CULTIVATING DATA OBSERVABILITY AS THE NEXT FRONTIER OF DATA ENGINEERING: A PATH TO ENHANCED DATA QUALITY, TRANSPARENCY, AND DATA GOVERNANCE IN THE DIGITAL AGE

https://doi.org/10.47743/jopafl-2023-30-19

SCHMUCK Matthias
Alexandru Ioan Cuza University of Iaşi, Doctoral School of Economics and Business Administration
Iaşi, Romania
matthias.schmuck@student.uaic.ro

Abstract: In the age of increasing process automation and data-driven decision-making, ensuring the reliability, transparency and usability of data is of paramount importance. In this context, the concept of "data observability" has aroused the interest of practitioners and there is also a lot of grey content on it. On the other hand, there is a lack of academic effort to define and build on the concept.
This conference paper will therefore examine the importance of "data observability" in modern data ecosystems. The focus is on the definition and characterisation of the concept, the differentiation from other concepts (e.g. data quality, data monitoring, data discovery, data operations) and why this concept appears to be so important in an increasingly data-driven world. In addition, the concept of "data observability" is categorised in the dynamically developing research field of data governance.
For this purpose, a multivocal literature review (MLR) was conducted, a form of systematic literature review (SLR) which, in addition to the published (formal) literature (e.g. journal and conference papers), also includes and brings together the grey literature (e.g. blog posts, videos and white papers).
The results show that the concept of "data observability" has the potential to revolutionise the way companies manage, analyse and derive insights from their data, ultimately leading to more informed and confident decision-making. Nevertheless, there is still plenty of room for further research into the specific contribution to better data and therefore better business processes and decisions.
Keywords: Digitalisation, data-driven company, data management
JEL Classification: L15, M15, O33

This Article was presented as a paper at the 15th edition of the Annual International Conference Globalization and Higher Education in Economics and Business Administration (GEBA 2023), which was held at the Alexandru Ioan Cuza University, Faculty of Economics and Business Administration in Iasi, Romania from the 19-21 October 2023.

Introduction

With advancing digitalisation and the increasing introduction and use of information and communication systems (ICT), companies are collecting and sto-ring data. As organisations collect seemingly endless streams of data from more and more sources, an ecosystem of data stores, pipelines and potential end users emerges. With each additional layer of complexity, the opportunities for data downtime - moments when data is incomplete, incorrect, unavailable or otherwise inaccurate - multiply. Such events tie up resources to resolve data quality issues (Moses, 2020; Forrester, 2018), prevent the introduction of artificial intelligence solutions (Mehrabi et al, 2019; Asay, 2017; Vartak, 2023; Coegil, 2023) or lead to the loss of customers (Violino, 2019). Twenty years ago, it was considered normal for business applications to fail or be unavailable. Today, enterprise applications have become business-critical, outages can be expensive and companies are investing accordingly to avoid service interruptions. This development in application technology must also apply to data technology if companies (want to and must) increasingly rely on data. To achieve this, companies must reach a new level of data reliability over the next few years, in which data engineers adapt the knowledge and experience of application engineers and apply it to their data processes. The more companies prioritise “data observability”, the more they will enter the new frontier of data engineering: the prevention of data downtime.

Methodology and material

A systematic literature review (SLR) approach in the form of a multivocal literature review (MLR) was used to answer the research questions and investigate the concept of "data observability" as a particular method of data engineering.

Research approach: This study is part of qualitative research.

Research method: We agree with Fink (2019) that SLR is "a systematic, explicit, and reproducible method for identifying, evaluating, and synthesising the existing body of completed and recorded work produced by researchers, scholars, and practitioners". However, many research topics today originate from practice (e.g. software industry), so it is only logical to include these voices. MLRs therefore recognise the need for "multiple" voices, rather than extracting evidence only from knowledge found in academic settings (formal literature). This is underlined by Ogawa and Malen (1991), in which they define MLR as follows: "Multivocal literatures are comprised of all accessible writings on a common, often contemporary topic. The writings embody the views or voices of diverse sets of authors (academics, practitioners, journalists, policy centres, state offices of education, local school districts, independent research and development firms, and others).". In addition to the published (formal) literature (e.g. journal and conference papers), the MLR also includes and brings together grey literature (e.g. blog posts, videos and white papers) (Garousi et al., 2019; Baysal et al., 2022; Gramlich et al., 2023; Janes et al., 2023; Trendowicz et al., 2023). To ensure the validity of our research design, i.e. the consistency between the defined and the conducted review process, we designed our study based on best practice guidelines defined in the relevant literature. These include in particular Garousi et al. (2019) for the overall design of the MLR and Kitchenham and Charters (2007) for parts of the formal literature and Adams et al. (2017) for the parts of the grey literature.

