International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246.IJETCSIT-V5I2P117
Eureka Vision Publication | Volume 5, Issue 2, 159-163, 2024

Original Article

Ontology and Future of Master Data Management in Respect to Data Quality

Susmit Sen
Independent Researcher and Industry SME.

Abstract - Companies are producing large volumes of data due to digital technologies, cloud computing, and AI usage. Big data management is becoming a problem for small and medium-sized businesses too. Master Data Management ensures platform-wide corporate data consistency. Regardless of their size, organisations must manage increasingly sophisticated electronic business data and operations. Complex entity connections are difficult to represent in traditional data. Enterprise information systems require technology-based capabilities to simplify and comprehend company data and procedures. System-wide data integration and understanding are promoted by semantic ontology frameworks. Three information system use cases that might be enhanced are given, along with semantic technology and ontology-based techniques. Ontology-driven Master Data Management improves data quality, semantic integration, and intelligent data governance in contemporary businesses. Trustworthiness and data quality are guided by provenance.

Keywords - Ontology, Master Data Management (Mdm), Data Quality Management, Semantic Data Integration, Knowledge Graphs, Data Governance.

1. Introduction

Current information systems, cloud platforms, and related technologies are used across industries, producing more digital data globally. Electronic data management and business procedures improved firms by automating product purchases and sales. By 2025, the International Data Corporation expects over 175 zettabytes of global data [1]. Thus, firms' information systems usually solely handle product and business process data. As companies grow, they require robust data management solutions to arrange and maintain correct information. However, managing more electronic data and procedures becomes increasingly complicated for organisations.

MDM standardises key company data across departments and digital platforms. Many debate this topic in terms of big data. These systems manage financial, customer, supplier, product, and personnel data. Large and complicated data collections that typical data applications cannot handle are called "big data". Gartner revealed that 60% of companies make bad judgements due to inconsistent, duplicated, or inadequate corporate data. Not all corporate information systems can handle big data management.

Ontology frameworks meaningfully explain data element connections. These frameworks allow systems evaluate data, not store it. Electronic data management and business procedures improved firms by automating product purchases and sales. This article discusses how ontology and semantic technologies enhance MDM and business data quality. Therefore, corporate information systems generally primarily handle product and business process data as data.

2. Master Data Management and Data Quality

Master Data Management manages, maintains, and distributes vital business data across corporate platforms. However, managing more electronic data and procedures becomes increasingly complicated for organisations. In databases, applications, and cloud platforms, organisations store the same data in different formats [2]. Not all corporate information systems can handle big data management. Differences in data reduce operational efficiency and decision accuracy. For cost efficiency, enterprise resource planning systems reflect all company operations.

Industry studies estimate that 30% of company data in organisational systems is duplicated or inconsistent across databases. Product-marketing descriptions, photographs, and supporting technical product data are often missing from enterprise resource planning systems. Bad data causes operational errors and makes company decisions questionable. IBM says poor data quality costs US firms $3.1 trillion yearly. These data are usually maintained by other information systems, such as product data management systems.

Data quality involves accuracy, completeness, consistency, and timeliness. Master Data Management systems centralise enterprise data for precise analytics and reporting. These systems allow firms to store and utilise product data uniformly within and outside the company with many functionalities. In modern digital ecosystems, standard database models fail to depict complex entity relationships. Even if they can imagine the optimal options, decision makers may not make them.

Today's data-driven world requires high-quality data. Accurate data helps companies make choices, improve operations, and satisfy customers [3]. Understanding data quality's aspects is essential for managing it. The correctness, completeness, consistency, timeliness, relevance, and originality of data may be analysed. Completeness guarantees that all relevant data is accessible, while accuracy ensures that data matches reality. Uniformity and coherence across data sets are consistency goals. Timeliness and relevance guarantee data is fresh and useful. Uniqueness eliminates redundant data, preventing confusion.

Problems including data input mistakes, inconsistent formats, obsolete information, and data duplication make data quality difficult to manage. These challenges might result from human data input, technological restrictions, or multi-source data integration.

3. Ontology Concepts in Data Management

In an information system domain, an ontology describes knowledge via entities, attributes, and relationships. Product information management systems may help firms and workers classify products, translate them, manage media assets, and output data to diverse media. Data management ontology models explain how logical meanings connect data. Unstandardised, proprietary software- or manual quality assessment methods hinder reusability, interoperability, and extensibility. This approach allows computers analyse context instead of storing values. Quality management literature includes ontology-based investigations.

