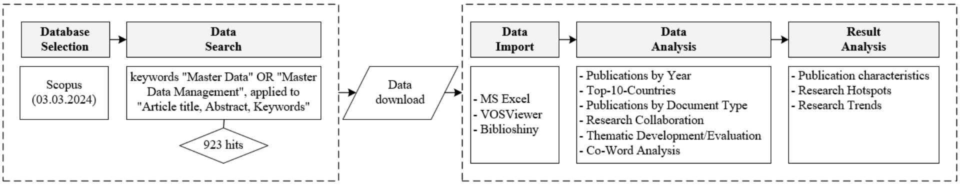
Figure 2 shows the analysis process in the context of this study. It begins with the definition of the purpose and objective of the study (step 1), followed by the selection of information sources and tool support (step 2), the collection of data relevant for evaluation (step 3) and the execution of the analysis (step 4). The study concludes with a discussion of the results of the analysis and the derivation of conclusions.



Source: contribution by the researcher adapted from Donthu *et al.* (2021)

Figure 2. Analytical process

For scientometric analysis (Figure 3), the researcher used the digital library Scopus (www.scopus.com), a popular data source for secondary data in research due to the volume of publications offered there (Popescul *et al.,* 2020; Jagals *et* *al.,* 2021; Necula and Păvăloaia, 2023). The date of the search was 03.03.2024 with the keywords "Master Data" OR "Master Data Management", applied to "Article title, Abstract, Keywords". Abbreviations such as "MD" or "MDM" were not used for the search, as these can also stand for other terms (e.g. metadata). No further restrictions were made at this point to obtain comprehensive results. The query resulted in 923 main hits, which were included in the scientometric analysis. Interestingly, there are almost four times as many articles on MDM in the field of informatics (number = 511) as in the field of economics (number = 117), which on the one hand reflects the multidisciplinary nature of MDM research and on the other hand shows that data (still) seems to be a technical topic, although data in general and MD in particular have a strong technical component.



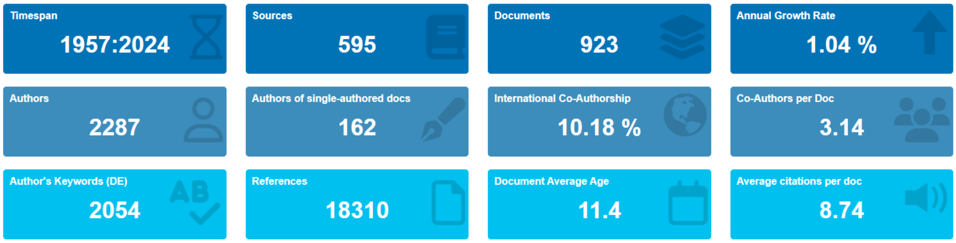
Source: contribution by the researcher

Figure 3. Strategy, material, analysis methodology

The results were exported in CSV/RIS format. The analysis was then performed in two ways: (a) the descriptive analysis was supported by Microsoft Excel (MS Excel), (b) the scientometric analysis was supported by VOSViewer (version: 1.6.20) and Biblioshiny.

4. ANALYSIS AND RESULTS

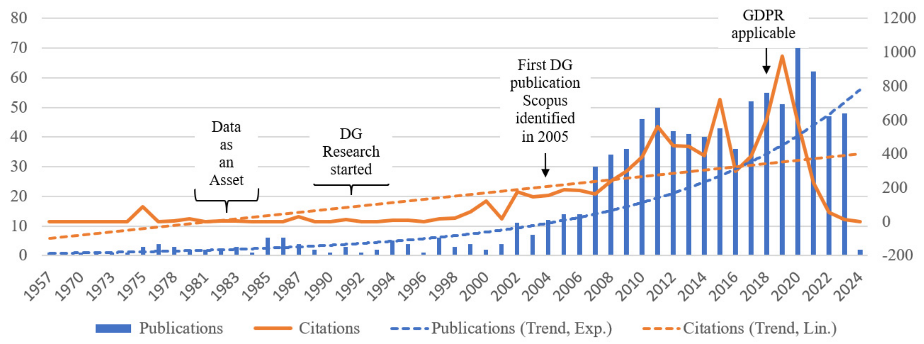
The sample (Figure 4) covers the period from 1957 to 2024 with a total of 923 documents in 595 sources (including books, articles, conference papers and reviews), with 2.287 Authors in total, an average age of the documents of 11.4 years (that seems to be old) and an annual increase of 1.04% (rather a marginal growth). A total of 2,054 keywords were declared in the documents for indexing and 18,310 references were cited (with an average of 8.74 citations per document). 2,287 different authors contributed to the research in MDM. Of the 923 documents, 162 were authored by single authors; this corresponds to a share of 17.5 %. The metric "international co-authorship", which has a value of 10.18% in the sample, reflects the strength of collaboration between researchers across national borders and is therefore a benchmark for international cooperation in research and science (OECD, 2019; Thakur *et al.,* 2011).