Research question(s): The current study provides answers to the following research questions: (Q1) What does the concept of "data observability" entail? (Q2) How does "data observability" differ from other concepts, e.g. "data quality", "data monitoring" or "data lineage"? (Q3) What are the benefits of “data observability”? (Q4) What challenges are discussed in the context of “data observability”? (Q5) How can “data observability” be categorised in the dynamically developing research field of Data Governance?

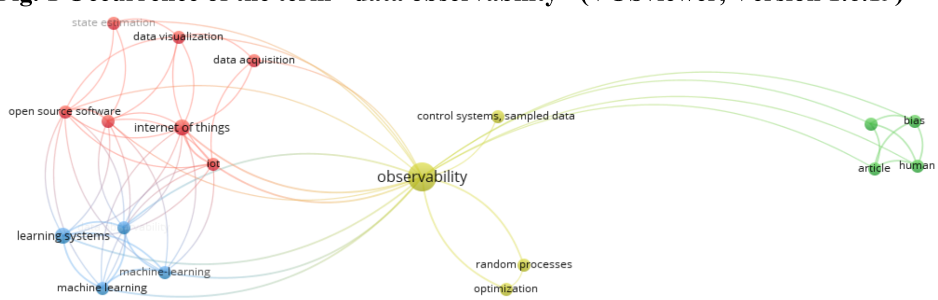
Source(s) of information: As part of the MLR, we considered various sources of formal and grey literature. The focus of the formal literature was on peer-reviewed publications of

primary studies in the databases Scopus, SpringerLink, Wiley, ACM Digital Library and Google Scholar. For the review of grey literature, we used Google Search, an established search engine for such a search (Adams et al., 2016; McGrath et al., 2006).
Inclusion and exclusion criteria: Inclusion and exclusion criteria were also applied. The documents selected on the basis of content comparison and relevance to the defined objectives of the current analysis were read, critically analysed and relevant information on the questions posed was extracted. The results are presented in the following section to facilitate the identification of findings and future directions in the literature reviewed.
Overall, the chosen review methodology provides an up-to-date and comprehensive insight into the concept that is attractive to practitioners and encourages further academic endeavours to define, develop and evaluate the concept in practice.

Results and discussion

Key bibliometric factors
To gain a general understanding of the topic of "data observability", we conducted a search in the Scopus database at the end of October 2023. We used the search term "data observability" to analyse the occurrence of keywords in this research. Fig. 1 shows the co-occurrence of various keywords, including "data acquisition", "data visualisation", "state estimation" and "bias", to name but a few. "Observability" is placed in the centre as a central concept, applied to the object "data".

Fig. 1 Occurrence of the term "data observability" (VOSviewer, Version 1.6.19)



(Source(s): Contribution of the author)

The coincidence was determined on the basis of a minimum threshold of two occurrences for each keyword, as the underlying list contained few data records.

Fig. 2: Thematic development (biblioshiny, an application for bibliometrix)

(Source(s): Contribution of the author)

Learning systems, i.e. machine learning (Feinberg et al., 2022), are the drivers in this field of research (Fig. 2). “data observability” uses machine learning to identify data problems, find the causes and assess the effects. Another topic is the visualisation of data. For example, Wang & Schneeweis (2022) propose a simple visualisation of data type and observability using solid and dotted lines and 2-colour palette to capture the differences between electronic health records and other registry data, which may have limited data continuity, and insurance claims data, which have enrolment files.

The concept (Q1)

The term "data observability" consists of two terms: "data" and "observability", where "observability" refers to the object "data" (has an effect on).

Data: According to DAMA UK (2023), data is "... a re-interpretable representation of information in a formalised manner suitable for communication, interpretation or processing". Data has special characteristics (Merkus et al., 2019), including versatility of format, volatility, fluidity or replicability (Zygmuntowski et al., 2021), which is why it must be treated in a special way.

Observability: The term "observability" was first coined in the 1960s as a method related to the measurement of a system based on its performance (Kálmán, 1960; Kálmán, 1963). Today, more than six decades later, “observability” has split into several specialised segments - from “application observability” to “security observability” and everything in between. “observability” is a measure of how well the internal states of an object (or system) can be deduced from knowledge of its external outputs. If the current state of an object (or system) can be assessed purely by information from the output values, “observability” is considered to be given (Pavlek and Kalpic, 2008).

