Main Data Inhibitors and Enablers for AI Applications

Sujit Wings[1], Janne Härkönen [1]

[1] *Industrial Engineering and Management, University of Oulu, Erkki Koiso-Kanttilankatu 1, 90014 University of Oulu, Oulu, Finland*

Abstract

Companies are increasingly leveraging AI (Artificial Intelligence), in attempts to gain competitive advantage. This paper focuses on the AI applications for analytics to enable automated decision-making. The AI applications are especially attractive to companies due to them potentially enabling process automation, and the wider adoption of RPA (Robotic Process Automation). These appear as the key drivers for reducing operational process expenses. The specific focus of this paper is on the data-related inhibitors and enablers for AI applications, as AI relies heavily on data. The methodology involves a literature review and an in-depth case study, involving a questionnaire covering roles in the data domain and the product domain, semi-structured interviews, and analyzing internal use-case descriptions. The findings indicate that data fragmentation is among the main inhibitors. Data fragmentation appears as the root cause for the low quality of two intrinsic data quality dimensions, namely completeness, and consistency. In addition, data fragmentation drives the cost of AI modeling up, as data scientists need to re-create data assets on a per-use-case basis. The findings also indicate that productized data assets could be the main enabler for leveraging AI applications as they not only ensure the quality of the intrinsic data quality dimensions (correctness, completeness, timeliness, and consistency), but also contribute to the re-use of data assets. The latter is a driver for both cost reduction of AI modeling and faster AI model iterations, which in turn is a driver for AI model quality.

Keywords

AI Applications, AI modeling, data assets, data sources, metadata management, master data management, productized data assets, data quality

1. Introduction

Companies are increasingly leveraging AI (Artificial Intelligence) applications in attempts to become more competitive [1]. While much research is concerned with the actual AI modeling and the approaches to decision-making [2], the work leading up to the actual modeling work should not be neglected. Studies have shown that data scientists spend a significant percentage of their working time searching for data and then grooming and cleaning it [3].

In general, AI has the promise of providing vast opportunities, and applications for data analytics with more accurate predictions for decision-making [4]. Learning algorithms are used for data analysis to extract meaningful patterns from data to aid decision-making [5]. The role of AI can be either support the decision-making by a human or replace the human role [6]. As the use of AI in decision-making is still evolving there are challenges to overcome, which relate to the human-AI interaction, ability of AI to adapt to a new environment, and those of legal and technical nature [7].

Process automation is one of the potential uses for AI that businesses can reduce time, costs and minimize manual work [8]. Robotic process automation (RPA) is an umbrella term for tools that aim to replace people by

TKTP 2023: Annual Symposium for Computer Science 2023,
June 13-14, 2023, Oulu, Finland

automation [9]. RPA operates on structured data via a combination of user interface actions and by mapping a process for the software robot to follow [10]. RPA uses software that mimics human actions while interacting with applications and carrying out rule-based tasks. A related concept, hyper-automation combines RPA, AI, machine learning (ML), and other technologies, a form of intelligent process automation with possibilities beyond task or process automation [11]. The difference between hyper-automation, and RPA could be seen as data-driven vs. process-driven, or intelligent RPA vs. symbolic RPA [12].

In companies, business processes, information systems, products, and data are very much interlinked [13-15]. Master data is essential for business, the same as data from processes that enrich it [13]. All company transactions are done against the master data [15]. Hence, it is essential that all enterprise data would be treated as strategic asset [16]. Nevertheless, the data sources are various, with different operational purposes [17]. Hence, data quality is a vital aspect of master data [18-19] and operating with valid data helps to improve company performance [20]. Similarly, data quality of the input data has significance for AI model accuracy [21]. For example, the consistency of data, whether the logical relationship is correct and complete, i.e., the equivalency of data in different storage locations [22], or any of data quality dimensions and related elements [23] have significance. Nevertheless, data quality is context dependent [24]. In addition, aside different sources of data, the different structures of data format can be challenging to handle if applying AI for a specific purpose [25]. The data source quality and understanding the major sources of lack of data quality have significance for intelligent automation [26].

Regardless of numerous studies existing in relation to AI applications, RPA, and data, the AI related process automation, related automated decision-making, and the related analytic models lack case studies to shed further light on the practical business applications. Hence, this paper focuses on the data related inhibitors and enablers for AI applications through an in-depth case study to determine the root causes for the data fragmentation and quality in this type of a use, and how the situation could be remedied. The research problem is divided into the following research questions:

1. What are the main data inhibitors for AI applications?
2. What are the main data enablers for AI applications?

2. Literature review

2.1. Data quality

Two strategies are mentioned for improving data quality, data-driven, and process-driven [27]. The data driven focuses on modifying the data value, and process-driven focuses on re-designing the process to improve data quality. In addition, there are numerous studies focusing on data quality dimensions to discuss them from a variety of perspectives [19, 28]. The data quality dimensions include completeness, consistency, correctness, timeliness, accuracy, accessibility, believability, ease of manipulation, free-of-error, relevancy, reputation, security, to name a few [19, 23, 27-29].

Master data is vital for business, and their data quality has significance [18-19] as operating with valid data helps to improve company performance [20]. From the perspective of AI models, the data quality of the input data is vital for model accuracy [21]. In terms of different data storage locations, the consistency of data, and the correctness and completeness of the logical relationship, and the equivalency of data in different storage locations are important [22]. As can be the with any of data quality dimensions and related elements [23], some having more emphasis over others, depending on the context [24] and use. Also, the different structures of data format is a relevant perspective if applying AI [25]. Overall, the data source quality is relevant for intelligent automation [26].

