PROCEEDINGS OF THE TECHNICAL UNIVERSITY OF SOFIA, ISSN: 2738-8549, 2738-8530 VOL. 75, NO. 3, YEAR 2025

From Data Complexity to Decision Intelligence: Modeling and Quality Challenges in SAP Environments

Reneta Duhova

Abstract - **Organizations operating within modern SAP (Systems, Applications and Products in Data Processing) landscapes are increasingly confronted with a range of complex challenges arising from the diverse and fragmented nature of enterprise data. This data originates from a variety of sources, including transactional systems, legacy databases and unstructured formats. The diversity of these sources creates major challenges in integration, scalability, data quality and regulatory compliance. Addressing these issues is essential to enable effective and sustainable decision-making processes. This work explores current advancements in SAP technologies, such as SAP Datasphere, SAP S/4 HANA and the SAP Business Technology Platform (BTP), which offer powerful capabilities for semantic integration, real-time analytics and enhanced automation. Following a review of recent studies and official SAP innovations, the analysis highlights the importance of continuous data quality management and modular governance structures. It also emphasizes the growing role of artificial intelligence (AI) and machine learning (ML) within decision intelligence frameworks.**

***Keywords*— data complexity, decision intelligence, SAP Datasphere, data modeling, proactive governance**

I. Introduction

Effective management of enterprise data has become a central concern for organizations navigating digital transformation. SAP environments support complex and mission-critical business processes. Data modeling, integration and quality assessment become increasingly difficult due to the heterogeneity of data sources, rapid data growth and the need for real-time insights. The field has shifted from traditional model-centric paradigms to more data-centric perspectives. The quality, structure and accessibility of data play a decisive role in determining the effectiveness of downstream analytics and decision-making. [1], [2].

This literature review examines current research and practical approaches concerning data complexity in SAP environments. It explores foundational theories of data quality, architecture frameworks and modern approaches to data modeling within SAP's ecosystem, including platforms such as SAP HANA, SAP BW/4HANA and the emerging SAP Datasphere. The review emphasizes recent developments that introduce semantic modeling, federated architectures and business-aligned governance models, all of which aim to reduce fragmentation and improve the agility and reliability of enterprise data systems [3].

Moreover, the review identifies concerns including system interoperability, regulatory compliance and the alignment of technical and organizational roles in data stewardship. Particular attention is given to how modern technologies such as machine learning, artificial intelligence and advanced analytics are integrated into SAP environments to enhance data-driven decision-making [1], [3], [4]. The aim is to synthesize current knowledge, uncover methodological gaps and inform the development of optimized, sustainable data modeling strategies tailored for SAP-centric organizations.

II. Data Complexity In Enterprise Systems And SAP Landscapes

Modern enterprise systems are increasingly affected by the growing complexity of data arising from diverse formats, fragmented storage and the demand for real-time analytics. In SAP landscapes, data stems from multiple sources – ranging from transactional ERP systems (e.g., SAP S/4HANA), legacy databases, third-party APIs and unstructured operational data – leading to challenges in semantic consistency, modeling efficiency and governance alignment. These issues are amplified by regulatory obligations, such as GDPR and the EU AI Act, which require high levels of transparency, completeness and traceability [1], [3], [6].

Historically, SAP data modeling platforms such as SAP BW and SAP HANA relied on structured, schema-first approaches focused on extract-transform-load (ETL) pipelines. However, the shift to cloud-native architectures and hybrid environments has introduced new requirements for federated access, semantic integration and business-aligned data management.

A growing dimension of complexity in enterprise data landscapes arises from SAP's announcement for the retirement and end of maintenance for SAP Data Intelligence. This change, part of SAP's broader strategic evolution toward unified, cloud-native platforms, has compelled organizations to reconsider their data orchestration and machine learning integration strategies. While SAP has initiated a transition period to support existing customers, the shift signals a movement away from centralized pipeline execution and metadata management previously handled within SAP Data Intelligence.

Received: 05.10.2025
Published: 26.10.2025
https://doi.org/10.47978/TUS.2025.75.03.001

Reneta Duhova is with the Technical University of Sofia, Sofia, 1000, 8 Kl. Ohridski Blvd, Bulgaria (reneta.duhova@tu-sofia.bg)

In response, enterprises are increasingly turning to SAP Datasphere, which offers federated integration, semantic modeling and embedded governance capabilities suited for modern, scalable data management. [7] The abrupt transition has ex-posed gaps in architectural readiness, intensified dependency on SAP Datasphere and highlighted the need for adaptable data modeling frameworks capable of accommodating rapid tool lifecycle changes [7], [8].