Source: contribution by the researcher with Scopus/Biblioshiny

Figure 4. Overview of the sample

Publication figures have risen continuously since the first publication in 1957 (Figure 5). This also applies to citations. The initial increase is moderate, but then increases noticeably from 2008 onwards. The reasons for this are (a) the growing awareness of data protection, (b) the increasing importance of data management in general (companies are becoming increasingly data-driven) and big data, and (c) the interest in analysing this data (e.g. 360-degree view of the customer). In recent times, cyber security, data security, mergers and acquisitions, compliance, risk management and the integration of corporate architectures are further drivers of this development Due to ongoing digitalization and the associated increase in the relevance of data in general and MD in particular, it can be assumed that the field of research will continue to enjoy the same level of attractiveness.



Source: contribution by the researcher with Scopus/MS Excel

Figure 5. Publications by year

TOP 10 COUNTRIES: With 604 publications, the TOP 10 countries (Table 1) account for 65.4% of the sample. The majority of publications are in Europe (271 publications or 44.9%). North America (147 publications or 24.3%) and Asia (186 publications or 30.8%) are roughly on a par.

Table 1. Top 10 countries

Country	Continent	Publications	Share
Germany	Europe	148	24.5%
United States	North America	147	24.3%
China	Asia	95	15.7%
United Kingdom	Europe	37	6.1%
Indonesia	Asien	35	5.8%
Switzerland	Europe	33	5.5%
India	Asia	31	5.1%
France	Europe	28	4.6%
Finland	Europe	25	4.1%
Japan	Asia	25	4.1%

Source: contribution by the researcher with Scopus/Biblioshiny

Table 2 shows the results by document type.

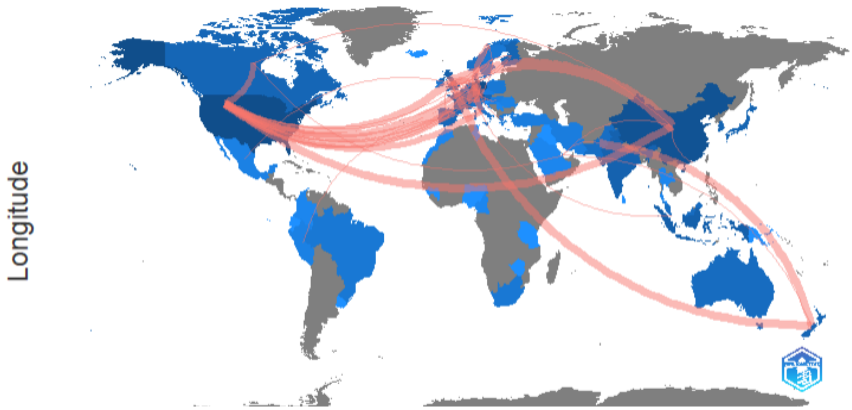
Table 2. Publications by document type

Document type	Publications	Share
Article	345	37.4%
Book	21	2.3%
Book Chapter	37	4.0%
Conference Paper	441	47.8%
Conference Review	32	3.5%
Data Paper	1	0.1%
Note	7	0.8%
Report	1	0.1%
Retracted	2	0.2%
Review	29	3.1%
Short Survey	7	0.8%
Total	923	100%

Source: contribution by the researcher with Scopus/MS Excel

With a share of 51.2%, conference objects are the largest group in the sample, followed by articles published in scientific journals with 37.4%. The other categories, including books and book chapters, make up a small proportion of the sample. For the researcher, this is an indication to concentrate on articles and conference contributions in the qualitative evaluation of previous publications (systematic literature analysis).