2.2. Data governance

Generally, responsibilities need to be assigned to have effective data practices and enable data quality. Data governance is necessary to assign roles and responsibilities for data organization-widely, including both IT and business departments [19]. Data governance has been defined in three levels:

organizational level, support function level and data set level, and includes regulations, practices, procedures, data and concept ownerships, responsibilities and roles, and the descriptions of the roles [30]. Further, product data management is seen to have a role in implementing policies, procedures and guidelines defined through data governance [19]. Recently published research agenda poses a question on how organizations should structure their business and technology architectures to support data engineering and data governance to support multiple AI components with different ecosystem conventions [31]. This is a relevant question that links the AI considerations and data governance and emphasizes the relevance of data governance in the context. The data quality in terms of a single AI component associates with false-positives, and false-negatives. Multiple AI components relate to multiple data conventions. i.e., sets of data principles and standards.

2.2.1. Data product, data asset, and productization

Data being used as fuel for applications may necessitate considering data as a product to affectively address business needs. The product and business perspectives, and data and technology perspectives must be addressed to keep data scientist close to products and business [32]. Considering data as a product is also linked to data utilization [33]. Retrieving data for AI for analytics is seen as a challenge as too much time is spent on preparing data and a move from handmade to industrialized is seen as necessary, which in turn further necessitates seeing data as a product [34]. In analytics, it is the data asset that needs to be complete and consistent [35]. Data assets have been prepared for access, have quality metrics, metadata describing them, and they follow a data model [36]. Eichler et al. [36] also explain the difference between data asset and data product. Assessing the value of a data asset is highlighted as relevant to realize the enormous economic value in data [37]. Data account including business attributes, management attributes, asset attributes, and technical attributes are also seen relevant for data asset, linking to the definition of data unit [38]. Data productization is indicated to help in clarifying data and their measurement [32,33]. Productization in general is defined as “the process of analyzing a need, defining, and combining suitable elements, into a product-like defined set of deliverables that is standardized, repeatable and comprehendible” [39]. Further, data assets are stated to have significance for digitalization [17], but there do not seem to be clear definitions for data productization, or productization of data assets. This even if the goal of a productized data asset is stated to support all use cases [40]. The importance of being able to reuse data is highlighted to enable creating value from data [41]. The reuse of data assets would enable reiterating analytics models faster as no separate data work would be needed. The AI applications that are based on a specific analytics model would be of higher quality when the false positive and false negative rate would be less. Productized data asset could be synthesized to be something between the lines of data being under release management and version management and being mostly backward compatible. The backward compatibility ensures the comparability of results gained via AI analytics, and that the AI model can be iterated. In addition, productization of data asset should mean that the maturity of documentation (service description), and the maturity of production (service level agreements, risk management, and such) are in order.

3. Research process

Figure 1 illustrates the research process. The study was carried out in a selected case company (Company A), which was selected based on the possibility to have intimate access to study the related matters. The company provides insurance related services. Data and related analytics are very much of essence to the company, and hence can provide valuable knowledge. The specific focus in this study is on analytics for AI applications. To further narrow the focus, it was decided to investigate the underlying enablers and inhibitors specifically from a data perspective. The identified data enablers and inhibitors were further investigated to identify their root causes. A set of drivers were identified to support the focus on data related inhibitors and enablers:

- Data is the primary source for AI modelling and AI applications.
- Most roles involved in AI applications in the case company are working in the data domain:
    a. Data Scientists
    b. Data Analysts
    c. Business Analysts
    d. Solution Analysts
    e. Product Owners (data warehouse managers)
    f. Data leads
    g. Data/information architects
- The quality of an AI model is primarily dependent on the data quality of the underlying data asset.

The decision to focus specifically on analytics for the purpose of automated decision-making and process automation was made based on the financial impact of these use cases. Decision-making in the investigated processes (e.g., claims handling and fraud detection) were personnel intensive and time consuming. Both factors drove up the process costs. Process automation was seen as a potentially significant cost saver.