SAP Datasphere introduces the concept of a business data fabric – an integrated approach to combining physical and virtual data sources while maintaining metadata integrity and domain-specific semantics [8],[9].

SAP Datasphere provides a modular environment for modeling, accessing and governing enterprise data. It enables both graphical and SQL-based modeling through features such as Analytic Models and Graphical Views, while offering isolated development and collaboration environments known as Spaces to support controlled and scalable data management [10]. These features contribute to a reduction in data silos, more agile data provisioning and improved alignment between business and IT.

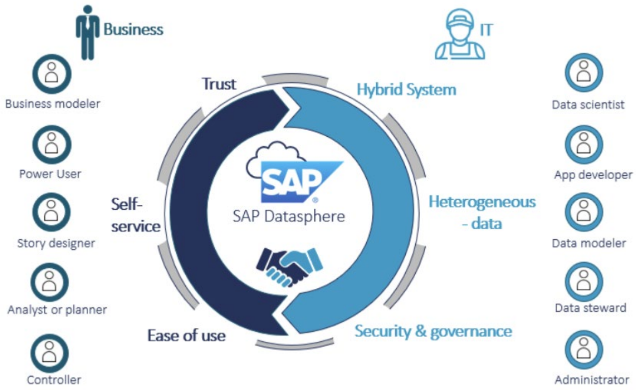

Fig. 1. SAP Datasphere environment.

Despite these advantages, several technical and organizational challenges remain. Studies report performance bottlenecks when modeling highly virtualized data scenarios, especially across distributed sources [11], [12]. Moreover, as noted by Rahman et al., organizations still face skill gaps and implementation complexity, particularly when transitioning from traditional ERP systems to cloud-native solutions like Datasphere [3], [13]. Nevertheless, SAP Datasphere represents a decisive shift towards a more flexible, interoperable and semantically enriched environment, addressing the core sources of complexity in modern SAP-based data ecosystems.

In summary, SAP Datasphere is an example of the evolution of SAP's response to data complexity by promoting a modular, real-time and semantically consistent approach to enterprise data integration and analytics. It supports the shift from isolated data silos to interconnected, governed and explainable data environments that meet the requirements of both business stakeholders and technical architects.

III. KEY CHALLENGES IN SAP DATA MODELING

Data modeling within SAP environments presents a unique set of challenges rooted in the architectural complexity of enterprise systems, the variety of integrated data sources and the evolving business and regulatory requirements. As organizations migrate from legacy SAP ERP systems to modern platforms such as SAP S/4HANA, SAP BW/4HANA, SAP BTP and SAP Datasphere, they must contend with architectural redesigns, data quality concerns, bottlenecks and the need for scalable solutions.

The deprecation of SAP DI has introduced new layers of technical and organizational uncertainty. Legacy pipelines orchestrated in DI must now be migrated to alternative SAP BTP components or restructured within Datasphere-compatible paradigms. This process often reveals hidden technical debt and insufficient governance structures, which were previously managed centrally within DI's metadata layer [7], [8], [13].

One of the most cited difficulties in SAP data modeling is integration complexity. Legacy systems often contain deeply embedded customizations and siloed logic, which are difficult to translate into the simplified, modular structures of modern SAP platforms. Migrating such models to SAP S/4HANA or Datasphere requires not only data trans-formation but also process re-engineering and business logic realignment – often with high risk and cost implications [3], [4], [13].

Semantic inconsistency and metadata fragmentation further complicate the modeling process. Inconsistent definitions of business terms, key figures and hierarchies across different modules and systems hinder cross-functional reporting and decision-making. SAP Datasphere aims to mitigate this through its semantic modeling layers (e.g., Analytic Models) but implementation remains challenging in multi-domain or multi-tenant environments where ownership of semantics is distributed [8], [14].

Another critical challenge is governance and data quality assurance. As highlighted by Mohammed et al., effective data quality assessment must consider multiple facets – data source, system, task context and human interpretation [1]. Without structured metadata management and data lineage tracking, SAP models can easily lose transparency and reproducibility, particularly when spanning hybrid on-premises and cloud architectures.

Furthermore, performance scalability is a recurring concern, especially when combining physical and virtual models within a single landscape. While Datasphere sup-ports real-time data federation, studies show that excessive reliance on virtual views over distributed sources can cause latency, inconsistent query behavior and memory overload in larger analytical scenarios [11], [12].