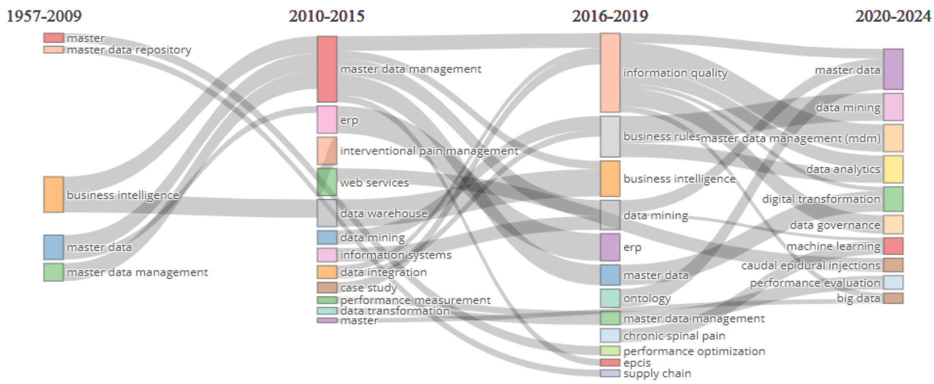
Figure 6 visualizes the researchers' work on a world map. The focus of research is clearly on industrialized countries, which seems understandable: the development of modern industries is heavily dependent on IT and therefore data as digitalization increases. However, all other countries are encouraged to get involved in MDM research.



Source: contribution by the researcher with Scopus/Biblioshiny

Figure 6. Research collaboration as a world map

Figure 7 shows the thematic development of the MDM research field in the period 1957-2024, divided into four sub-periods (1957-2008; 2009-2015; 2016-2019 and 2020-2024) as a so-called Sankey diagram (Aria and Cuccurullo, 2020).

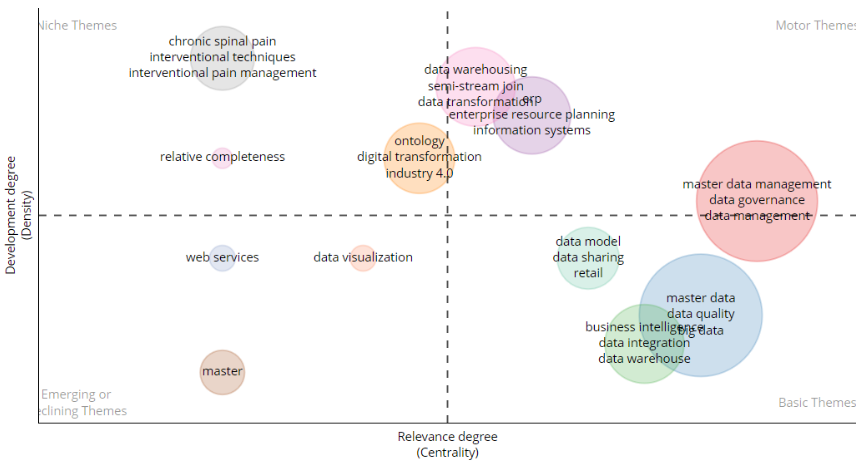


Source: contribution by the researcher with Scopus/Biblioshiny

Figure 7. Thematic development (Sankey diagram)

Each node represents a set of different topics over time, characterized by the keyword with the highest frequency, where the size of the node is proportional to the number of keywords. The flow between the nodes illustrates the direction of the topic development, the width of the edge is proportional to the index of inclusion between two related topics, leading to an increase in connections between the topics over time. The diagram confirms the correlation between advancing digitalization and the associated increase in the importance of data in operational organizations, especially MD, also in light of the corresponding research area DG. MDM is a driver over the entire observation period and can be seen as one of the "seeds" for DG.