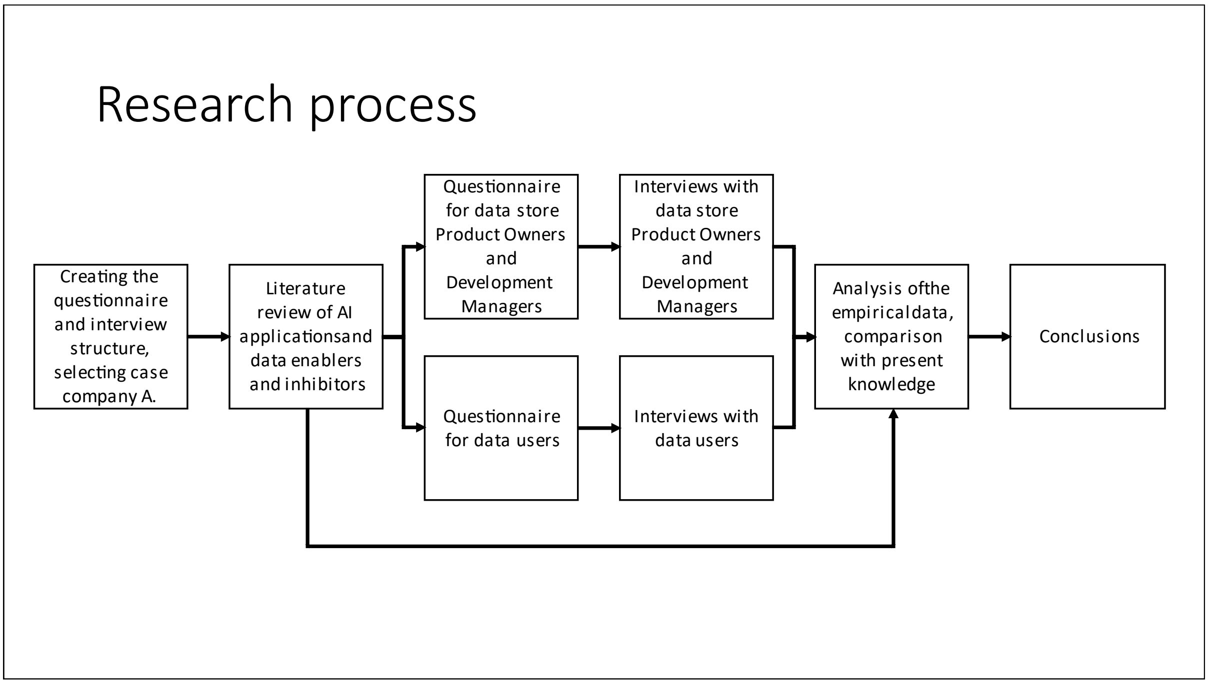


Figure 1: Outline of the research process

To investigate AI applications for analytics to enable automated decision-making, and the data-related inhibitors and enablers for AI applications further, two questionnaires were sent out to the participants. One questionnaire was sent to data storage owners, who are the custodians of data in their respective process or data domain. A second questionnaire was sent out to data consumers, who are using data, among other purposes, for AI applications. Figures 2 and 3 show the questionnaire structure. The questionnaire follows a black-box model to eliminate the need for prior training as there is a limited visibility to all process details, and this allows taking a position on completeness of data flows and data quality.

The results of the questionnaires were analyzed both quantitatively and qualitatively. The quantitative analysis was visualized as a maturity score ranging from 0 (low maturity) to 1 (high maturity). The maturity scores were color coded in three categories according to Table 1

Table 1
Color codes for the maturity scores

Color	Maturity score range
Green	0,75 and above
Yellow	0,35 – 0,74
Red	below 0,35