Skill shortages and change management often slow down modeling efforts. As the modeling paradigm shifts from rigid ETL pipelines to agile, semantic and self-service approaches, organizations require cross-functional teams capable of understanding both business logic and technical architecture. This dual requirement has created a talent gap, particularly for roles that bridge data architecture, governance and modeling in SAP-specific environments [3], [8], [14].

Successful data modeling in SAP systems requires more than technical tooling, it demands organizational alignment, process redesign, strong data governance frameworks and ongoing competency development. Addressing these challenges holistically is essential for leveraging the full potential of SAP's modern data platforms in support of decision intelligence and digital transformation.

TABLE I
EVOLUTION FROM TRADITIONAL ETL-BASED TO MODERN FEDERATED SAP MODELING

Dimensions	Traditional SAP BW / HANA (ETL-Centric)	Modern SAP Datasphere (Semantic & Federated)
Architecture	Layered, schema-based architecture centered on structured ETL workflows and centralized data warehousing.	Cloud-native, service-oriented design supporting federated access and semantic abstraction across heterogeneous sources.
Data Integration	Sequential ETL pipelines often lead to data replication and latency.	Virtualized integration via semantic joins and metadata mapping minimizing replication.
Governance	Governance applied after modeling, relying on manual stewardship and periodic validation.	Embedded, metadata-driven governance integrated across modeling layers and Spaces domains.
Data Quality	Reactive cleansing and limited lineage tracking.	Continuous quality control with rule-based validation and semantic tagging.
Ownership	Centralized within IT or BI units, limiting active business participation.	Decentralized, domain-based ownership enabling self-service modeling and collaborative stewardship.
AI / ML Integration	External, dependent on separate data science environments and manual orchestration.	Natively embedded through SAP HANA PAL/APL and AI Core for in-database learning and automation.
Scalability	On-premises scaling constrained by hardware capacity and vertical optimization.	Elastic, cloud-based scalability with dynamic resource allocation across hybrid environments.
Value Proposition	Operational stability, structured control, and consistency in reporting.	Agility, transparency, explainable analytics within governed and interoperable data landscapes.

IV. MODERN TECHNOLOGIES AND METHODOLOGIES FOR DATA MODELING

In response to increasing data complexity, SAP has continuously evolved its data modeling technologies, shifting from rigid, structured ETL pipelines to more agile, semantically enriched and cloud-native architectures. SAP data modeling practices are now centered around real-time processing, semantic layering and federated data access, aiming to provide a unified view of enterprise data while maintaining scalability, governance and business alignment.

One of the most impactful advancements in recent years is the introduction of SAP Datasphere, which integrates modern data modeling concepts such as graphical modeling, analytic models, semantic joins and business-layer abstraction. These tools allow users to create reusable, modular data models that incorporate both physical and virtual data from heterogeneous sources without replication [8], [11], [14]. Datasphere supports hybrid modeling, where SAP BW Bridge connects traditional BW models to the cloud, preserving investment in legacy logic while enabling new real-time applications [14].

Enterprises are increasingly incorporating modular decision intelligence frameworks that unify semantic modeling, embedded quality controls and AI components. The interplay between SAP Datasphere, SAP AI Core and strategic methodologies such as the SAP Data & Analytics Advisory Methodology offers a promising pathway to build resilient, explainable and business-aligned data architectures [15].

SAP has also adopted metadata-driven modeling methodologies, which rely on semantic tagging, data lineage tracking and domain-specific catalogs. These techniques enable self-service analytics by allowing business users to explore trusted data models without needing deep technical expertise [6], [9], [10]. The Spaces concept in Datasphere fosters decentralized data ownership and governance by assigning isolated development environments to specific business domains [9], [10], [14].

Complementing these developments is the SAP Data & Analytics Advisory Methodology, a strategic framework based on TOGAF (The Open Group Architecture Framework) and SAP EAF (Enterprise Architecture Framework), which guides enterprises through architecture design, capability mapping and governance alignment in four structured phases. The methodology emphasizes the importance of aligning business outcomes with data architecture and highlights best practices for data product design, use-case modeling and road mapping [6].

The methodology consists of four distinct phases (fig. 1):

- *Phase I: Scoping and Baseline Analysis*

This initial phase involves a comprehensive definition of the project scope and the articulation of specific objectives. A thorough assessment of the current data landscape, including existing architectures, systems and capabilities, is conducted to establish a baseline understanding. Furthermore, this phase identifies critical gaps, potential weaknesses and opportunities for enhancement within the organization's data environment.