Data observability: The concept of " data observability " is an approach inspired by "Development and Operations" (DevOps) to ensure data quality (Moses, 2023) and is defined differently in the literature, as is very often the case with concepts that have derived from practice. According to Mohler and Hwang (1988), "data observability" refers to the ability to reconstruct a process state x(t) from measured data y(t) in a selected time period t E [to, t1]. This very mathematical description does not appear to be very clear. The definition by Moses (2023) from 2019 is better, according to which "data observability" is the ability of an organisation to fully understand the state of the data in its systems in order to reduce the frequency and impact of data outages. "data observability" thus focuses on reducing error rates in the data (Strod, 2021). Moses and Strod recognise the concept of "data observability" as a human-task-technology system (Heinrich, 1993, p. 173), which essentially comprises a series of activities (better: processes, practices) as well as people and technologies whose (socio-technical) combination enables you to identify, correct and solve data problems in near real time (Databand, 2023). The concept is based on five pillars (Moses, 2023): (a) data freshness (How fresh is your data?), (b) data schema (How is your data organised?), (c) data scope (How complete is your data?), (d) data quality (How reliable is your data?) and (e) data provenance (How are the data assets in your data pool connected upstream and downstream and how are responsibilities defined?). Together, these components provide a valuable insight into the quality and reliability of your data.

Delimitations (Q2)
In the following section, the concept of "data observability" is differentiated from other concepts.
Data observability vs. data quality: “data quality” is a measure of the quality of data for its intended use in operational and analytical applications and refers to the accuracy, completeness, consistency and timeliness of the data (LightsOnData, 2023; Walker, 2022; Reno, 2023; Atlan, 2023a). On the other hand, “data observability” enables the monitoring and investigation of systems and data pipelines to develop an understanding of the state and performance of the data. Both concepts overlap but work synergistically together to ensure confidence in the data (Atlan, 2023a).
Data observability vs. data monitoring: Both terms are often used interchangeably in practice because they fulfil similar functions to help identify data problems; however, they differ fundamentally in their definition (Travkin, 2022; Krishnamoorthy, 2021; Rakibe, 2023). “data monitoring” can only alert a data specialist to data problems, but not necessarily tell them how to fix them. “data observability”, on the other hand, can uncover the cause of the problem and provide details. In “data monitoring”, problems or irregularities are usually anticipated on the basis of defined criteria in order to measure the perceived problems. “data observability”, on the other hand, involves collecting metrics across the entire IT landscape in order to proactively detect potential anomalies. “data observability” measures all expenditure across multiple applications and systems, which “data monitoring” cannot do.
Data observability vs. data lineage: Both concepts are related, but serve different purposes in terms of data management and analysis (Atlan, 2023b; Bergh, 2023). “data lineage” focuses on understanding the historical movement of data between source and destination, while “data observability” is about real-time monitoring as the data flows through the systems. “data lineage” provides a detailed, step-by-step view of the data flow and the data transformations taking place within it, while “data observability” looks at aggregated metrics and insights about data systems and processes to identify patterns and trends. “data lineage” is represented by data flow diagrams, the metadata of which is usually stored in a separate system from the data itself, while “data observability” is represented by KPIs and alerts in dashboards.

Data observability vs. data detection: Both concepts are about the visibility of and information about data in real time (Immuta, 2023). “data detection” is about identifying anomalies, specific incidents or risks at a specific point in time, while “data observability” is about the same goal, but measured over time in order to recognise patterns and trends. “data detection” is therefore an upstream prerequisite for “data observability”.

Data observability vs. data discovery: The concept of "data discovery" is primarily used to accommodate data documentation and knowledge sharing in one place, i.e. to provide a single "source of truth" (Kim, 2023). This is done primarily through discovery-oriented data catalogues, which are to be distinguished from control-oriented catalogues, as are common in companies that have to manage exactly who has access to which data and for how long (de Leyritz, 2023). These catalogues are part of metadata management (Cittadin, 2022) and contain information that is in turn taken into account in “data observability” (e.g. value descriptions, data type descriptions).

Data observability vs. data operations (data ops): "data observability" serves as the monitoring basis for "data ops", an established discipline for building and managing data flows. "data ops” applies the principles of Dev Ops, agile software development and total quality management to data and data flows to deliver timely, accurate data to the organisation. "data ops” includes testing, continuous integration and delivery (CI/CD) and orchestration. "data observability" provides insights that make each of these elements more effective.

Data observability vs. data tests: both concepts complement each other (Atlan, 2023c). "data tests” are used to validate the content, structure and integrity of data with the aim of ensuring that the data is processed correctly at all stages of the data pipeline and is therefore accurate and reliable. “data observability” is about gaining insights into the health and performance of the entire data system in order to understand, diagnose and resolve problems in real time. "data tests” are reactive and are only carried out at specifically defined points in time (at regular intervals), while data observability is proactive, permanent and continuous. "data tests” focus on defined use cases and test conditions, while “data observability” does not address such framework conditions. "data tests” provide binary results ("test passed" or "test failed"), while “data observability” provides multidimensional insights.

