

## MACHINE LEARNING-DRIVEN DATA QUALITY MANAGEMENT IN SAP MDG

*Arun Chinnannan Balasubramanian*  
*Master data management for Supply Chain and Finance*  
*Verizon Communications,*  
*Basking Ridge, USA*  
*arun.chinnannanbalasubramanian@verizonwireless.com*

---

### *Abstract*

*The exponential growth of data in the digital era has underscored the critical importance of maintaining high-quality master data within enterprises. SAP Data Quality Management (DQM) offers a comprehensive suite of tools designed to ensure data integrity and provide actionable insights through data mining and trend analysis. This paper delves into the functionalities of SAP DQM, emphasizing its capabilities in data profiling, trend analysis, and predictive modeling. By examining applications. Also illustrate how organizations can leverage SAP DQM to enhance master data quality and drive informed business decisions.*

*Keywords* DQM, SAP MDG, Data Quality, Data Mining, Data Quality Trend Analysis, Rule management, BRF+. Machine learning, Rule mining, PAL, KORD, K-Optimal.

## **I. INTRODUCTION**

### **1.1 Background**

Master data serves as the foundational information shared across organizational processes, such as customer, product, business partner (Customer and Vendor), and Financial master data(Like Profit center, Cost center, GL and Hierarchies). High-quality master data directly impacts operational efficiency and decision-making. However, maintaining its quality is challenging due to the complexity and volume of modern datasets.

SAP DQM with Machine Learning capabilities provide tools that automate data validation and offer deep insights into data patterns and quality trends, ensuring consistency and accuracy. Data Stewards and Information Stewards has the need to constantly enhance and elevate the data standards according to the organization needs and standards.

### **1.2 Problem Statement**

Key challenges faced by organizations include:

- Inaccuracies and inconsistencies in master data leading to operational inefficiencies.
- Difficulties in maintaining high data quality across centralized or decentralized systems.
- Lack of visibility into data quality trends over time.

This paper will walk you through an advanced way to orchestrate a quality domain that can sustainably upkeep the Master Data towards its goldenness. Rule management that sets the base for further rule mining through Machine Learning. Thereby also paves a way to permanently

monitor the quality trends and make the informed decisions by Data stewards.

### 1.3 Literature review

**1. Efficiency of SAP MDG DQM:** DQM incorporates advanced tools for rule-based validation, automated workflows, and real-time monitoring. These features allow organizations to define quality dimensions and rules tailored to their needs, ensuring data compliance with established standards. Integration with SAP HANA Predictive Analytics Library enhances its anomaly detection capabilities, enabling proactive data quality management. According to SAP's resources, these functionalities improve operational workflows, reduce redundancies, and support strategic decision-making [3].

Case studies have demonstrated the tangible benefits of SAP MDG DQM. For instance, organizations implementing the platform observed significant improvements in data quality, consistency, and efficiency. Interactive dashboards and quality analytics provide real-time insights, further bolstering decision-making processes.

**2. Integration with Enterprise Systems:** The seamless integration of SAP MDG DQM with other SAP solutions, such as SAP Analytics Cloud and SAP Data Intelligence, is a cornerstone of its efficiency. This integration enables organizations to use advanced analytics tools for monitoring and improving data quality. By visualizing trends and identifying anomalies, enterprises can implement comprehensive data governance strategies to ensure consistency across their systems.

SAP DQM is a powerful toolset that addresses the challenges by enabling:

- Automated data profiling and validation after the rule setup.
- Insights into data quality trends using analytics and ML.
- Predictive modelling for proactive issue resolution.

## II. DEFINITION AND CORE FEATURES OF DQM, FOR RULE MINING AND ADAPTION

SAP DQM is a solution designed to define, monitor, analyze, and improve data quality within enterprise systems. Its key features include the following in a broader perspective. Rule setup and repository. Rule mining with ML, Data quality trend analysis and adaptation.

**1. Rule setup and repository:** DQM rule management, leverage a centralized rule repository to collaboratively define, catalog, and implement data quality rules.

The MDG rule repository offers a unified platform for managing master data quality rules effectively. It allows you to document and catalog rules comprehensively, including detailed descriptions, business relevance, and associated stakeholders. With lifecycle management tools, such as rule statuses and collaboration capabilities powered by SAP CoPilot, teams can work together seamlessly to refine and manage rules. Moreover, the repository integrates with BRFplus, enabling direct implementation of rules while ensuring full visibility into their business impact, usage scenarios, and technical execution. This approach enhances transparency and efficiency across all aspects of rule management.