- *Phase II: Business Outcomes and Solution Requirements*

In the second phase, the focus shifts to translating strategic business goals into clearly defined, measurable business outcomes. This involves a thorough analysis of relevant use cases and gathering of clear, detailed requirements for data products and solution architectures that effectively support the desired business outcomes. Additionally, this phase emphasizes the integration of proposed solutions within existing business processes and workflows.

- *Phase III: Capability Mapping and Solution Architecture*

The third phase involves the development of a capability map that correlates identified business needs with the requisite technical capabilities necessary to fulfill them. Various architectural alternatives are explored, assessed and refined to ensure alignment with strategic imperatives and operational feasibility. The outcome of this phase is the validation and formalization of a target data architecture that optimally supports the envisioned business and technical requirements.

- *Phase IV: Data Governance and Roadmap Development*

The final phase addresses the establishment of robust data governance frameworks, encompassing data quality management, security controls and compliance considerations. Clear specifications of roles, responsibilities and organizational structures for governance are articulated. This phase culminates in the formulation of comprehensive, actionable roadmaps and project plans that guide the phased implementation of data and analytics initiatives.

Collectively, these phases provide a disciplined pathway for organizations to transition from an as-is data state toward a strategically aligned, governed and value-driven data architecture and analytic capability.

In addition, ML and AI-powered augmentation are now embedded within SAP modeling platforms. SAP HANA supports Python- and R-based ML pipelines directly in the database engine, while SAP Analytics Cloud provides tools such as Smart Discovery and Predictive Modeling for automated insight generation and forecasting [16]. These capabilities enable the development of decision intelligence frameworks that combine semantic models, real-time data ingestion and explainable ML algorithms to support advanced analytics scenarios in SAP environments.

However, the adoption of these modern approaches requires cultural and operational shifts. Organizations must foster cross-disciplinary teams, integrate data governance early in the design process and establish clear policies for metadata management and model reuse. Without such foundations, even the most sophisticated modeling tools may fall short of delivering sustainable business value [2], [5], [6].

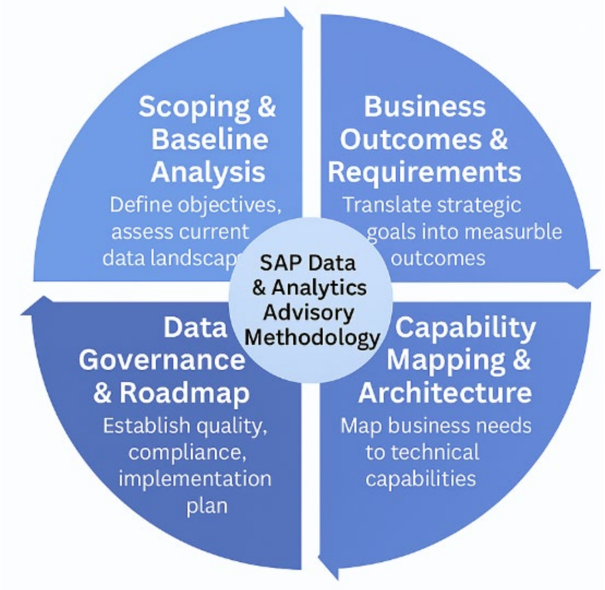

Fig. 2. SAP Data and analytics advisory methodology.

In summary, SAP's modern data modeling stack – spearheaded by Datasphere and enriched by advisory methodologies and AI components, represents a shift toward more flexible, business-driven data architectures. These innovations lay the groundwork for scalable analytics and effective data governance in complex enterprise landscapes.

V. Data Quality Assessment: Standards And Techniques

High-quality data is essential for the success of analytical systems, predictive modeling and digital transformation initiatives. In SAP environments where data is often distributed across multiple systems and domains, the complexity of maintaining data quality (DQ) is significantly increased. Ensuring accuracy, consistency, completeness and relevance of data becomes a critical requirement for generating reliable insights and complying with regulations such as GDPR and the EU AI Act.

Contemporary research suggests that effective data quality assessment must go beyond surface-level checks and instead consider the multidimensional nature of data in enterprise systems. Mohammed et al. propose a five-facet framework for DQ assessment, encompassing the data itself, the source, the system, the task and the human factors involved [1]. This holistic model allows organizations to evaluate data quality from both technical and organizational perspectives, considering lineage, storage architecture, user interaction and task-specific requirements.