To identify clusters and groups in the documents, the researcher identified the topics of engine, niche, base and decline. The results are shown as a thematic map in Figure 8. It classifies research topics according to their degree of centrality - as the extent of the relationship between different topics - and density - as a symbol of progress in the research field to assess importance and connectivity (Esfahani *et al.,* 2019; Mobin *et al.,* 2023). Recurring keywords in the map indicate clusters that form the basis for research in the quadrants (Cobo *et al.,* 2011). Biblioshiny supports this type of analysis and visualization of bibliographic data without researchers having to do their own coding.



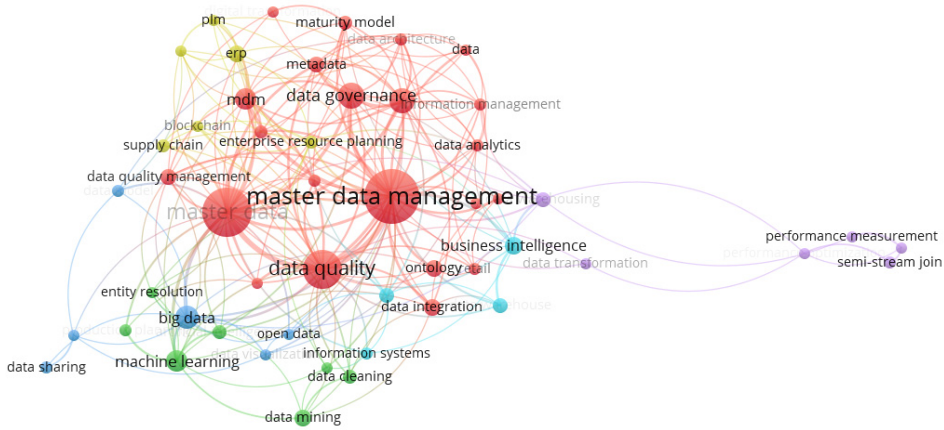
Source: contribution by the researcher with Scopus/Biblioshiny

Figure 8. Thematic evolution

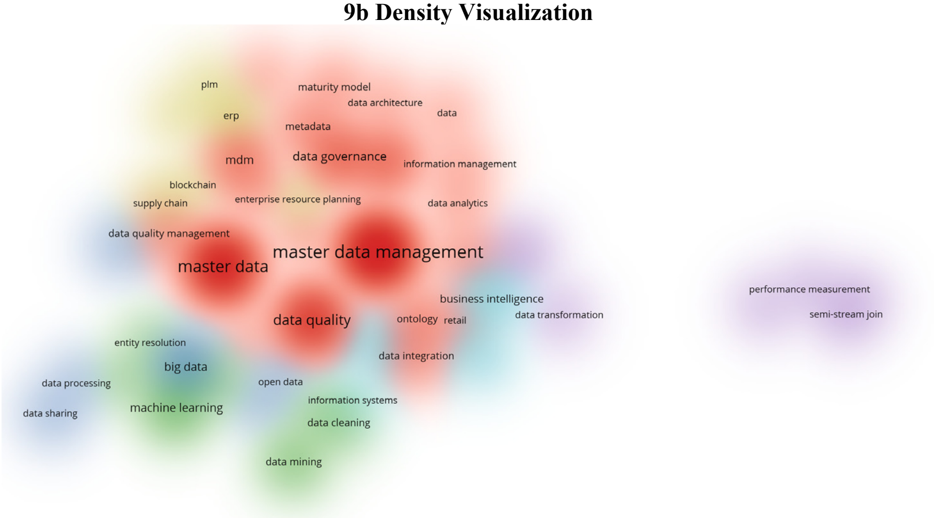
MDM, DG, classic DM, but also specific applications of information systems, such as enterprise resource planning (ERP) or data warehousing, are the driving topics in this research area (Motor Themes). The basis of these drivers are classic DM topics such as data model, data quality, data warehouse, business intelligence and data integration (basic topics). This seems logical, as 80% of the data in ERP systems is MD, e.g. for configuring the ERP, and MD is also qualitative information in analytical applications, according to which key figures are broken down. Digital transformation is still a niche topic but is generating additional pressure in the aforementioned engine topics, as companies will become increasingly data-driven as a result. In contrast, web services and data visualization appear to be declining topics.