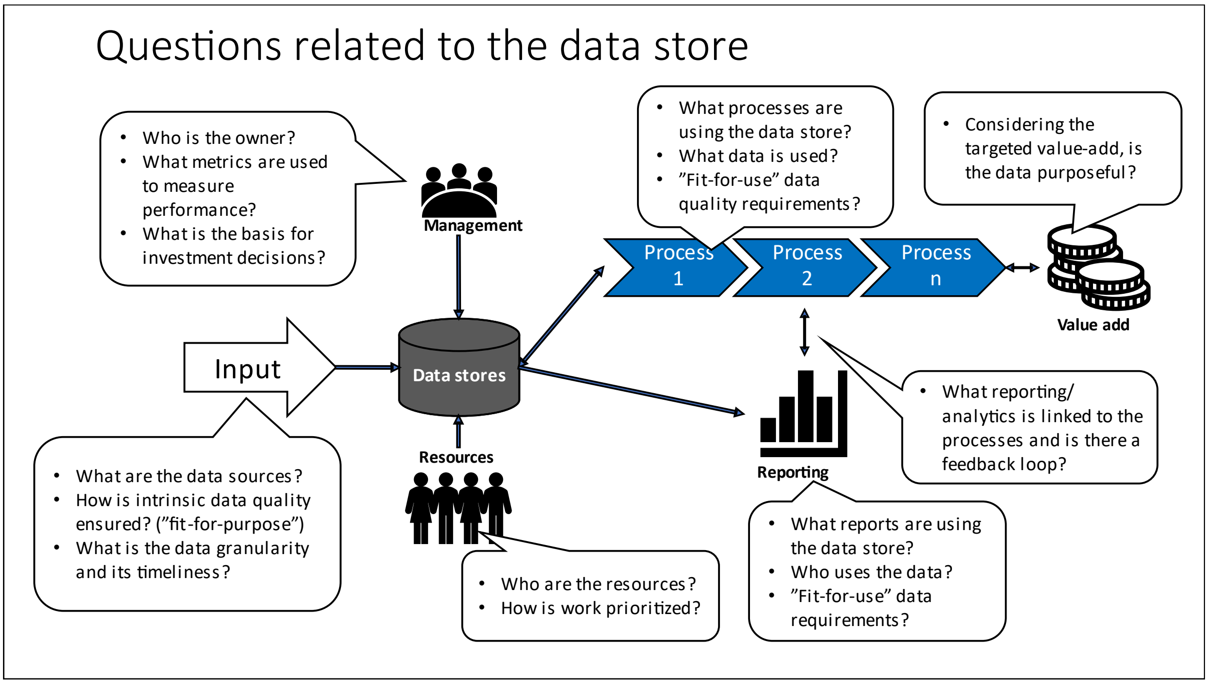


Figure 2: Questions related to the data store

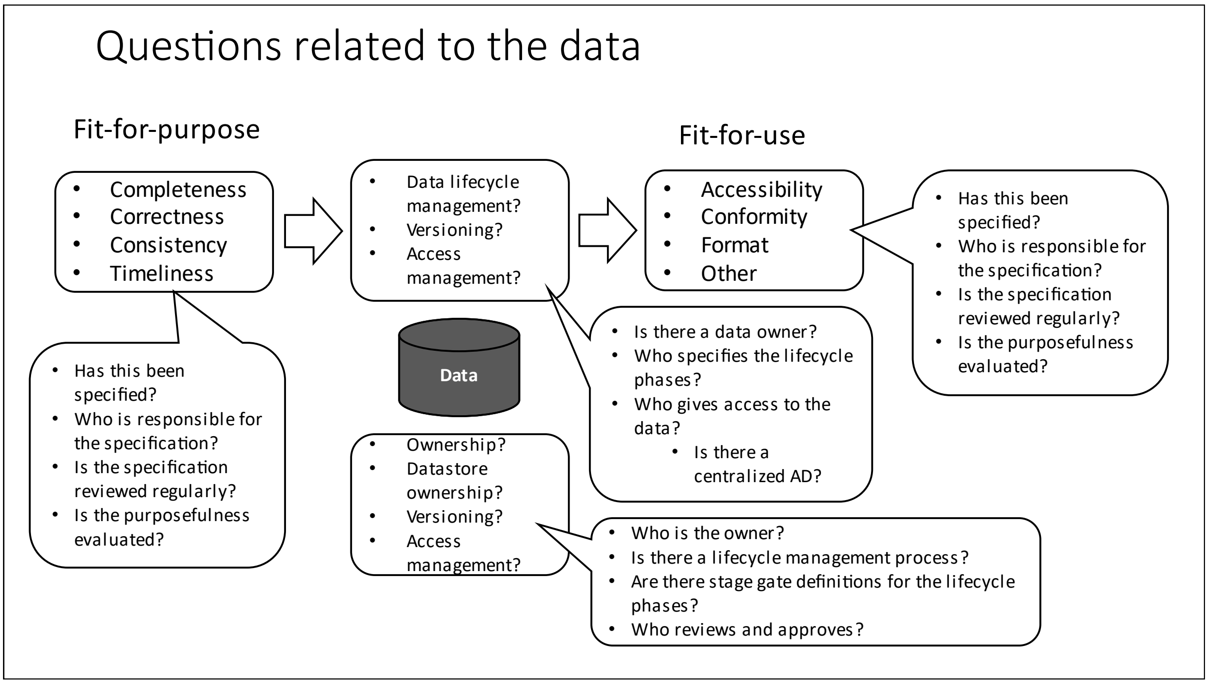


Figure 3: Questions related to the data

To gain further insights a series of semi-structured interviews [42] with key personnel were conducted. The interviewees included the following roles from the operational business side: Tribe Lead, Business Lead, Process Owner, and Process Developer. The following roles from the data management side were also included: Product Owner, Data Scientist, Data Analyst, Business Analyst and Solution Analyst. The interviewees were selected based on the relevance of their role to AI applications and related matters.

The questionnaires and semi-structured interviews were complemented with company internal documentation that included use-case descriptions, process charts and further process documentation, data schemas and other documentation related to the data stores.

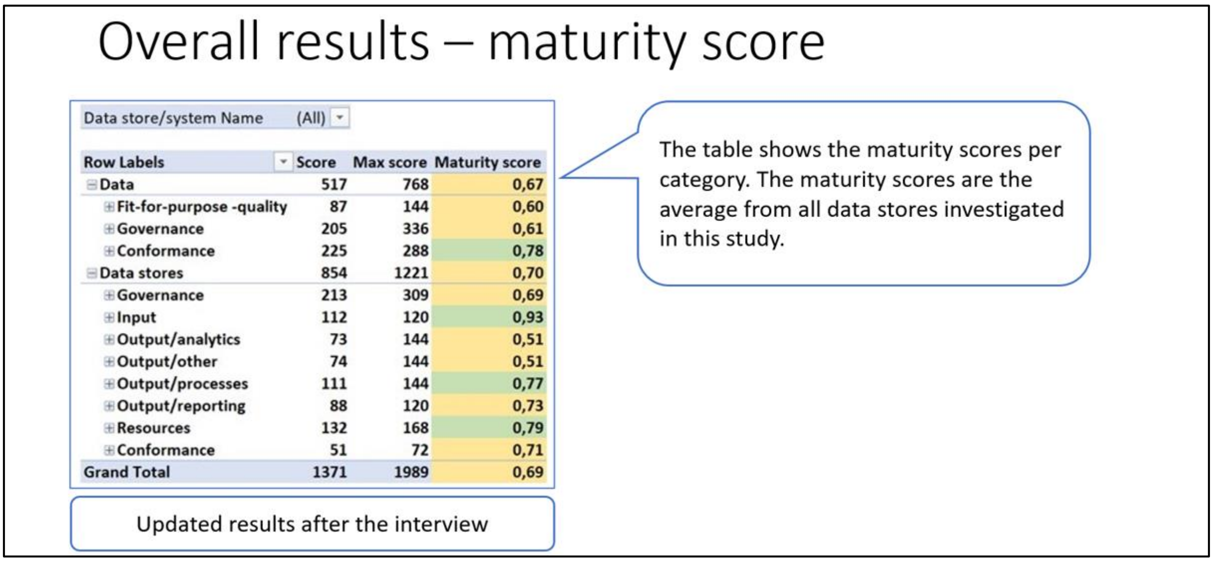
3.1. Case company A

The selected case company A is a Finnish insurance company that is part of a larger financial services group. The financial services group is the market leader in Finland and consists of a banking division (divided into a corporate banking and retail banking area), a life insurance division and a NLI (Non-Life Insurance) division. Case company A forms the NLI division of the financial services group.

The case company was selected due to the practical relevance to the studied topic and the ease of access to key personnel and internal documentation. Another factor was the ongoing large scale system renewal that was a major driver for process re-design and related AI applications.

The questionnaire results are visualized as maturity scores, ranging from 0 (low maturity) to 1 (high maturity). Figure 4 shows the overall average maturity scores of the investigated data stores (9 in total). The color coding is according to Table 1. The maturity levels in most of the categories are low (yellow), but not critical. High maturity levels are achieved only in a few areas. Conformity with laws and regulations (shown as conformity category) has a high (green) maturity level, which is not surprising. Conformity is a must in a highly regulated industry such as insurance. Therefore, controls are in place to ensure conformity. As a consequence, also the input has a high maturity level. The conformity requirements are spelled out in the work procedures for customer service, insurance agents and partners. Because the data stores and the data gathered there have been originally designed to serve a designated process, also the output/processes category has a high maturity level

4. Results



Row Labels	Score	Max score	Maturity score
Data	517	768	0,67
Fit-for-purpose -quality	87	144	0,60
Governance	205	336	0,61
Conformance	225	288	0,78
Data stores	854	1221	0,70
Governance	213	309	0,69
Input	112	120	0,93
Output/analytics	73	144	0,51
Output/other	74	144	0,51
Output/processes	111	144	0,77
Output/reporting	88	120	0,73
Resources	132	168	0,79
Conformance	51	72	0,71
Grand Total	1371	1989	0,69

Figure 4: Overall results of the maturity scores per category

The lowest maturity level (0,51 – yellow) is in the category output/analytics. This is not surprising, as analytics is newest of all the data related use cases. The category output/reporting has a higher maturity level (0,73 – only 0,2 short of scoring green), because reporting is an established use case.