In SAP systems, traditional DQ assessments have often been relegated to periodic data cleansing and validation within ETL pipelines or SAP Information Steward. However, with the shift toward semantic and federated modeling, as seen in SAP Datasphere, quality assessment must be embedded into the modeling process itself. Datasphere introduces metadata-driven governance, lineage tracking and business semantic tagging, allowing organizations to track data transformations across layers and evaluate DQ within context-specific use cases [2], [9], [10].

Datasphere's integration with tools like SAP Data Intelligence and Collibra enhances data governance through automated profiling, cataloging and rule-based validation [7], [9], [10]. These capabilities support the implementation of data quality policies directly within the platform by enabling continuous monitoring of key metrics such as accuracy, timeliness, completeness and conformity. Furthermore, Datasphere's Spaces structure provides isolated governance domains, which allow business and IT units to enforce quality standards independently while maintaining overall architectural coherence [14].

From a methodological standpoint, SAP also incorporates DQ into its enterprise advisory frameworks. In Phase IV of the SAP Data & Analytics Advisory Methodology, data governance and quality roadmaps are developed to align technical architectures with business goals, regulatory requirements and organizational roles [6]. This phase also recommends assigning data stewards and product owners to enforce accountability for quality across the data lifecycle.

Despite these advances, several practical challenges persist. Federated data remains harder to validate in real time, external source reliability is difficult to track and quality standards often vary across organizational units [1], [2], [12]. Addressing these challenges requires not only better tooling, but also cultural shifts toward proactive governance, cross-functional ownership and transparent documentation of modeling assumptions and rules.

In conclusion, modern SAP platforms have evolved to embedded data quality assessment into the data modeling lifecycle through semantic modeling, integrated metadata management and automated quality policies. The combination of methodological frameworks and platform-native governance capabilities offers a robust foundation for ensuring data trustworthiness in complex enterprise environments.

VI. Application of AI and ML in SAP Context

The incorporation of AI and ML within enterprise information systems is fundamentally transforming the methodologies by which organizations derive actionable insights, automate decision-making processes and identify anomalous patterns. Within SAP environments, the application of ML has evolved from isolated predictive models into embedded, scalable architectures that support real-time analytics and decision intelligence. These innovations are particularly relevant in the context of SAP Datasphere, which serves as a foundation for harmonizing structured, semi-structured and federated data to feed intelligent applications.

A major advancement is the utilization of in-database ML capabilities in SAP HANA, particularly through the Predictive Analysis Library (PAL) and the Automated Predictive Library (APL). These components enable users to execute classification, regression, clustering and time series forecasting directly within the HANA engine, thus reducing data movement and improving execution latency. These libraries are accessible via Python and R bindings, enabling seamless integration with data science workflows and real-time applications connected to SAP Datasphere [16].

To support more complex ML workflows and bring-your-own-model (BYOM) strategies, SAP Data Intelligence provides a powerful orchestration environment. It allows data scientists to train, deploy and monitor ML models built in TensorFlow, Scikit-learn, or PyTorch and connect them to SAP data pipelines. The platform offers graphical pipelines for ML training, inference and retraining - making it a natural complement to Datasphere in hybrid modeling scenarios [17], [18]. It also supports model versioning, metadata management and runtime scaling.

For cloud-native and enterprise-scale AI deployment, SAP AI Core – part of the SAP Business Technology Platform (BTP) – offers full ML Operations lifecycle management, including Kubernetes-based model deployment, integration with Git-based model repositories and multi-environment inference exposure. SAP AI Core connects to Datasphere and HANA as data sources and provides REST APIs and event-driven triggers to embed ML into business processes [18]. Its companion service, AI Launchpad, offers a unified dashboard for managing models, artifacts and lifecycle states.

Recent advancements show the potential of embedding ML-based audit models directly within SAP systems to support real-time fraud detection, anomaly classification and proactive internal control enhancement. For example, SoftMax regression models have been successfully applied to classify billing documents by audit risk dimensions such as suspicious pricing patterns, weekend processing, or destination-based red flags. Such use cases demonstrate how ML can bridge the gap between raw data and meaningful internal audit interventions [2], [19]. This positions SAP not only as a data platform, but also as a live risk-monitoring infrastructure that aligns with modern continuous auditing strategies.

Although SAP Analytics Cloud (SAC) remains a useful front-end for consuming AI insights through Smart Discovery and Predictive Forecasting, its modeling flexibility is more limited compared to the options enabled by HANA ML, Data Intelligence and Datasphere. In this context, SAC acts as a consumption and visualization layer, while the intelligence layer resides in the backend platforms that connect through Datasphere's semantic and physical integration.