On the next level of semantic richness, taxonomy classifies words hierarchically. It uses super- and sub-relations to define word relationships [4]. These relations rank these concepts by generality. Thesauruses extend taxonomy. A thesaurus describes any word relationship. Topic maps are ontology-like models. Topic maps are abstract models and data formats for knowledge systems. Thesaurus-based connections define topic linkages. Additionally, occurrences may incorporate other papers within topic maps.

An enterprise database ontology framework may define customer, product, supplier, and transaction relationships. As product complexity and consumer demand for customisable items rise, firms must manage increasingly sophisticated product information. Systems automatically connect datasets using semantic correlations. Several ontology-based data quality management tools exist. Research shows that ontology-based integration strategies improve semantic data interoperability by 40% over database integration methods. Note that some of these solutions concentrate on certain domains and do not solve geographic data quality management concerns.

Ontology systems employ Resource Description Framework, Web Ontology Language, and SPARQL. It formalises user-defined quality criteria using a Data Quality Management language. Through organised knowledge graphs, these technologies help robots examine dataset links. Many firms cannot adopt such an information system into their IT structure because data and business process integration are too expensive and time-consuming. Ontology frameworks improve data classification and semantic coherence in data management systems [5]. The system's SfO construction for user rules and spatial quality dimensions is the primary differentiator.

4. Ontology Driven Master Data Management

Ontology-driven MDM systems structure company data by semantic linkages, not databases. Existing methods for managing data quality across domains are useful, but spatial data quality management solutions are needed. These systems define business entities and connect data sources logically. Data redundancy and discrepancies in an enterprise's product information might result. Remote digital environments may help businesses manage data. Thus, keeping, looking for, and displaying product information may be costly.

Semantic data management technology may improve data integration by 30%, according to Forrester Research. Ontology-based studies have been done for geographical data quality management. Master Data Management solutions use ontology frameworks to locate similar things across databases [6]. Any domain has rules to evaluate data against its specification rules. This function improves corporate data correctness and avoids duplicates. The authority, or constitution, is dynamic yet stable, an open-world regulatory structure.

Using the ontology, various semantic technologies that have not been used in complicated and big product data management settings may be used. Semantic technologies use ontology alignment, data quality evaluation, and business process integration.

This strategy has now become popular for annotating webpages with product data that search engines like Google and Yahoo can analyse. The semantic-based Aletheia architecture integrates organised and unstructured data. This strategy uses a service hub to share and alter data and separate certain and uncertain knowledge repositories. This effort converted BMEcat to GoodRelations. This method emphasises using the same data structures throughout the product life cycle. We provide a semantic master data management system reference architecture and research methodology.

Customer records in different databases may have different names. Therefore, syntactical and semantical constraints or defaults are needed to prevent duplicate statements or misinterpretations in complicated and/or massive data sets [7]. Ontology-based systems may use semantic links to identify these records. When constructing a customisable product, remember that picking one feature may exclude others. This improves corporate data governance and consolidation.

5. Impact on Data Quality

Organisational operational and strategic data quality involves accuracy and reliability. Other central-system-dependent controls. Poor data quality causes operational inefficiencies, reporting errors, and financial losses. More similar rules may be added to the system without affecting earlier handling. According to Experian, 91% of organisations experienced negative business outcomes from inaccurate or inadequate data [8]. Semantic Web and Open-World Assumption expertise are not anticipated of domain experts, according to the implementation.

Ontology frameworks improve data quality by connecting data objects consistently. Customers who wish to customise a product or get information require this information, as do staff who manage product data. Since semantic models standardise data definitions, departments may perceive information identically. Thus, complicated data must be captured, managed, and presented clearly. This united knowledge improves corporate communication and reduces misunderstandings. Finally, whether big data management systems for data and business process integration include user viewpoints has not been examined.

SDQO is the framework's core ontology, while SfOs are its specification ontologies. The objective of SDQO-based SfO design is a reusable, domain-independent framework. SfOs are intended to meet user needs and utilise class hierarchies for domain experts to update and manipulate. A GUI generates SfO ontologies to formalise rules [9]. CSV files are created from domain expert contributions. An SfO ontology is created by optimising this file.