4.1. Insufficient data governance and low data quality

Figure 5 shows the data governance, intrinsic data quality and conformance maturity levels. Again, the maturity levels for conformity are high (green) due to need to conform with laws and regulations. However, data governance and intrinsic data quality maturity levels are low. In the interviews especially the unclear data ownership and lack of an end-to-end data lifecycle management process were raised as the main governance issues.

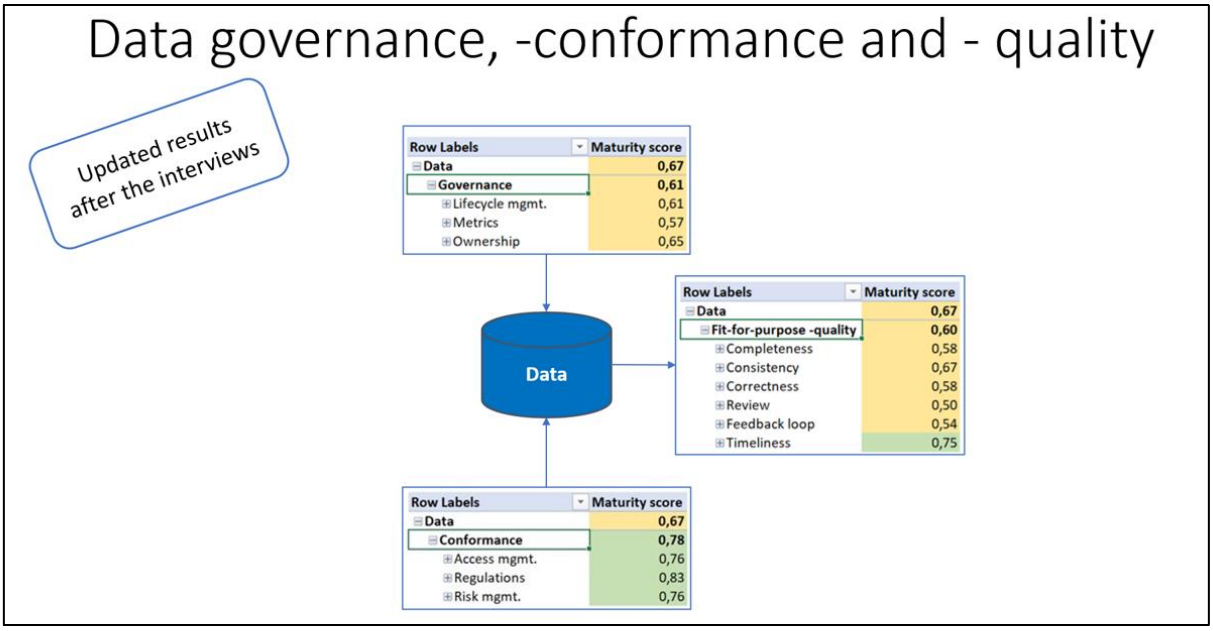


Figure 5: Data governance, -conformance and intrinsic quality

4.2. Low maturity for analytics use cases

Figure 6 shows that the maturity levels for analytics use cases are especially low, compared to other use cases. In the follow-up interviews the reasons became apparent. Analytics use cases are the most recent use-case category. In addition, analytics use cases often require data from multiple data stores. This leads to challenges with data consistency, as the data structure and format are different.

Another challenge is the low visibility of analytics use cases for the people maintaining the data store or managing the data. Process use cases (a process utilizes specific data from the data store) are the most transparent, as most data stores have originally been created to serve a specific process. There is a continuous exchange between the process managers and the data store product owner. Reporting use cases also provide a good visibility, because reports are regularly recurring and updates to the reports are discussed with stakeholders. In the case of analytics use-cases the people maintaining the data store might be involved in the first analytics use case specification meeting. However, the analytics model is then iterated multiple times and their involvement is much reduced or non-existent.