An example rule for product master is setup, that has the base table as KSSK and has some pre-conditions set by the user. Once this rule is set up and approved, it can be adapted to Change

request, Mass consolidation or both. The actual rule set will have to be defined which will internally invoke Business Rules Framework(BRF+). This is a rule engine framework to dictate the rules in a less-code format. Once set up the rule can be monitored for its performance.

Figure 1, has a conformity score of 83.1 for a rule set for product on whether that product has to be challenged for kit material based on the conditions mentioned in the business reason, This also which tells the rest of the percentage is not adhering to the rule or an outlier.

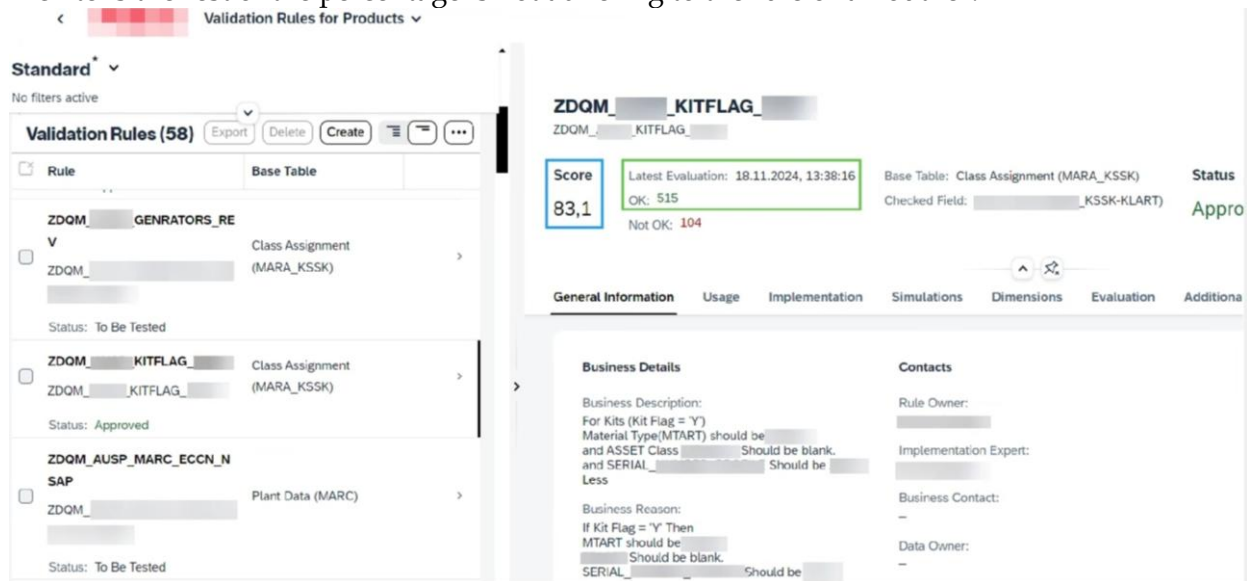


Figure 1: Validation rule repositories on the left & relative Evaluation scores on the right side.

Figure 2, gives an overview of a Rules configuration cockpit, with scope expression and condition expression. Rule expression is set in the same rule definition with other administrative details.