SDQO includes spatial quality evaluation general classes. Specifications commonly establish class connections using spatial rules. For instance, "building" cannot overlap with "sea", "lake", their subclasses, or other buildings. Specifications use separate spatial rules for each relation. The framework optimises shared rules across classes using spatial linkages. This optimisation offers one-step efficient evaluation. Java libraries like OWLAPI map SfO and SDQO. SDQO ontology rules are implemented using relations, and quality evaluation follows. The suggested methodology generates data quality reports automatically.

A provincial basis map case study tests the framework. We detected all overlapping buildings, parcels, roads that pass over buildings, and buildings outside cadastral parcels in the classes without introducing ontology errors.

Ontology-based validation systems may discover inaccurate records because semantic linkages give data accuracy requirements. While "volume" refers to the continual growth of data to be processed, "variety" focuses on the rising number of data kinds. Ontology validation finds product records without supplier connections. "Velocity" refers to data production and processing speed to fulfil stakeholder needs. Studies suggest ontology-based governance systems reduce data errors by 35% [10]. A GUI allows domain experts or non-Semantic Web users to develop and modify Specification Ontology.

6. Integration with Artificial Intelligence

Modern data management systems employ AI, ontologies, and semantic knowledge graphs. Since data sets come from many sources and may not meet quality requirements, the International Business Machines Cooperation (IBM) recommends adding "veracity" to volume, variety, and velocity to quantify data dependability. Ontology structures may be automatically extended by AI patterns in massive datasets. Semantic technology may assist analyse complicated information and idea linkages by finding context. This reduces semantic framework maintenance human effort. Semantic technologies let people and robots comprehend and communicate data and complicated ideas.

These tools could automate big data management and the integration of different data sources into big data management systems by making data integration experts more efficient, error-free, and faster. This will reduce data integration resources compared to a less automated technique, which is unsuitable for large data management [11]. Ontology alignment has not yet been proved to be applicable to information systems for complicated and massive product data. Ontology matching complicated product information architectures should increase data automation and business process integration. This will also advance big data management and data-business process integration research.

Machine learning models may classify new data using ontology definitions. Data governance refers to the choices and responsibilities for data management operations, conducted according to stated models. Organisational data governance and validation improve with automated classification. This definition excludes data management activities to implement these choices. McKinsey Global Institute estimates AI-driven data management may increase operational efficiency by 20%. Several models define data governance, including stakeholders and their obligations.

AI systems can understand corporate dataset context and linkages using ontology knowledge graphs [12]. Semantic technology may use metadata to discover information and documents more effectively. This lets powerful reasoning systems identify irregularities in complex information situations. Additionally, metadata may be connected across data sources. This helps companies maintain data quality. However, formal metadata representation needs standardised criteria.

7. Future of Ontology Based Master Data Management

AI, distributed data governance, and semantic technologies will influence Master Data Management systems. These rules are needed for information systems, applications, and workstations to communicate metadata. Industry estimates predict 70% of enterprise data management systems will employ semantics and knowledge graphs by 2030. Therefore, the Resource Description Framework (RDF) may be used to describe online resources (meta-information). These technologies may help organisations manage massive structured and unstructured data. Due to more stakeholders and big data operations, its complexity has expanded in recent years. Ontology-driven Master Data Management systems will benefit from cloud computing's scalability storage and processing [13]. Accessing and managing data is challenging owing to their dispersal across organisations, corporations, and diverse sources. Businesses will integrate semantic data governance and cloud analytics. Data management is complicated by the multiplicity of sources and administrative contexts.

Existing solutions for complicated and massive product data integration and big data management involve manual tasks like decoding databases and assigning data fields and kinds. Ontology-based big data management will use semantic technologies to automate data integration and administration. This strategy reduces manual integration interference, reducing manual errors. Ontology-based big data management helps users/enterprises make decisions during integration by restricting their options.

Ontology frameworks will help digital platforms communicate because semantic links allow computers to understand common meanings. Ontologies frequently provide conceptualisations explicitly [14]. Healthcare, manufacturing, transportation, and banking require data exchange; therefore, this capability is vital. Ontology defines ideas and interactions between them, while specification representation offers a formal semantic of the specification.

Ontologies are promoted by semantic ontologies. The models will be discussed in order of increasing semantic richness, although none of them match ontologies.