Data observability vs. data cleansing: Both concepts are about ensuring the integrity and reliability of a data ecosystem (Atlan, 2023d). While “data cleansing” is about identifying and correcting inaccuracies, inconsistencies and duplicates within a data set, “data observability” focuses on monitoring and gaining insights into the behaviour, performance and quality of data as it flows through processes, pipelines and systems in real time. “data observability” is permanent and continuous, while “data cleansing” takes place during the data analysis process.

Data observability vs. data availability: "data availability" means that a company ensures that all business-relevant data is available to its employees, partners and other end users at any time of day (24/7) and at any location. This applies to both the accessibility and continuity of information. Data that cannot be accessed quickly can prevent the provision of services, which costs a company time and revenue. "data observability" presupposes "data availability", because "data observability" focusses on the object "data"; data that is not available cannot be observed.

Data observability vs. observability data: Both "observability data" and "data observability" deal with different aspects of the data lifecycle and complement each other (Cribl, 2023). For example, the lifecycle of "data observability" deals with the collection, storage, analysis and visualisation of company data, while "observability data" manages the forwarding of data to different locations in different versions. "data observability" is analytical, while "observability data" is more focussed on operations.
Overall, this list of concepts and their distinction from the concept of "data observability" shows that the boundaries are often fluid, i.e. not clearly defined.

Advantages (Q3)
In addition to the benefits already mentioned in the previous section, we have extracted the following potential benefits of "data observability" from the formal and grey literature (Table 1):

Table 1: Advantages of "data observability"

Advantages	Supported by ...
Increasing confidence in the data by identifying inconsistencies or data errors so that organisations can make more confident data-driven decisions.	Kutay (2023)
Ensuring the quality, reliability and consistency of data in the data pipelines by providing a 360-degree view of the data ecosystem.	Kutay (2023)
Tracking relationships that organisations didn't know they should be looking for and discovering data issues before they impact the business.	Kutay (2023); Magnusson (2023)
Increasing the maturity level of company data by simplifying root cause analyses.	Kutay (2023); Magnusson (2023)
Automation of safety management in the company (real-time identification of problems, automation of the triage process.	Magnusson (2023)
Improving the reliability of process flows when capturing and processing your data (more robust data environment).	Zeenea, 2023
Tracking the status of interfaces and enabling rapid response to (un)expected changes	TDWI, 2023
Recognising any changes in data schema and data logic	TDWI, 2023
Insights into the size of data deliveries and the ability to react quickly to (un)expected or missing data deliveries	TDWI, 2023
Enabling the traceability of responsibilities for data in decentralised teams	TDWI, 2023

(Source(s): Contribution of the author)

"data observability" helps to build trust through transparency, reduce effort and therefore costs, and optimise the data infrastructure.

Challenges (Q4)
"data observability" poses a number of challenges. Below, an overview of the main challenges recognised in the formal and grey literature is provided (Table 2):

Table 2: Challenges of „data observability"

The challenges	Supported by ...
Existing data silos (compartmentalised and inaccessible data) make it difficult to integrate all systems into the monitoring solution.	Magnusson (2023)
Different data models complicate the implementation of "data observability".	Magnusson (2023)

The challenges	Supported by ...
Different types of data storage in the systems and regulatory constraints (e.g. retention guidelines) influence the scalability of "data observability" (and thus generate high costs).	Magnusson (2023)
"data observability" technologies overlap with monitoring and data quality technologies. This fact must be taken into account when selecting tools, as the tools available on the market differ considerably in terms of the underlying technologies, the depth of observability and the scope of coverage.	Feinberg et al (2022)
The concept of "data observability" does not guarantee the correctness of data. It only observes, i.e. provides information on whether the data is delivered as it should be delivered.	Feinberg et al (2022)
The tools available on the market are aimed at data engineers, but not business users. The latter may therefore not find their requirements reflected in them.	Feinberg et al (2022)
"data observability" requires a data culture within the company, i.e. management support and an organisation that is prepared to learn, implement and follow the prescribed procedures.	Bange, 2022
""data observability" does not replace or supersede other components of data management, such as data quality management and master data management (MDM).	Pratt, 2022