Collectively, these technologies contribute to the development of Decision Intelligence frameworks – modular systems that integrate semantic modeling, explainable ML and near-real-time analytics. In SAP-centric environments, such frameworks rely on the federated and governed data access provided by SAP Datasphere [9], [10], combined with embedded ML in SAP HANA (PAL/APL) [25] and orchestrated ML pipelines via SAP Data Intelligence and AI Core [26], [27]. This integration enables advanced applications such as fraud detection, anomaly flagging, supply chain forecasting and compliance monitoring [2], [13].

Decision Intelligence frameworks constitute a sophisticated approach to data-driven decision-making by integrating contemporary technologies into modular, coherent systems. Within SAP-centric environments, these frameworks enhance decision quality, speed and transparency by leveraging key capabilities.

Semantic Modeling: SAP Datasphere provides a semantic layer through graphical and analytic models that unify heterogeneous data sources, ensuring consistent and accurate data interpretation across the enterprise.

Explainable Machine Learning (ML): These frameworks incorporate explainable ML models, embedded in SAP HANA or managed via SAP AI Core, which facilitate transparent and interpretable algorithmic decisions essential for trust and regulatory compliance.

Near-Real-Time Analytics: Support for timely decision-making is enabled by processing data streams and transactions nearly in real time, critical for applications such as fraud detection and anomaly identification.

Federated and Governed Data Access: SAP Datasphere's federated architecture allows querying distributed data without excessive replication, under strict governance and compliance frameworks including GDPR.

Applications of Decision Intelligence frameworks span fraud detection, anomaly flagging, supply chain forecasting and compliance monitoring, enabling proactive risk management and operational optimization.

Strategically, the modularity of these frameworks allows incremental scalability and adaptation. SAP Datasphere's governance combined with SAP's AI and ML tools enables transparent, business-aligned decision systems. These systems ensure continuous auditing, fraud prevention and regulatory compliance in complex enterprise landscapes.

SAP's AI and ML ecosystem is rapidly maturing, with a clear trend toward cloud-based, integrated and semantically aware architectures. SAP Datasphere plays a pivotal role as the unified data foundation that connects modeling, governance and intelligence across the enterprise landscape.

VII. Key Challenges and Research Gaps in SAP Data and Analytics Architecture

The digitization and evolution of enterprise architectures that are driven by cloud migration, real-time analytics and AI integration has exposed a set of enduring practical challenges alongside research imperatives within SAP environments. This summary points out important gaps that require careful study and attention from organizations:

1. *Complex Integration of Legacy and Cloud Systems*

Organizations continue to face substantial difficulties integrating heterogeneous legacy infrastructures with modern cloud-native platforms such as SAP Datasphere and SAP BTP. This integration complexity encompasses both technical interoperability and business process realignment, impeding unified data accessibility and analytics.

2. *Fragmented Metadata and Semantic Data Modeling*

The persistence of fragmented metadata management and inconsistent semantic standards across various data domains creates barriers to effective data governance, lineage tracking and model interoperability. Addressing semantic heterogeneity remains imperative for achieving coherent enterprise data fabrics.

3. *Scalability and Performance Constraints in Federated and Virtualized Models*

Real-time federation of distributed data sources within hybrid and multi-cloud SAP landscapes presents performance challenges including query latency, inconsistent behavior and memory bottlenecks. Research on optimizing and benchmarking federated modeling performance is notably lacking.

4. *Standardization Deficits in Model Quality and Explainability Assessment*

There is a pronounced absence of widely accepted standards and methodologies for evaluating the quality, robustness and explainability of semantic data models and AI/ML applications within SAP-native environments. This limits reproducibility and trust in data-driven decision systems.

5. *Underdeveloped Explainable AI (XAI) Frameworks in Enterprise Contexts*

The integration of explainable machine learning approaches within SAP ecosystems, particularly in sensitive domains such as auditing, compliance and risk management is insufficiently explored. Enhancing XAI capabilities is critical for transparency, accountability and regulatory adherence.

6. *Misalignment Between Data Governance and Decision Intelligence Architectures*

Frameworks that comprehensively align SAP's data governance mechanisms, including data quality controls, compliance protocols and stewardship responsibilities with AI-powered decision intelligence architectures, are presently inadequate. This gap undermines the development of automated decision-making processes that are trustworthy, auditable and compliant.