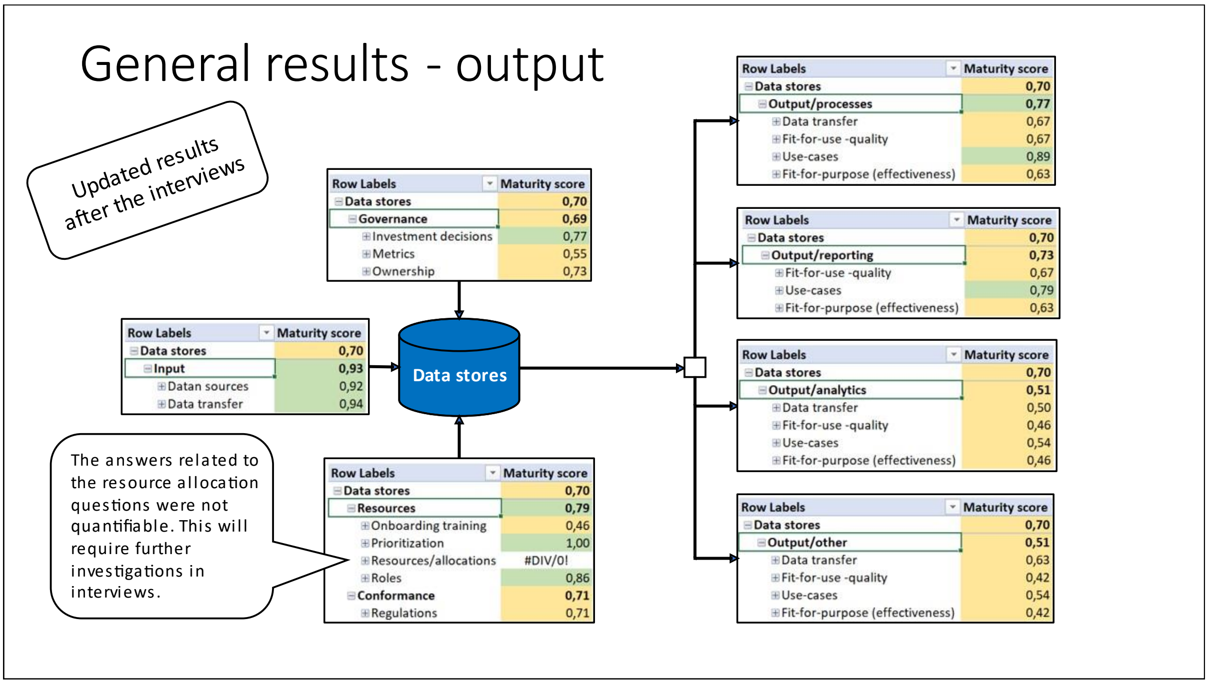


Figure 6: General results: maturity levels for different use-case categories

4.3. Limited knowledge of data coverage

Figure 7 below shows how many data stores the data users are using. It is remarkable that two thirds of the users use only three (or less) of the nine data stores. This is significant, as analytics use-case for the purpose of AI applications utilize multiple data stores to enable meaningful pattern recognition and correlations.

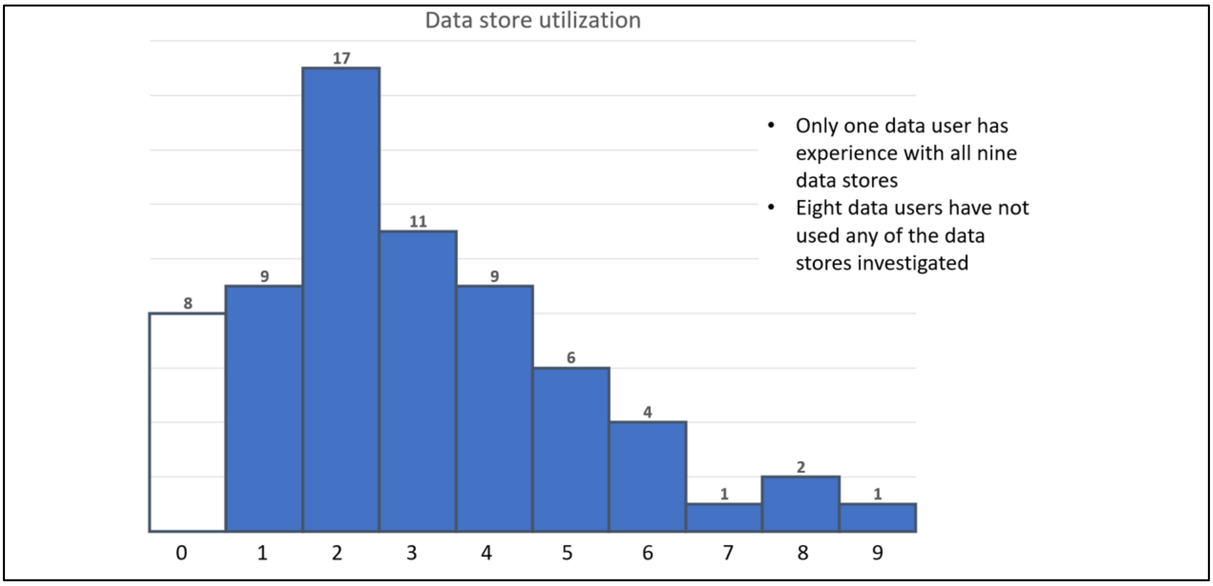


Figure 7: Data store utilization

During the interviews, the data users mentioned that they utilize only a limited number of data stores, because they had used them before and were familiar with the available data. The threshold for using multiple data stores as sources for analytics use cases was especially high. Because of the differing data structures and formats across various data stores, it requires a lot of effort to create a consistent data set for analytics use case.

5. Findings

It was found as expected that poor data quality is one of the main inhibitors for AI applications. However, when investigating the root causes of the poor data quality in the context of AI applications, a multi-faceted picture emerged. AI applications are at the end of a long chain of data collection, -storage, -publishing and -utilization. Each step is affecting data quality through the processes, systems and utilized tools. Because of the previous business focus on process efficiency, data flows across these various process steps have not been considered.

A specific AI application need is the requirement to quickly iterate AI models to achieve high quality results. AI model quality is primarily measured in terms of false positives/negatives. The requirement for quick iteration can only be met if the underlying data assets are re-usable and can be versioned.

The data quality dimensions that affect AI applications the most were intrinsic data quality dimensions, primarily completeness and consistency and secondarily correctness and timeliness. The root causes for the poor intrinsic data quality were manifold.

5.1. Data fragmentation as the Main Inhibitor

The fragmentation of data and data sources has been identified as the main inhibitor to intrinsic data quality, affecting mainly completeness and consistency dimensions. The reason for data fragmentation is historical.

5.1.1. Process fragmentation leads to system fragmentation

Case Company A has been structured according to functions, which reflect the company's main processes.

Main processes at Case Company A

- Customer lifecycle management
- Financial and regulatory reporting
- Claims management
- Insurance product lifecycle management

These processes are the responsibility of dedicated departments. Within the departments there are individual teams that handle one or multiple process phases within the main processes. So far, these departments have enjoyed a large measure of autonomy. The departments have chosen support systems that specifically serve their needs and did not pay attention to the interoperability of the different systems. The abundance of systems, which might even be specific to a process phase, leads to an abundance of data stores.

5.1.2. Impact on intrinsic data quality

The system landscape has grown organically, and its focus has been on serving specific processes and process phases. Because the teams/departments responsible for these processes have process efficiency as their main performance metric, the data collected and utilized is (partial-)optimized for the process.

- no unnecessary data has been collected (i.e., data that doesn't serve the specific process)
- data consistency across processes or process phases has not been considered

This has impacted mainly two intrinsic data quality dimensions, namely completeness and consistency.