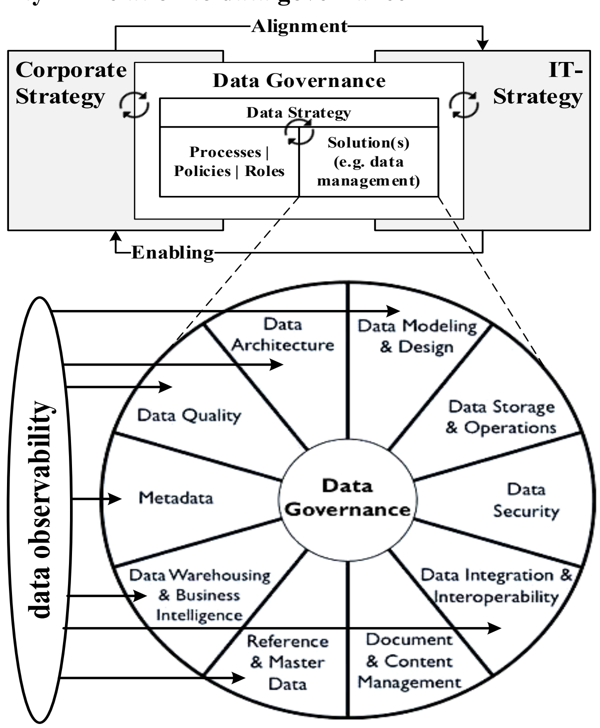
(Source(s): Contribution of the author)

Despite the challenges described above, "data observability" must become an essential component of a data-driven company.

Relationship to data governance (question 5)

"data observability" and Data Governance are two key concepts (Fig. 3) that play an important role in ensuring the accuracy, reliability and efficiency of data-driven (enterprise) systems and thus form the basis for data-driven organisations.

Fig. 2: "Data observability" in relation to data governance



(Source(s): Author's contribution, with application of Krcmar, 2005, p. 316 and DAMA, 2017)

The starting point is the mutual relationship between the "corporate strategy" and "IT strategy" tasks, a relationship known as "business-IT alignment" in the sense of a control loop. The object of the "corporate strategy" is the entire company, including its corporate processes and organisation. The results of the "corporate strategy" are specifications for all downstream levels (alignment). In contrast, the object of the "IT strategy" is the operational information system, consisting of a business process model and a configuration plan for the resources to be used in the information system. The latter requires a definition of the degree of automation of processes and technology utilisation. In this respect, technological options are developed that are to be utilised in the "corporate strategy" in order to improve the company's competitiveness in the long term (enabling).
Data Governance is the closed-loop process of managing the availability, usability, integrity and security of the data used in an organisation (Wende & Otto, 2007; Otto, 2012; Weber et al, 2009; Newman & Logan, 2006; Gregory, 2011). Based on a “data strategy” that supports the "corporate strategy" in a closed-loop manner by identifying the business objectives and key drivers related to data as an operational asset, “processes”, “policies” and “roles and resposnibilities” are defined and “solutions” are selected to ensure that corporate data is collected, stored, processed and shared in a secure and compliant manner in a closed-loop manner to ensure that corporate data is accurate and trustworthy for decision makers. One of these solutions is “data management”, which manages the entire lifecycle of an organisation's data by executing and activating the rules and policies described in data governance, thus answering the question "How do you manage corporate data?". To make the concept of “data management” more tangible, we use a well-known data management framework, the DAMA DMBOK (DAMA, 2017). One component of this framework is the "DAMA wheel", which concretises the knowledge area of "data management" with processes such as "data integration", "data warehousing & business intelligence" or "data architecture". In this environment, the concept of "data observability" has points of contact with different knowledge areas of "data management" and has a formative effect on them.
As a result, it can be stated that Data Governance is the framework concept and "data observability" is a solution process itself as well as a formative component of other solution processes in “data management”.

Summary

Through this study, we have succeeded in providing an insight and overview into the topic of "data observability" as a method of modern data engineering. With this in mind, we have applied the literature review as a research method to a number of papers identified in the formal and grey literature. In particular, we included the grey literature because the concept currently enjoys a great deal of attention in practice and a large number of sources can be found about it. This is also reflected in Gartner's assessment of "data observability" as hype in the "Hypecycle for Data Management" (Feinberg et al., 2022).
We were able to capture the concept, differentiate it from other concepts and identify the benefits and challenges. The research had a special focus on categorisation in the research field of data governance; in this respect, this study is unique.

The results of the study show that "data observability" can be a suitable means of increasing trust in company data. Future research will focus on the practical side of implementation and utilisation in specific corporate environments, particularly in the application of self-learning artificial intelligence systems. In particular, the value contribution of "data observability" solutions must be evaluated over time and whether the design and provisioning effort is worthwhile in relation to the goal of providing better data for better business processes and decisions.