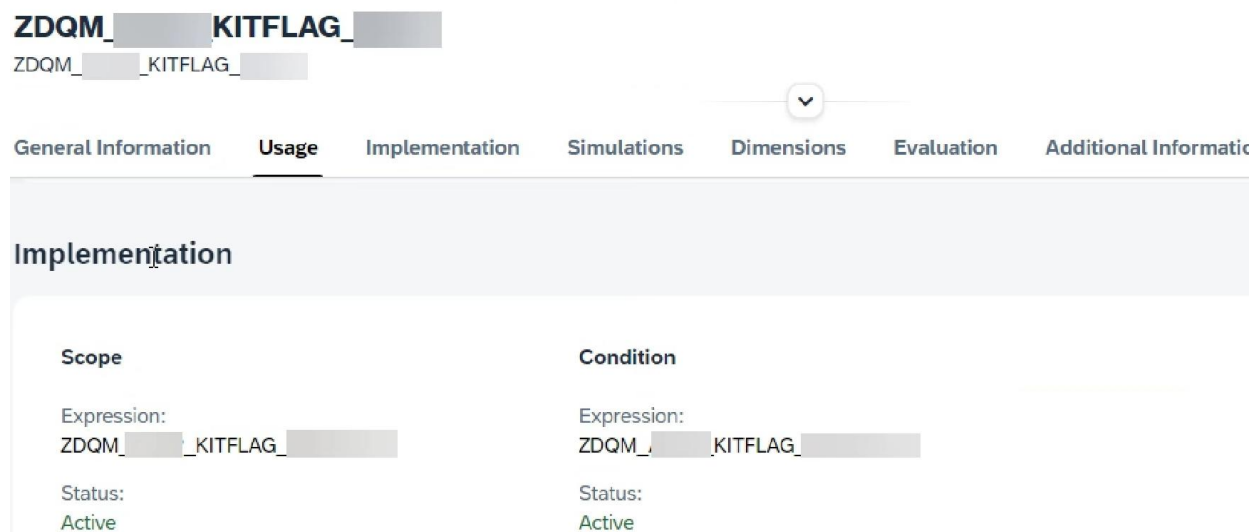


Figure 2: Rule with scope and condition expression.

**2. Rule Mining with ML:** DQM integrated capability of Rule mining can be one of ways to unearth the data gaps in an environment like multiple disparate systems. Mining can bring out the master data or the process with identifying patterns, interdependent tasks, and inconsistencies of a

master data. To achieve this advanced capabilities like SAP Hana Predictive analytics libraries and Machine learning(ML) are used [4]. Rule repository and its usage is mined with these above said anomalies with various algorithms to detect duplicates, associative analysis, missing dependencies, unidentifiable outliers. These identifiable anomalies can then be turned to an actionable work item for data stewards(DS) review.

**3. Data quality trend analysis:** Technically, DQM trend analysis relies on historical quality scores to uncover patterns and predict risks. Aggregate scoring tracks changes over time, helping to identify recurring problems and systemic issues. For example, a drop in address accuracy scores over several cycles might indicate inefficiencies in data entry processes, prompting targeted actions like rule enhancements or workflow improvements. Additionally, SAP DQM incorporates predictive analytics to anticipate future risks, enabling proactive measures to address potential issues before they affect operations. These capabilities make DQM an indispensable tool for interpreting and improving master data quality systematically, empowering organizations to achieve higher data governance standards [6].

Figure 3, is the data quality evaluation dashboard for products master. It scored overall 92.5. The trend of the score has been declining since October. This trend helps to find the reason, if it's because of any major event, like new projects that spiked that quality trend. If we look at the 'incorrect product data', it's sorted by the 'Cross plant status' to show which materials are most impacted. Approved & Inventoried materials are highly impacted. We can actually drill down to the next level by clicking the bar, which can lead to what ruleset contributed more.

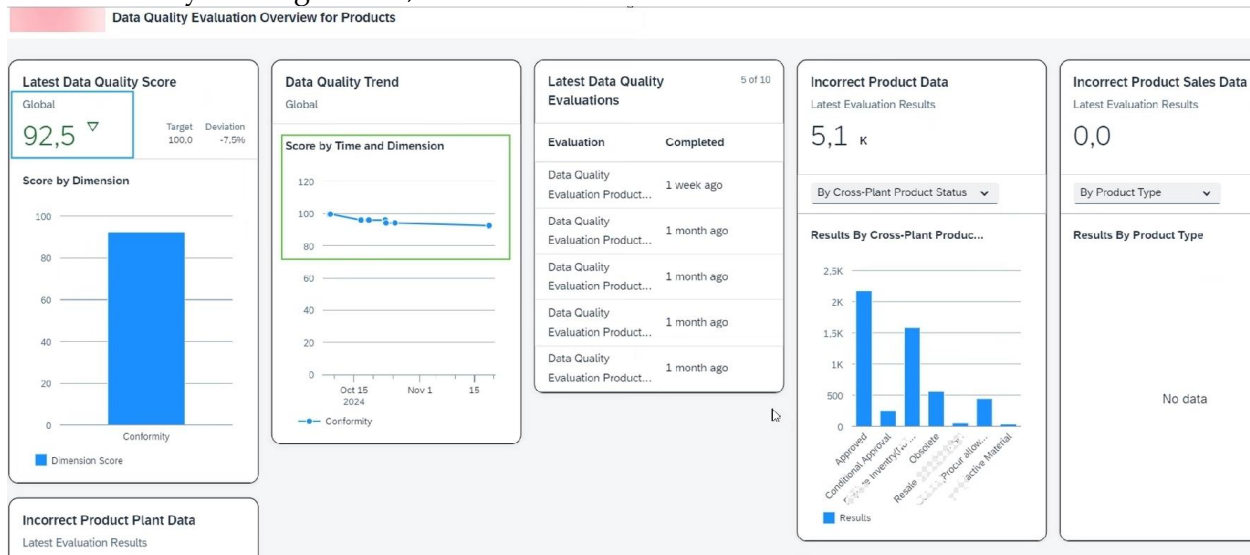


Figure 3: Data quality evaluation for the Products.