5.2. Productized Data Assets as the Main Enabler

AI applications and the underlying analytics use cases that enable them would greatly benefit from productized data assets that are re-usable and provide good data quality. Productized data assets are the logical continuation from the growth of data utilization in organizations.

Originally data utilization was the domain of dedicated experts that accessed databases directly to generate mainly financial reports. As data utilization within organizations grew, data from one or multiple databases were aggregated into datamarts that served specific use cases.

The trend is towards "Democratization of Data," where data are no longer the domain of technical experts but are made both accessible and usable to business users throughout the organization. For this purpose, data must be productized to ensure controlled and consistent use of the data.

This paper's contribution is to show that productized data assets are a main enabler in this "Democratization of Data" trend.

5.2.1. Impact on data quality

The productization of a data asset would primarily impact two intrinsic data quality dimensions: consistency (foremost) and up to a certain extent also completeness, as the data asset would combine data from multiple data stores and group them into a data domain. However, the intrinsic data quality dimensions correctness, timeliness and the larger part of the completeness dimensions are influenced by the processes that gather the data.

5.2.2. Impact on re-usability

The biggest impact a productized data asset would have, is on the re-usability of the data. Through productization, the data asset would be under version- and release management, ensuring that the data set remains compatible. This is of importance for analytics models, as these require multiple iterations to reach a sufficient maturity. Productized data assets thus enable a faster iteration of analytics models, improving their quality and thus the quality of the related AI applications.

6. Conclusion

Leveraging AI for competitive advantage, and the use of AI applications for automated decision-making necessitate understanding the potential data-related inhibitors and enablers for these applications. Specifically, the analytics models the AI applications are based on were found to be of low maturity in the case study. Analytics being a relatively new use case, the needs of analytics are currently not paid enough attention. Due to AI heavily relying on data, this study specifically focused on the related inhibitors and enablers. Poor data quality was found to be the main inhibitor for AI applications. The fragmentation of data and data sources is a challenge. The interoperability of systems, and the abundance of systems and data stores necessitate attention. This is necessary to address the completeness and consistency of data. Data fragmentation appears as a root cause for deficiencies in data quality for completeness and consistency. Productized data assets were found to be a key enabler for AI applications and the underlying analytics use cases. The use of productized data assets would have an impact on data quality and the reusability of data assets. The implication of productized data assets, and specifically the reusability of data would be driving the cost reduction of AI modeling, and enabling faster AI model iterations, which would drive the quality of AI models.

As a follow-up research question to this study, it would be interesting to investigate how the root-causes of data fragmentation could be remedied. The research question would yield new knowledge, as it touches two fields, namely data governance and process management. There is little research regarding the transition from treating data as a process resource to treating data as an asset.

7. Acknowledgements

The authors would like to thank Case Company A for granting access to the material and making this case study possible. A warm thank you is also extended to the people who have answered the questionnaires and were available for the interviews.

8. References

[1] X. Zhou, Z. Yang, M.R. Hyman, G. Li, Z.H. Munim. "Guest editorial: Impact of artificial intelligence on business strategy in emerging markets: a conceptual framework and future research directions", International Journal of Emerging Markets, 17.4 (2022): 917-929.

[2] D. Zhdanov, S. Bhattacharjee, M.A. Bragin. "Incorporating FAT and privacy aware AI modeling approaches into business decision making frameworks", Decision Support Systems, 155 (2022): 113715.

[3] M. Kim, T. Zimmermann, R. DeLine, A. Begel. "Data Scientists in Software Teams: State of the Art and

Challenges," IEEE Transactions on Software Engineering, 44.11 (2018): 1024-1038.
[4] S. Makridakis. "The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms, " Futures, 90 (2017): 46-60.
[5] K. Pahwa, S. Chauhan, Big Data and Machine Learning in Healthcare: Tools & Challenges in 3rd International Conference on Advances in Computing, Communication Control and Networking, ICAC3N, Greater Noida, India, 2021, pp. 326-330.
[6] M.H. Jarrahi. "Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making," Business Horizons, 61.4, (2018): 577-586.
[7] Y. Duan, J.S. Edwards, Y.K. Dwivedi. "Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda," International Journal of Information Management, 48 (2019): 63-71.
[8] A.K. Thakur. "Artificial Intelligence (AI) in Information and Communication Technology (ICT): An Overview," International Journal of Research and Analysis in Science and Engineering, 1.3 (2021): 34-40.
[9] W.M.P. van der Aalst, M. Bichler, A. Heinzl. "Robotic Process Automation. Business & Information Systems Engineering, 60 (2018): 269–272.
[10] C. Tornbohm. (2017) Gartner market guide for robotic process automation software. Report G00319864. Gartner
[11] S. Madakam, R.M. Holmukhe, R.K. Revulagadda. The Next Generation Intelligent Automation: Hyperautomation. Journal of Information Systems and Technology Management, 19 (2022): e202219009.
[12] LV. Herm, C. Janiesch, H.A. Reijers, F. Seubert, From Symbolic RPA to Intelligent RPA: Challenges for Developing and Operating Intelligent Software Robots, in: A. Polyvyanyy, M.T. Wynn, A. Van Looy, M. Reichert (Eds.), Business Process Management (BPM 2021), Lecture Notes in Computer Science, Springer, Cham. 12875 (2021): 289-305.
[13] H. Hannila, A. Tolonen, J. Harkonen, H. Haapasalo. "Product and supply chain related data, processes and information systems for product portfolio management," International Journal of Product Lifecycle Management, 12.1 (2019): 1-19.
[14] R. Silvola, A. Tolonen, J. Harkonen, H. Haapasalo, T. Mannisto. "Defining one product data for a product," International Journal of Business Information Systems, 30.4 (2019): 489–520.
[15] S. Wings, J. Harkonen. "Decentralised or centralised management of data and products: influence on revenue-generating processes," International Journal of Management and Decision Making, 22.1 (2023): 74–105.
[16] P. Aiken, J. Billings, Monetizing Data Management. Unlocking the Value in Your Organization's Most Important Asset, Technics Publications, LLC USA, 2013.
[17] H. Hannila, S. Kuula, J. Harkonen, H. Haapasalo "Digitalisation of a company decision-making system: a concept for data-driven and fact-based product portfolio management," Journal of Decision Systems, 31.3 (2022): 258-279.
[18] B. Otto, Y.W. Lee, I. Caballero. "Information and data quality in networked business", Electronic Markets, 21.2 (2011): 79–81.
[19] R. Silvola, J. Harkonen, O. Vilppola, H. Kropsu-Vehkapera, H. Haapasalo. "Data quality assessment and improvement, International Journal of Business Information Systems, 22.1 (2016): 62–81.
[20] J.S. Arlbjørn, C.Y. Wong, S. Seerup. "Achieving competitiveness through supply chain integration," International Journal of Integrated Supply Management, 3.1 (2007): 4–24.
[21] L. Bertossi, F. Geerts. "Data Quality and Explainable AI," Journal of Data and Information Quality, 12.2 (2021): 1-9.
[22] A. Silberschatz, H. Korth, S. Sudarshan, Database System Concepts, Higher Education Press. Beijing, 2006.
[23] L. Cai, Y., Zhu. "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era," Data Science Journal, 14 (2015): 1-10
[24] V. Simard, M. Rönnqvist, L. Lebel, N. Lehoux. "A General Framework for Data Uncertainty and Quality Classification",