Acknowledgements
Funding: This research received no external funding. Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Adams, J., Hillier-Brown, F. C., Moore, H. J., Lake, A. A., Araujo-Soares, V., White, M. and Summerbell, C. (2016). Searching and synthesising 'grey literature' and 'grey information' in public health: Critical reflections on three case studies. Systematic Reviews, 5(1), article 164. https://doi.org/10.1186/s13643-016-0337-y
2. Adams, R. J., Smart, P. and Huff, A. S. (2017). Shades of grey: Guidelines for working with the grey literature in systematic reviews for management and organisational studies. International Journal of Management Reviews, 19(4), pp. 432-454. https://doi.org/10.1111/ijmr.12102
3. Asay, M. (2017). 85% of big data projects fail, but your developers can help yours succeed. [online] Available at: https://www.techrepublic.com/article/85-of-big-data-projects-fail-but-your-developers-can-help-yours-succeed/ (Accessed 15 October 2023).
4. Atlan (2023a). Data Quality and Observability - Why Are They Important? [online] Available at: https://atlan.com/data-quality-and-observability/#:~:text=Data%20quality%20refers%20to%20the,to%20ensure%20trust%20in%20data. (Ac-cessed 15 October 2023).
5. Atlan (2023b). Data Lineage & Data Observability: Why Are They Important? [online] Available at: https://atlan.com/data-lineage-and-data-observabili-ty/#:~:text=In%20summary%2C%20data%20lineage%20focuses,and%20efficiency%20in%20different%20ways. (Accessed 15 October 2023).
6. Atlan (2023c). Data Observability vs Data Testing: 6 Points to Differentiate. [online] Available at: https://atlan.com/data-observability-vs-data-testing/ (Accessed 15 October 2023).
7. Atlan (2023d). Data Observability vs Data Cleansing: 5 Points to Differentiate. [online] Available at: https://atlan.com/data-observability-vs-data-cleansing/ (Accessed 15 October 2023).
8. Bange (2022). Data Culture: Definition, Challenges & Measures. [online] Available at: https://barc.com/de/data-culture/ (Accessed 15 October 2023).
9. Baysal, M.V., Özcan-Top, Ö. and Betin-Can, A (2023). Blockchain technology applications in the health domain: a multivocal literature review. J Supercomput 79, pp. 3112-3156. https://doi.org/10.1007/s11227-022-04772-1
10. Bergh, C. (2023). "You Complete Me," said Data Lineage to Data Journeys. [online] Available at: https://datakitchen.io/you-complete-me-said-data-lineage-to-dataops-observability/ (Accessed 15 October 2023).
11. Cittadin, B. (2022). Data Observability and Data Discovery: Better Together. [online] Available at: https://www.siffletdata.com/blog/data-observability-and-data-discovery-better-together (Accessed 15 October 2023).
12. Coegil (2023). Why 85% of AI Projects Fail to Deliver Expected Business Value & How to Improve Success. [online] Available at: https://www.linkedin.com/pulse/why-85-ai-projects-

fail-deliver-expected-business-value-how-improve?utm_source=share&utm_medium=guest_desktop&utm_campaign=copy (Ac-cessed 15 October 2023).