Statisticians demonstrate that ontology-driven data governance improves data accuracy and efficiency. A glossary has the least semantic richness among these models. Master Data Management solutions will employ ontology and AI increasingly. Ontologies aid framework implementation. This will enhance business decision-making and intelligent data governance. A glossary lists terms alphabetically with meanings but no relationships.

8. Conclusion

Master Data Management systems help firms manage core data online. Therefore, implementing these frameworks is a challenging undertaking. Complex data entity connections are difficult to portray in typical database models. Ontologies are the most semantically rich knowledge representation paradigm. Data integration and interpretation are promoted by semantic ontologies. The models will be discussed in order of increasing semantic richness, although none of them match ontologies.

Statisticians demonstrate that ontology-driven data governance improves data accuracy and efficiency. A glossary has the least semantic richness among these models. Master Data Management solutions will employ ontology and AI increasingly. Ontologies aid framework implementation. This will enhance business decision-making and intelligent data governance. A glossary lists terms alphabetically with meanings but no relationships.

ACKNOWLEDGEMENT

I appreciate my friends' distractions and support when I felt like quitting up. This trip would not have been possible or entertaining without my university friends' chats and support, which have improved my research. Their presence and guidance have helped me understand my study matter.

References

[1] C. Batini and M. Scannapieco, "Data and Information Quality: Dimensions, Principles and Techniques," Springer International Publishing, 2016, doi: https://doi.org/10.1007/978-3-319-24106-7.
[2] D. Loshin, "Master Data Management," Morgan Kaufmann / OMG Press, 2009, ISBN: 978-0-12-374225-4.
[3] D. Brickley and R. V. Guha, "RDF Schema 1.1," W3C Recommendation, Feb. 2014. [Online]. Available: https://www.w3.org/TR/rdf-schema/.

[4] N. F. Noy and D. L. McGuinness, “Ontology Development 101: A Guide to Creating Your First Ontology,” Stanford Knowledge Systems Laboratory Technical Report KSL-01-05, 2001. [Online]. Available: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf.
[5] B. Stvilia and D. J. Lee, “Data quality assurance in research data repositories: a theory-guided exploration and model,” Journal of Documentation, Jan. 2024, doi: https://doi.org/10.1108/jd-09-2023-0177.
[6] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, “Methodologies for data quality assessment and improvement,” ACM Computing Surveys, vol. 41, no. 3, pp. 1–52, Jul. 2009, doi: https://doi.org/10.1145/1541880.1541883.
[7] T. C. Redman, “Data Quality: The Field Guide,” Digital Press, Boston, MA, 2001, ISBN: 978-1-55558-251-0.
[8] S. Karkošková, “Data Governance Model To Enhance Data Quality In Financial Institutions,” Information Systems Management, vol. 40, no. 1, pp. 1–21, Mar. 2022, doi: https://doi.org/10.1080/10580530.2022.2042628.
[9] Z. Pan, Y. Gao, F. Ponci, and A. Monti, “Semi-Automatic Ontology Development Framework for Building Energy Data Management,” IEEE Access, vol. 11, pp. 111991–112003, Jan. 2023, doi: https://doi.org/10.1109/access.2023.3323335.
[10] J. B. Correia, F. Rodrigues, N. Santos, M. Abel, and K. Becker, “Data Management in Digital Twins for the Oil and Gas Industry: beyond the OSDU Data Platform,” Journal of Information and Data Management, vol. 13, no. 3, Sep. 2022, doi: https://doi.org/10.5753/jidm.2022.2506.
[11] J. Martins, H. S. Mamede, and J. Correia, “Risk Compliance and Master Data Management in Banking – A novel BCBS 239 compliance action-plan proposal,” Heliyon, vol. 8, no. 6, p. e09627, Jun. 2022, doi: https://doi.org/10.1016/j.heliyon.2022.e09627.
[12] K. Beck et al., “Manifesto for Agile Software Development,” Agile Alliance, 2001. [Online]. Available: https://agilemanifesto.org.
[13] Arthur et al., “Process-Data Quality: The True Frontier of Process Mining,” Journal of Data and Information Quality, vol. 15, no. 3, pp. 1–21, Sep. 2023, doi: https://doi.org/10.1145/3613247.
[14] C. Stach, “Data Is the New Oil–Sort of: A View on Why This Comparison Is Misleading and Its Implications for Modern Data Administration,” Future Internet, vol. 15, no. 2, pp. 71–71, Feb. 2023, doi: https://doi.org/10.3390/fi15020071.