IFAC-PapersOnLine, 52.13 (2019): 277-282.
[25] A.S.H. Al-MSloum, I.M. Alharbi. "Application and trends in information management system using artificial intelligence", Materials Today: Proceedings, (2021): https://doi.org/10.1016/j.matpr.2021.03.308
[26] S. Anagnoste, Robotic Automation Process - The operating system for the digital enterprise Proceedings of the International Conference on Business Excellence, 12.1 (2018): 54-69.
[27] C. Batini, C. Cappiello, C. Francalanci, A. Maurino, "Methodologies for data quality assessment and improvement," ACM Computing Surveys, 41.16 (2009): 1-52.
[28] F. Sidi, P.H.S. Panahy, L.S. Affendey, M.A. Jabar, H. Ibrahim, A. Mustapha, Data quality: A survey of data quality dimensions, in 2012 International Conference on Information Retrieval & Knowledge Management, Kuala Lumpur, Malaysia, (2012): 300-304.
[29] L.L. Pipino, Y.W. Lee, R.Y. Wang. "Data Quality Assessment", Communications of the ACM, 45.4 (2002): 211–218.
[30] R. Vilminko-Heikkinen, S. Pekkola, Establishing an Organization's Master Data Management Function: A Stepwise Approach, in 46th Hawaii International Conference on System Sciences, Wailea, HI, USA, (2013): 4719-4728.
[31] Y.K. Dwivedi, L. Hughes, E. Ismagilova, G. Aarts, C. Coombs, T. Crick, Y. Duan, R. Dwivedi, J. Edwards, A. Eirug, V. Galanos, P.V. Ilavarasan, M. Janssen, P. Jones, A.K. Kar, H. Kizgin, B. Kronemann, B. Lal, B. Lucini, R. Medaglia, K. Le Meunier-FitzHugh, L.C. Le Meunier-FitzHugh, S.Misra, E. Mogaji, S.K. Sharma, J.B. Singh, V. Raghavan, R. Raman, N.P. Rana, S. Samothrakis, Ja. Spencer, K.Tamilmani, A. Tubadji, P. Walton, M.D. Williams. "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," International Journal of Information Management, 57 (2021): 101994.
[32] E.G. Sands "How to build great data products", Harvard Business Review. October 30, 2018.
[33] S.A. Olaleye, A.G. Adusei, Data economy through productization: a conceptual paper, in Twenty Ninth World Business Congress, JAMK University of Applied Sciences Jyväskylä, Finland. in Eds. E. Kaynak, T.D. Harcar. Advances in global business, vol 29.
[34] C.S. Langdon, R. Sikora, Creating a Data Factory for Data Products. In: et al. Smart Business: Technology and Data Enabled Innovative Business Models and Practices. WeB 2019. Lecture Notes in Business Information Processing, 403. (2020), Springer, Cham.
[35] R., Eichler, C. Gröger, E. Hoos, H. Schwarz, B. Mitschang, From Data Asset to Data Product – The Role of the Data Provider in the Enterprise Data Marketplace. In: Barzen, J., Leymann, F., Dustdar, S. (eds) Service-Oriented Computing. Summer SOC 2022. Communications in Computer and Information Science, 1603 (2022), Springer, Cham.
[36] T. Fisher, The Data Asset: How Smart Companies Govern Their Data for Business Success, Wiley, Hoboken, NJ, USA, 2009
[37] W. Yanlin, Z. Haijun, Data Asset Value Assessment Literature Review and Prospect. Journal of Physics: Conference Series, 1550 (2020): 032133.
[38] X. Liu, Y. Jiang, L. Pan, "Application of Data Assets," 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, China, 2018, 2414-2418.
[39] J. Harkonen, H. Haapasalo, Hanninen, K., "Productisation: A Review and Research Agenda", International Journal of Production Economics, 164 (2015): 65–82.
[40] U. Kruhse-Lehtonen, D. Hofmann, How to Define and Execute Your Data and AI Strategy. Harvard Data Science Review, 2.3 (2020): 1-9.
[41] B.H. Wixom, G. Piccoli, J. Rodriguez, Fast-track data monetization with strategic data assets, MIT Sloan Management Review, 62.4 (2021): 1-4
[42] R. Merton, M. Fiske, P. Kendall, The Focused Interview: A Manual of Problems and Procedures, 2nd Ed., Free Press, New York, 1990.