In the next step, the co-word analysis was carried out (Figures 9 and 10).

9a Network Visualization

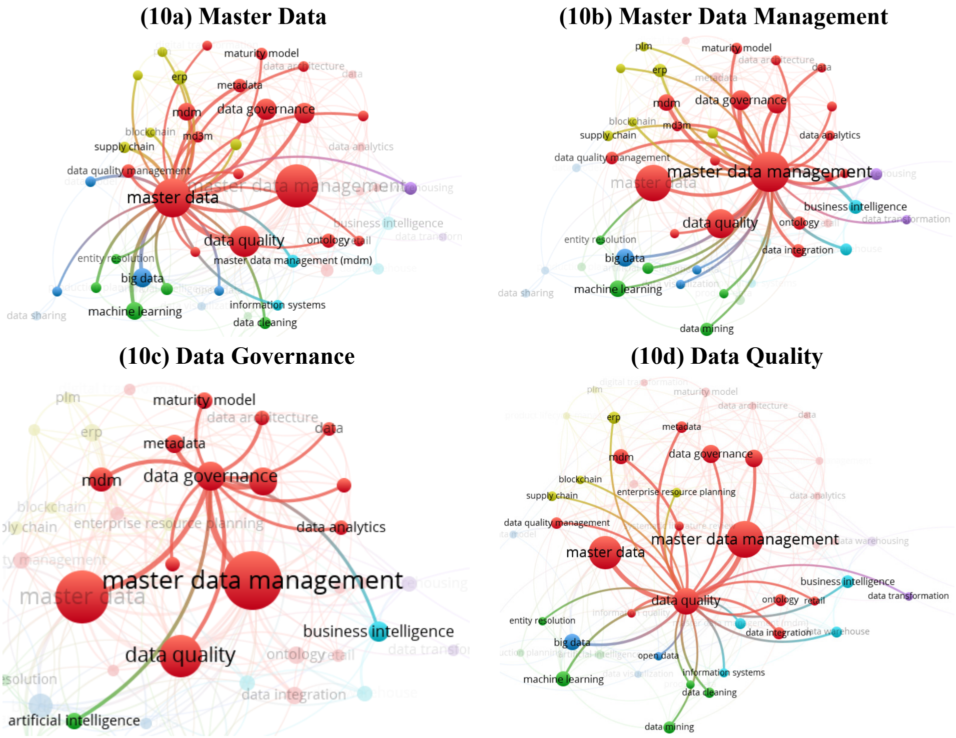


9b Density Visualization



Source: contribution by the researcher with Scopus/VOSviewer

Figure 9. Bibliographic linking Author keywords (all links)



Source: contribution by the researcher with Scopus/VOSviewer

Figure 10. Bibliographic linking of author keywords (details)

This involved checking how often certain words occur together in a selection of scientific publications. From this, term clusters can be formed. Topics at the center of the co-word network belong to the so-called "mainstream topics", while peripheral clusters can represent new or innovative topics in the research field. The connections between the clusters mean that keywords are mentioned together in sources. Noise in the data can occur due to ambiguities of words/homonyms or the occurrence of several words for one and the same term/synonyms (Havemann, 2009). For the co-word analysis, 2,056 author keywords were included. Only keywords with a minimum frequency of occurrence of 5 (threshold value) were included; 56 keywords met this restriction.

Figure 9a shows that MD, MDM and Data Quality (DQ) are the most important terms used to form clusters. DG and DQ are drivers for MDM. Importing the data also revealed a heterogeneous structure with a few clearly definable clusters (Figure 9b). The red cluster represents data management with topics such as data quality, metadata, maturity model or DG. The green cluster combines topics relating to modern forms of business analytics (e.g. data mining, machine learning). The yellow cluster refers to operational applications in which MD plays a prominent role, e.g. Product Livecycle Management (PLM) or ERP. The blue cluster is Big Data. However, peripheral topics such as performance measurement (purple) can also be identified.

Figure 10 shows the connections between the most important (top 4) identified topics. The figures show impressively that the topics are related to each other resp. influence each other: the connections between the keywords overlap and show significant relationships, represented by the strength of the connections.