13. Cribl (2023). Observability Data vs Data Observability: What's the Difference? [online] Available at: https://cribl.io/blog/observability-data-vs-data-observability-whats-the-difference/ (Accessed 15 October 2023).
14. DAMA (2017). The Data Management Association. The DAMA Guide to The Data Management Body Knowledge. 2nd Edition. DAMA International.
15. DAMA UK (2023). Data. In: DAMA Glossary of data management terms. [online] Available at: https://www.dama-uk.org/Glossary (Accessed 15 October 2023).
16. Databand (2023). What is Data observability? Everything you need to know. [online] Available at: https://databand.ai/data-observability/ (Accessed 15 October 2023).
17. de Leyritz, L. (2023). How Do Data Discovery and Observability Interact? [online] Available at: https://www.castordoc.com/blog/data-discovery-and-data-observability-how-do-they-interact (Accessed 15 October 2023).
18. Feinberg, D., Russom, P. and Showell, N. (2022). Gartner Hype Cycle for Data Management, 2022.
19. Fink, A. (2019). Conducting Research Literature Reviews - From the Internet to Paper. SAGE Publications Inc.
20. Forrester (2018). Data Performance Management Is Essential To Prove Data's ROI. [online] Available at: https://www.forrester.com/report/data-performance-management-is-essential-to-prove-datas-roi/RES83344 (Accessed 15 October 2023).
21. Garousi, V., Felderer, M. and Mäntylä, M. V. (2019). Guidelines for including grey literature and con-ducting multivocal literature reviews in software engineering. Information and software technology, 106, pp. 101-121. https://doi.org/10.1016/j.infsof.2018.09.006
22. Gartner (2018). How to Create a Business Case for Data Quality Improvement. [online] Available at: https://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement (Accessed 15 October 2023).
23. Gramlich, V., Guggenberger, T., Principato, M.; Schellinger, B. and Urbach, N. (2023). A multivocal literature review of decentralised finance: Current knowledge and future research avenues. Electron Markets 33, 11. https://doi.org/10.1007/s12525-023-00637-4
24. Gregory, A. (2011). Data governance - Protecting and unleashing the value of your customer data assets. J Direct Data Digit Mark Pract 12, pp. 230-248 (2011). https://doi.org/10.1057/dddmp.2010.41
25. Heinrich, L.J. (1993). Business informatics. Introduction and foundation. Oldenbourg, Munich.
26. Immuta (2023). Data Observability vs. Data Detection. [online] Available at: https://www.immuta.com/guides/data-security-101/data-observability-vs-data-detection/ (Accessed 15 Oc-tber 2023).
27. Janes, A., Li, X. and Lenarduzzi, V. (2023). Open tracing tools: Overview and critical comparison. Journal of Systems and Software 204 (2023) 111793. https://doi.org/10.1016/j.jss.2023.111793
28. Kalman, R.E. (1960). On the general theory of control systems. In: IFAC Proceedings Volumes. 1: pp. 491-502. https://doi.org/10.1016/S1474-6670(17)70094-8
29. Kalman, R.E. (1963). Mathematical Description of Linear Dynamical Systems. Journal of the Society for Industrial and Applied Mathematics, Series A: Control. 1 (2): pp. 152-192. https://doi.org/10.1137/0301010
30. Kim, S. (2023). Data Discovery vs. Data Observability: Understanding the Differences for Better DataOps. [online] Available at: https://www.selectstar.com/blog/data-discovery-vs-data-observability-understanding-the-differences-for-better-dataops (Accessed 15 October 2023).
31. Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical report. EBSE 2007-001. Keele University and Durham University Joint Report.
32. Krcmar H. (2005). Information Management. 4th edition. Springer-Verlag, Berlin 2005

33. Krishnamoorthy, M. (2021). Data Observability vs Data Monitoring - What is the Difference? [online] Available at: https://www.qualdo.ai/blog/monitoring-vs-observability-what-why-and-how/ (Accessed 15 October 2023).
34. Kutay, J. (2023). Data Observability: What Is It and Why Is It Important? [online] Available at: https://www.striim.com/blog/data-observability-what-is-it-and-why-is-it-important/#:~:text=With%20good%20data%20observability%2C%20organisations,prevent%20data%20downtimes%20and%20inconsistencies (Accessed 15 October 2023).
35. LightsOnData (2023). Data Observability vs. Data Quality: A Comprehensive Discussion. [online] Available at: https://www.lightsondata.com/data-observability-vs-data-quality-a-comprehensive-discussion/ (Accessed 15 October 2023).
36. Lupu, D., & Tiganasu, R. (2024). Does education influence COVID-19 vaccination? A global view. Heliyon, 10(3).
37. Magnusson, A. (2023). Data Observability Explained. [online] Available at: https://www.strongdm.com/blog/data-observability (Accessed 15 October 2023).
38. McGrath, Y., Sumnall, H., Edmonds, K., McVeigh, J. and Bellis, M. (2006). Review of grey literature on drug prevention among young people. London/Manchester: National Institute for Health and Clinical Excellence. ISBN: 1-84629-184-4.
39. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A. (2021). A survey on bias and fair-ness in machine learning. ACM computing surveys (CSUR), 54(6), pp. 1-35. https://doi.org/10.48550/arXiv.1908.09635
40. Merkus, J.; Helms, R. and 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, ISBN 978-989-758-372-8; ISSN 2184-4992, SciTePress, pp. 143-154. https://doi.org/10.5220/0007411901430154
41. Mohler, R. R. and Hwang, C. S. (1988). Nonlinear data observability and information. Journal of the Franklin Institute, 325(4), pp. 443-464. https://doi.org/10.1016/0016-0032(88)90055-5
42. Moses, B, (2020). How to Calculate the Cost of Data Downtime Introducing a better way to measure the financial impact of bad data on your company. [online] Available at: https://towardsdatascience.com/how-to-calculate-the-cost-of-data-downtime-c0a48733b6f0 (Accessed 15 October 2023).
43. Moses, B. (2023). What is Data Observability? 5 Key Pillars To Know. [online] Available at: https://www.montecarlodata.com/blog-what-is-data-observability/ (Accessed 15 October 2023).
44. Newman, D., & Logan, D. (2006). Governance Is an Essential Building Block for Enterprise Information System. Gartner Research, May.
45. Ogawa, R.T. and Malen B. (1991). Towards rigour in reviews of multivocal literatures: applying the exploratory case study method. Rev. Educ. Res. 61 (3) (1991) pp. 265-286.
46. Otto, B. (2012). Managing the business benefits of product data management: The case of Festo. Journal of Enterprise Information Management, 25(3), pp. 272-297.
47. Pavlek, S. and Kalpic, D. (2008). Observability of information in databases. New Spins in Data Ware-housing for Credit Risk Management. In: Proceedings of the Third International Conference on Software and Data Technologies, pp. 361-368. https://doi.org/10.5220/0001892903610368.
48. Pratt, C. (2022). Data observability. [online] Available at: https://www.techtarget.com/searchdatamanagement/definition/data-observability (Accessed 15 October 2023).
49. Rakibe, M. (2023). Data Observability vs. Monitoring vs. Testing. [online] Available at: https://www.dataversity.net/data-observability-vs-monitoring-vs-testing/ (Accessed 15 October 2023).
50. Reno, G. (2023). Differences Between Data Quality and Data Observability. [online] Available at: https://firsteigen.com/blog/differences-between-data-quality-and-data-observabili-ty/#:~:text=Data%20quality%20is%20concerned%20with,users%20to%20data%20quality%20issues. (Ac-cessed 15 October 2023).