Below trend shows the overall product for one year., Likewise trends can be viewed by Rule, Plants, dimensions and much more. Thereby gives any cross-sectional view of a data.

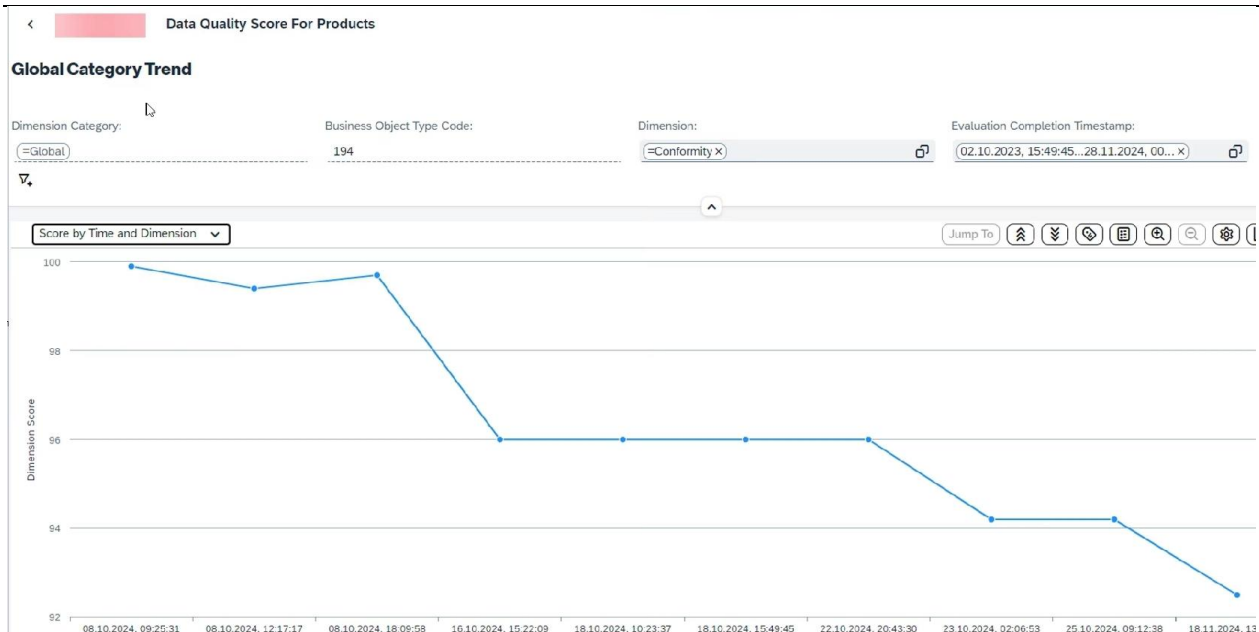


Figure 4: Data quality trend by timeline and dimensions.

Figure 5 shows the Products performance by manufacturer & Cross-plant status. Each of the rule ID can be inspected on the adherence level. Evaluations can be drilled down by different criteria.