7. *Insufficient Automation in Real-Time Data Quality Management*

Although metadata-driven governance tools in platforms like SAP Datasphere offer promising outcome, their capacity for automated, scalable and context-aware data quality assessment in dynamic, distributed environments require further investigation and enhancement.

8. *Increasing Demand for Cross-Disciplinary Expertise*

The increasing complexity of modern SAP data landscapes demands the development of cross-functional competencies that integrate technical modeling expertise with AI governance and business strategy. The synthesis of these capabilities is recognized as a prerequisite for effective data stewardship and the advancement of sustainable analytics maturity [2], [6], [15].

This comprehensive analysis integrates prevailing industry challenges with emerging academic perspectives, offering a solid basis for future research and strategic efforts aimed at enhancing SAP modeling and analytics capabilities and governance frameworks within an increasingly dynamic digital environment for powering accurate and effective decision-making processes.

VIII. Future Work

Within the context of this research, the existing modular Decision Intelligence architecture tailored for SAP Datasphere and SAP Business Technology Platform (BTP) serves as a foundational reference model. This architecture integrates four core layers – semantic modeling, data quality governance, AI/ML capabilities and orchestration - designed to provide standardized, governed and explainable data environments supporting real-time intelligence, continuous auditing, fraud detection and smart compliance [18].

This literature review aims to address and scope future work focusing on enhancing proactive governance practices and audit mechanisms, process and data monitoring, specifically through the development and integration of advanced machine learning models aimed at flagging fraudulent activities within complex SAP data systems.

While this architecture provides a solid structural basis, future work should focus on incremental, achievable contributions that extend its practical utility in enterprise settings:

- *Targeted refinement of AI/ML approaches*

Future work should prioritize benchmarking explainability methods in concrete SAP fraud detection cases. It should further adapt lightweight anomaly-detection models for hybrid data environments to balance performance and transparency.

- *Strengthening governance frameworks through metadata-first practices*

Instead of proposing entirely new frameworks, future work could pilot early-stage metadata management policies within SAP Datasphere Spaces, evaluating their impact on cross-disciplinary collaboration and governance continuity, especially after the deprecation of SAP Data Intelligence.

- *Embedding audit mechanisms into orchestration workflows*

A realistic step is to design monitoring templates in SAP AI Core or orchestration pipelines that capture audit logs, exception handling and compliance flags in alignment with GDPR and EU AI Act requirements.

- *Operational strategies for semantic and lineage consistency*

Rather than attempting to resolve semantic inconsistency in full, future contributions should focus on developing repeatable guidelines and mappings (e.g., business term catalogues, lineage tracking prototypes) that can be validated across two or three SAP domain areas (finance, supply chain, HR).

- *Pragmatic data integration improvements*

Future work could prioritize use-case driven integration pilots that combine one internal and one external data source (e.g., SAP ERP with a regulatory dataset), measuring the effect on data quality, completeness and timeliness in real-time analytics.

Overall, these refinements situate future research in a pragmatic trajectory: testing and operationalizing targeted components of the Decision Intelligence framework rather than attempting wholesale transformation. By advancing explainability benchmarks, metadata-driven governance pilots, orchestration-based audit templates, semantic consistency guidelines and use-case–driven integration strategies, this research can contribute to the progressive realization of resilient, transparent and business-aligned SAP data ecosystems. The research highlights current best practices but also addresses critical gaps in active governance and real-time fraud detection, contributing to the realization of resilient, transparent and business-aligned SAP data ecosystems.

IX. Conclusion

This research reviewed the main challenges and recent developments in data modeling and governance within SAP environments. The study outlined that persistent data complexity, integration of legacy systems and inconsistent metadata structures remain key challenges to ensuring reliable and accurate analytics. The analysis also revealed that the retirement of SAP Data Intelligence has forced many organizations to rethink their data orchestration, governance and machine learning workflows. These developments confirm that sustainable decision-making requires not only technical upgrades but also well-structured governance and strong organizational coordination.

Modern SAP platforms such as SAP Datasphere, SAP S/4HANA and SAP BTP, offer important advances for managing these challenges. They support semantic modeling, federated integration and real-time analytics, helping organizations connect data from multiple systems into one governed view. When combined with AI and ML, these technologies enable the creation of Decision Intelligence frameworks that link data quality, governance and analytics into a single, explainable process for decision support.

At the same time, the study emphasized that technology alone is not enough. Effective data management also depends on people, skills and collaboration between technical teams, business users and governance experts. Building cross-functional competencies and promoting shared responsibility for data quality are key to making these frameworks work in practice.