51. Strod, E. (2021). Data Observability and Monitoring with DataOps. [online] Available at: https://datakitchen.io/data-observability-and-monitoring-with-dataops/ (Accessed 15 October 2023).
52. TDWI (2023). Data observability in analytics projects. Which techniques you can use to monitor data flows seamlessly. [online] Available at:
53. https://www.youtube.com/watch?v=4yM1LDX05NA (Accessed 15 October 2023).
54. Travkin, P. (2022). Data Observability vs. Data Monitoring; What's the Difference? [online] Available at: https://solutionsreview.com/data-management/data-observability-vs-data-monitoring-whats-the-difference/ (Accessed 15 October 2023).
55. Trendowicz, A., Groen, E. C., Henningsen, J., Siebert, J., Bartels, N., Storck, S. and Kuhn, T. (2023). User experience key performance indicators for industrial IoT systems: A multivocal literature review. Digital Business, Vol. 3, Issue 1, 100057. https://doi.org/10.1016/j.digbus.2023.100057
56. Vartak, M. (2023). Achieving Next-Level Value From AI By Focusing On The Operational Side Of Machine Learning. Forbes Technology Council COUNCIL POST. [online] Available at: https://www.forbes.com/sites/forbestechcouncil/2023/01/17/achieving-next-level-value-from-ai-by-focusing-on-the-operational-side-of-machine-learning/ (Accessed 15 October 2023).
57. Violino, B. (2019). Companies are losing revenue opportunities and customers because of bad data practices. New report shows that poor data structure seems to be a significant barrier at many organisations. [online] Available at: https://www.zdnet.com/article/companies-are-losing-revenue-opportunities-and-customers-because-of-bad-data-practices/ (Accessed 15 October 2023).
58. Walker, E. (2022). Data Observability Vs Data Quality: What makes them different? [online] Available at: https://www.datasciencecentral.com/data-observability-vs-data-quality-what-makes-them-different/ (Accessed 15 October 2023).
59. Wang, A.V. and Schneeweiss, S. (2022). A Framework for Visualising Study Designs and Data Observability in Electronic Health Record Data. Clinical Epidemiology, pp. 601-608- https://doi.org/10.2147/CLEP.S358583
60. Weber, K., Otto, B. and Österle, H. (2009). One size does not fit all - A contingency approach to data governance. Journal of Data and Information Quality, 1(1), pp. 1-27. [Online] Available at: https://www.alexandria.unisg.ch/67793/1/a4-weber_external.pdf [accessed on November 01, 2023].
61. Wende, K. and Otto, B. (2007). A Contingency Approach To Data Governance. Iciq, pp. 163-176.
62. Zeenea (2023). All about Data Observability. [online] Available at: https://zeenea.com/de/alles-uber-data-observability/#:~:text=The%20first%20advantage%20of%20data,you%20recognise%20problems%20very%20early%20on%C3%BCh. (Accessed 15 October 2023).
63. Zygmuntowski, J J., Zoboli, L. and Nemitz, P.F. (2021). Embedding European values in data governance: a case for public data commons. Internet Policy Review, 10(3). https://doi.org/10.14763/2021.3.1572