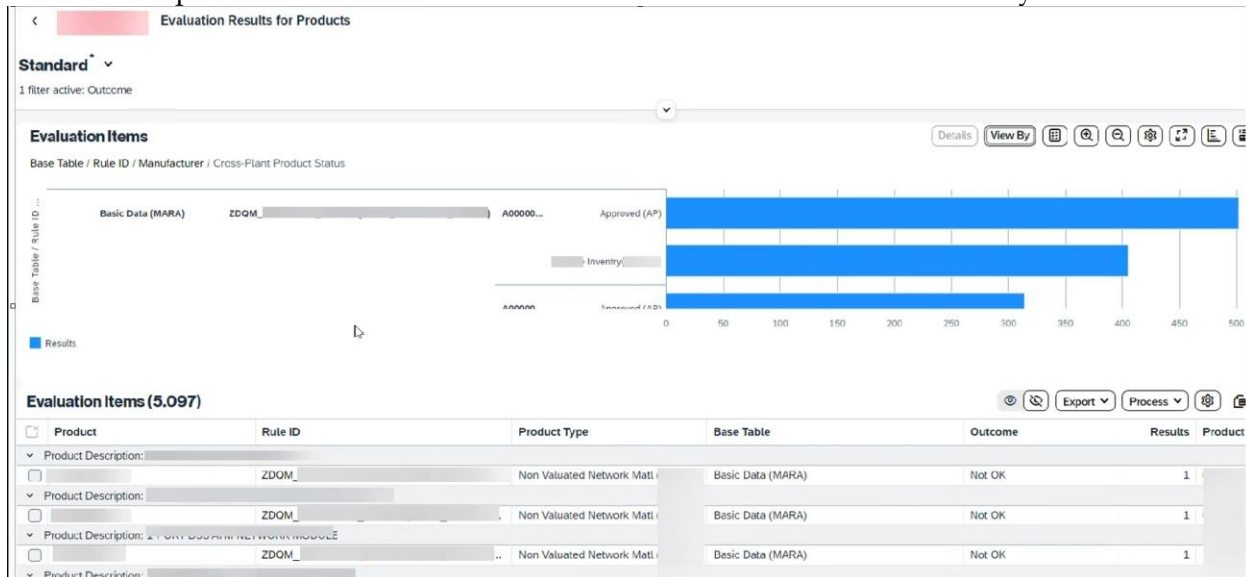


Figure 5: Evaluation drill down by different criteria.

### III. RULE MINING TECHNIQUES IN SAP DQM

DQM integrates Rule mining capabilities to enhance master data quality by uncovering hidden patterns, dependencies, and inconsistencies within datasets. Leveraging advanced technologies like SAP HANA Predictive Analytics Library (PAL) and machine learning, SAP MDG DQM identifies implicit rules and relationships, such as common data attributes or frequently occurring anomalies, which can then be used to inform rule management and validation processes. Data

mining techniques in SAP MDG include clustering, association analysis, and outlier detection to analyze large volumes of data and detect duplicate records, inconsistent hierarchies, or missing dependencies. These insights are crucial for building robust validation rules and workflows that proactively address data quality issues. For example, in a business partner context, mining might reveal that customers from a specific region consistently use a particular tax code, prompting the creation of a derivation rule to ensure consistency. With these capabilities, SAP MDG DQM transforms raw data into actionable intelligence, enabling organizations to maintain high-quality master data and drive operational excellence.

### 1. Algorithms:

SAP HANA PAL provided algorithms are classification (Gradient Boosting Classifier, Auto Classifier), clustering (Auto Supervised Clustering, Auto Unsupervised Clustering), regression (Gradient Boosting Regressor, Auto Regressor), time series (Auto Time Series).

DQM uses the K-Optimal Rule Discovery(KORD) to find the outliers, data match, Rule inconsistencies. KORD follows the idea of generating association rules with respect to a well-defined measure, instead of first finding all frequent itemsets and then generating all possible rules.

Figure 6, 'Process flow of Rule mining and adaptation' explains the lifecycle. Rule mining is an extension of rule management and It is a prerequisite.

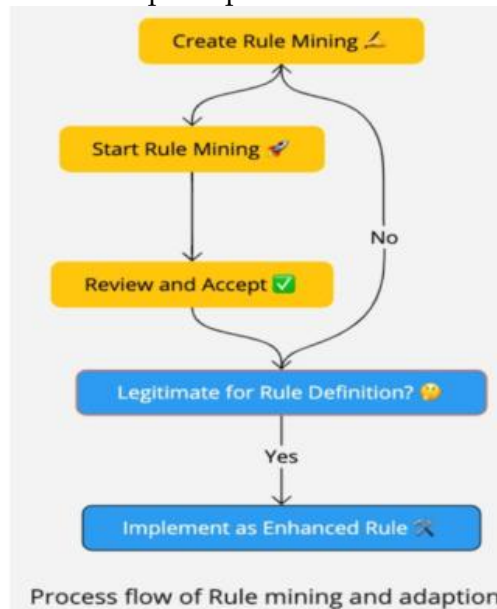


Figure 6: Process flow of Rule mining

### 2. Field selection for Filters in Rule Mining.

When you define a run for rule mining you use filters to determine the data to be analysed. SAP delivers a predefined set of fields that you can use for these filters. If the delivered settings do not fit to your requirements you can change these settings in the following Customizing activity.