5. SUMMARY AND OUTLOOK

The aim of this scientometric analysis was to understand the interaction between MDM and DG, including the impact on company success. To do this, the researcher examined 923 publications on master data management indexed in the Scopus online library over a period from 1957 to March 2024. The researcher addressed various research questions, including the current state of research, the key scientometric factors and the most important topics in the field of MDM research.

The results impressively show that MD and its management (MDM) make an important contribution to digital transformation and therefore to the success of the company. They influence the quality of operational and analytical applications. The published literature covers four central aspects: MD, MDM, DQ and DG. The terms "Master Data" and "Master Data Management" are the most common keywords in the same ratio, but DQ and DG are the drivers. Europe was the main contributor in this research area, followed by North America and Asia.

Future topics in the MDM research field include the establishment of a data culture, or rather a data-driven culture. It enables companies to innovate faster, continuously improve through real-time feedback and response and make smarter business decisions. Conversational technologies, including AI-powered voice assistants, chatbots or intelligent personal assistants, i.e. AI applications in general, act as disruptive forces in MDM because they force companies to keep trustworthy, well-organized and metadata-rich data at hand. The acceptance of these applications is heavily dependent on how fast, clean and efficient the company data is. Multi-domain data (including digital assets and geodata) will also trigger MDM. The challenge here is to intelligently network the domains. The share of cloud-based MDM solutions will increase as the need for application migrations and the associated consolidation of data sets grows. In this context, Gartner introduces the term "headless MDM" into the discussion (Gartner, 2021), in which the end user interfaces to the master data can be fully customized to support the end user's business processes in all MDM workflows, including administration. The topics offer very good starting points for future research.

However, the study also has some limitations. At first, only one library (Scopus) was used. An extension to other libraries, e.g. Web of Science or SpringerLink, would make sense. Furthermore, the text analysis was restricted to the keywords assigned by the authors. It would have to be checked whether the extension to all indexed keywords would yield more detailed findings, or whether the sample is already saturated with the chosen selection.

The next step for the researcher is to analyse the publications from a qualitative point of view. For this purpose, the 923 studies are further narrowed down in a suitable manner according to known techniques. The aim is to extract the success factors of MDM from these studies and to evaluate the factors and mechanisms that influence MDM. They serve as a starting point for the subsequent conception of an MDM success framework based on the Design Science Research (DSR) approach. Expert interviews in a real company context (German SME) will serve to evaluate the framework.

References

1) Abraham, R., Schneider, J. and vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management,* 49(2019), pp. 424-438.
2) Allen, M. and Cervo, D. (2015). *Multi-Domain Master Data Management: Advanced MDM and Data Governance in Practice*. Waltham: Morgan Kaufmann.
3) Aria, M. and Cuccurullo, C. (2017). Bibliometrix: An R-tool for Comprehensive Science Mapping Analysis. *Journal of Informetrics*, 11, pp. 959-975.
4) BearingPoint (2016). *Reifegrad und Relevanz des Stammdatenmanagements Stammdaten als unabdingbare Basis für die Digitalisierung, Prozesseffizienz und Compliance.* [online] Available at: https://www.bearingpoint.com/de-de/insights-events/insights/stammdatenmanagement/ [Accessed 03.03.2024].
5) Brindha, T. and Murugesapandian, N. (2016). Scientometrics Tools and Techniques: An Overview. *Shanlax International Journal of Arts, Science & Humanities,* 4(2), pp. 90-92.
6) Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E. and Herrera, F. (2011). Science mapping software tools: review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, 62(7), pp. 1382-1402.
7) Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. and Lim, W.M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *J. Bus. Res.*, 133, pp. 285-296.
8) Esfahani, H., Tavasoli, K. and Jabbarzadeh, A. (2019). Big data and social media: A scientometrics analysis. *International Journal of Data and Network Science*, 3(3), pp. 145-164.
9) Gartner (2021). *What is modern MDM?* [online] Available at: https://www.gartner.com/en/documents/4005840 [Accessed 03.03.2024].