The paper concludes that addressing data complexity in SAP environments is both a technical and strategic task. Modern solutions must combine automated quality checks, semantic consistency and explainable AI models within clear governance structures. By adopting these principles, enterprises can transform SAP from a traditional transactional system into a decision-driven platform that supports transparency, compliance, and continuous improvement. Future work should focus on testing these frameworks in real enterprise settings, refining explainability methods and expanding proactive governance to strengthen trust and agility in data-driven decision-making.

Acknowledgment

The research work presented in the paper is funded by research grant 252ПД0007-09 financed by the Research and Development Sector of TU–Sofia.

References

[1] S. Mohammed, L. Ehrlinger, H. Harmouch, F. Naumann and D. Srivastava, “The Five Facets of Data Quality Assessment,” arXiv preprint arXiv:2403.00526, 2024.

[2] A. Mohammed et al., “Multi-faceted Data Quality Assessment in Hybrid SAP Landscapes,” *J. Data Governance*, vol. 12, no. 1, pp. 34–50, 2023.

[3] M.A. Rahman, J. Bhowmik, M.S. Ahamed and R. Rahman, “Opportunities and Challenges in Data Analysis Using SAP: A Review of ERP Software Performance,” *Int. J. Manag. Inf. Syst. Data Sci.*, vol. 1, no. 4, pp. 50–67, 2024.

[4] V. Penumarthi, “Mastering SAP’s Data Models: Navigating Complex Architectures for Effective Business Intelligence,” *Int. J. Commun. Netw. Inf. Secur.*, vol. 13, no. 3, pp. 554–562, 2021.

[5] Kyndryl, “Navigating the Business, Human and Technical Challenges of SAP Transformation,” Whitepaper, Feb. 2025.

[6] SAP SE, “SAP Data & Analytics Advisory Methodology,” SAP Official Documentation, Dec. 2024.

[7] SAP SE, “What’s New in SAP Data Intelligence Cloud – End of Maintenance,” SAP Help Portal, Apr. 2024. [Online] Available: https://help.sap.com/docs/data-intelligence-cloud/sap-data-intelligence-whats-new/what-s-new-in-sap-data-intelligence

[8] SAP SE, “SAP Datasphere: The Foundation for a Business Data Fabric,” SAP Whitepaper, 2023. [Online]. Available: https://www.sap.com/documents/2023/03/sap-datasphere-business-data-fabric.html

[9] SAP SE, “SAP Datasphere: Deep Dive – A Deeper Look Into How SAP Datasphere Enables a Business Data Fabric,” SAP News Center, Mar. 8, 2023. [Online] Available: https://news.sap.com/2023/03/sap-datasphere-business-data-fabric/

[10] SAP SE, “SAP Datasphere Features,” SAP, 2024. [Online] Available: https://www.sap.com/products/data-cloud/datasphere/features.html

[11] P. Banothu and M. De Castro Fialho, “Modernizing Data Integration: SAP Datasphere’s Integration with Data Visualization Tools Versus Traditional Data Warehouse Architectures,” *Int. J. Creative Res. Thoughts*, vol. 12, no. 11, 2024.

[12] BARC, “SAP Datasphere Review: Strengths, Limitations and Market Positioning,” BARC Report, Q1 2025.

[13] S. R. Athota, “Navigating the SAP Data Analysis Landscape: Opportunities and Challenges,” *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 11, no. 1, pp. 2738–2743, 2025.*

*[14] SAP SE, “Le*arning Journey: Modeling in SAP Datasphere and SAP BW Bridge,” SAP Learning Hub, 2024.

[15] D. Smith and Y. Zhao, “Strategic Frameworks for SAP Data Governance: Accelerating Digital Transformation,” *Enterprise IT Review*, vol. 15, no. 4, pp. 210–225, 2023.

[16] SAP SE, “Machine Learning Capabilities in SAP HANA: PAL and APL Overview,” SAP Developer Center, 2024.

[17] SAP SE, "SAP Data Intelligence Cloud – ML Orchestration and BYOM Integration," SAP Documentation Portal, 2024.
[18] SAP SE, "SAP AI Core and AI Launchpad: Manage the Full Lifecycle of AI Models," SAP BTP Documentation, Apr. 2025.
[19] R. Duhova, A. Duhov, M. Lazarova and P. Georgieva, "Continuous Auditing with Machine Learning: Early Flagging of Risky Billing Documents from SAP", in *2025 13th Int. Scientific Conf. Computer Science*, 13–15 Sept. 2025, Sozopol, Bulgaria.