### 3. Customizing path:

Cross-Application Components > Processes and Tools for Enterprise Applications > Master Data

---

Governance > Classic Mode in SAP MDG Consolidation and Mass Processing > Configure Process Models and Field Properties > Specify Table and Field Properties.

Predictive models in SAP DQM leverage machine learning to forecast potential data quality issues. For example, by analysing historical patterns, organizations can predict the likelihood of incomplete supplier records and take preventive actions.

#### IV. TREND ANALYSIS

Trend analysis tracks data quality metrics over time to identify patterns and measure the effectiveness of governance initiatives. SAP DQM dashboards enable users to monitor key indicators like data completeness and consistency.

During the data quality evaluation process, the system records the results of applying predefined rules to the data. It also computes scores for each rule, representing the proportion of valid data within the dataset. Multiple rules can be organized into data quality dimensions, which are further categorized under broader data quality categories. This hierarchical structure enables the aggregation of rule-level scores into higher-level KPIs, facilitating comprehensive data quality reporting. The system provides overview pages for monitoring the current state of data quality and analysing trends. These reports include drill-down capabilities, allowing users to trace issues to specific dimensions or rules and identify the root causes of data quality issues effectively [2].

#### V. METRICS AND KPIS

SAP DQM monitors these foundational metrics, which will churn quality turnout for any organization.

- **Completeness:** The percentage of records with all required fields populated.
- **Accuracy:** Alignment of data with real-world conditions.
- **Consistency:** Uniformity of data across systems.
- **Timeliness:** Frequency of updates to reflect current information[1].

#### VI. CHALLENGES AND LIMITATIONS

Only a heterogeneous system can have an easier way of implementing this quality control through DQM. If the organization ecosystem contains other legacy products, even that has to be plumbed. Also the Data stewards have to be trained to make use of these tools. Sometimes that is even challenging, as it could not be the core function of the user.

SAP DQM could incorporate generative AI to identify hidden patterns and anomalies. Recent innovations in SAP MDG include the integration of generative AI for assisted changes and summarization of changes in central governance. This feature provides users with natural language summaries of data modifications, enhancing transparency and facilitating efficient data management [5]. But the organization should grow to the level of adoption.

## VII. CONCLUSION

SAP MDG DQM has proven to be an effective solution for managing master data quality challenges. Its rule-based validation, integration with analytics tools, and anomaly detection capabilities enable organizations to achieve higher operational efficiency and better data governance. Although challenges remain in its implementation, technological advancements continue to enhance its value as a critical tool for enterprise data management. Value propositions for SAP DQM implementation are as follows.

- The DQM by itself sets the agenda for fine tuning the gaps, by providing the score cards. It becomes easier for Data Stewardship to focus on where to attend, rather pulling reports & analysing the facts.
- Rule definition need not take a long strained coding or deployment. It's more of a framework that can integrate to the Change request process of a master data entity.
- Data stewardship nowadays has become a perilous task with vast amounts of dynamic data. Businesses have changed operating their models, like for example, supply chain of manufacturing companies have to revisit their logistics every few years to meet the customers demand. Either 3PL segregation, same day delivery, importing hazardous products, all these lie on a massive set of master data. Unless an organization keeps this golden, it has to fight other operational difficulties.
- Leveraging the advanced ML, rule derivations and trend analysis is a must. Also these platforms are future-ready with AI and Gen-AI capabilities to be adapted.
- The results of DQM outweigh the struggles on the path of implementation of this kind of system in any landscape. If the organization is having more heterogeneous systems that integrate with complex businesses, these kinds of tools are the only way out.

## REFERENCES

1. SAP SE. "SAP Data Quality Management Overview". 2021.
2. Andreas Seifried. "Master Data Quality Management with SAP Master Data Governance on SAP S/4HANA 1909". 2019.
3. SAP SE. "Data Quality Management with SAP Master Data Governance on SAP S/4HANA." product information, pp. 18. 2023.
4. Kefw. "Deep Insight into Your Master Data with MDG Data Quality Analytics 2.0". In Oct 2023.
5. Mark "Benefit From SAP Master Data Governance on SAP S/4HANA 2023 Feature Pack Stack 02". Oct, 2024.
6. SAP AG. 'Data Quality Management with SAP Master Data Governance on SAP S/4HANA', product information, pp. 29-30. Oct 2023.