10) Grosser, T. (2019). *Die Erfolgsfaktoren für zufriedenstellende Datenqualität und Stammdaten.* [online] Available at: https://barc.com/de/die-erfolgsfaktoren-fuer-zufriedenstellende-datenqualitaet-und-stammdaten/ [Accessed 03.03.2024].
11) Havemann, F. (2009). *Einführung in die Bibliometrie,* Berlin: Auflage. http://d-nb.info/993717780
12) Jagals, M., Karger, E. and 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), pp. 105-120.
13) Khatri, V. and Brown, C.V. (2010). Designing Data Governance. *Communications of the ACM*, 53(1), pp. 148-152.
14) Lee, S.U., Zhu, L. and Jeffery, R. (2018). Designing data governance in platform ecosystems. In: Bui, T.X. (ed.), *51st Hawaii International Conference on System Sciences: Proceedings.* Atlanta (GA, USA): Association for Information Systems (AIS), pp. 5014-5023.
15) Loshin, D. (2009). *Master Data Management*. MK Morgan Kaufmann.
16) Mittelstand-Digital (2022). *Stammdatenqualität – die Basis für erfolgreichen Handel. Bedeutung, Technologien und Management. Leitfaden*. [online] Available at: https://digitalzentrumhandel.de/wp-content/uploads/2022/04/leitfaden-stammdatenerzeugung.pdf [Accessed 03.03.2024].
17) Mobin, M.A., Mahi, M., Hassan, M.K., Habibm M., Akter S. and Hassan T. (2023). An analysis of COVID-19 and WHO global research roadmap: knowledge mapping and future research agenda. *Eurasian Econ Rev,* 13, pp. 35-56. https://doi.org/10.1007/s40822-021-00193-2
18) Nalimov, V. and Mulcjenko, B. (1969): *Scientificometrics. Study of the development of science as an information process*. Moskau: Science.
19) Nalimov, V. and Mulcjenko, B. (1971). *Measurement of Science: Study of the Development of Science as an Information Process*. Washington DC: Foreign Technology Division.
20) Necula, S.C. and Păvăloaia, V.D. (2023). AI-Driven Recommendations: A Systematic Review of the State of the Art in E-Commerce. *Applied Sciences.*, 13(9), pp. 5531.
21) OECD (2009). International co-operation in science. In: *OECD Science, Technology and Industry Scoreboard*. Paris: OECD Publishing.
22) Otto, B. (2009). Funktionsarchitektur für unternehmensweites Stammdatenmanagement. Bericht Nr.: BE HSG/C CDQ/14, Institut der Wirtschaftsinformatik, St. Gallen.
23) Otto, B. (2011). A morphology of the organisation of data governance. In: *ECIS 2011 Proceedings*. Helsinki, p. 272.
24) Popescul, D., Radu, L.D., Păvăloaia, V.D. and Georgescu, M.R. (2020). Psychological Determinants of Investor Motivation in Social Media-Based Crowdfunding Projects: A Systematic Review. *Front. Psychol.*, 11, pp. 588121.
25) PWC (2018). *Master Data Management im Handel und in der Konsumgüterindustrie. Wie aus dem „neuen Öl“ effizient Energie gewonnen wird.* [online] Available at: https://www.pwc.de/de/digitale-transformation/mdm-im-handel-und-in-der-konsumgueterindustrie.html [Accessed 03.03.2024].

26) Scheuch, R., Gansor, T. and Ziller, C. (2012). *Master Data Management. Strategie, Organisaton, Architektur. TDWI Europe Edition.* Heidelberg: Dpunkt Verlag.
27) Schmidt, A. (2010). *Entwicklung einer Methode zur Stammdatenintegration*. Berlin: Logos Verlag GmbH Berlin.
28) Thakur, D., Wang, J. and Cozzens, S. (2011). What does international co-authorship measure? In: Melkers, J., Monroe-White, T. and Cozzens, S. E. (eds.), *Atlanta Conference on Science and Innovation Policy,* Atlanta: IEEE Explore, pp. 1-7.
29) The Economist (2017). *The world's most valuable resource is no longer oil, but data. The data economy demands a new approach to antitrust rules*. [online] Available at: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data?_ga=2.6241139.1130524973.1709478685-1692148278.1709478684 [Accessed 03.03.2024].