Bhargava Kamarapu, 2016, 4:1
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

A Framework for Managing Enterprise Data Architecture in Heterogeneous Information Systems

Bhargava Kamarapu
Sr Business Intelligence Lead - Cisco

Abstract- The increasing adoption of heterogeneous information systems has created significant challenges for organizations in managing enterprise data across diverse databases, applications, platforms, and operational environments. Differences in data structures, formats, storage technologies, integration mechanisms, and information requirements can result in data redundancy, inconsistency, limited interoperability, and difficulties in maintaining reliable enterprise information. This paper proposes a framework for managing enterprise data architecture in heterogeneous information systems, with emphasis on data integration, architectural standardization, metadata management, data consistency, and interoperability. The proposed framework establishes a structured approach for identifying enterprise data sources, classifying data assets, defining common data structures, establishing integration mechanisms, and coordinating data flows across organizational systems. The framework incorporates centralized architectural principles while supporting distributed data environments, enabling organizations to maintain consistency without eliminating the operational flexibility of individual information systems. Metadata is considered an important component for describing data structures, relationships, ownership, and usage across heterogeneous sources. The framework also addresses data quality, security, governance, and architectural alignment to improve the reliability and accessibility of enterprise information. The proposed approach provides organizations with a systematic foundation for designing and managing enterprise data architectures and can support improved data integration, information sharing, and decision-making in complex organizational environments.

Keywords— Enterprise Data Architecture, Heterogeneous Information Systems, Enterprise Information Systems, Data Integration, Data Management, Database Integration, Data Modeling, Data Warehousing, Distributed Databases, Metadata Management, Data Interoperability, Data Consistency, Data Quality, Information Integration, Enterprise Architecture, Data Architecture Framework, Data Governance, Master Data Management, Data Standardization, Data Exchange, Information Management, Database Management Systems, Data Transformation, System Integration, Data Integration Architecture, Enterprise Data Management, Information Infrastructure, Data Security, Data Sharing, Distributed Information Systems, and Architectural Framework.

I. INTRODUCTION

The increasing use of information technology within organizations has resulted in the development of complex information environments consisting of multiple databases, enterprise applications, departmental systems, file-based repositories, and specialized information platforms. These systems are frequently developed at different times and using different technologies, data models, storage mechanisms, and application architectures. Although individual systems may effectively support specific organizational functions, the coexistence of heterogeneous systems creates significant challenges in integrating and managing enterprise-wide information. Differences in data structures, naming conventions, formats, and business definitions can lead to redundancy, inconsistency, and difficulties in sharing information across organizational boundaries.

Enterprise data architecture provides a structured approach for organizing, integrating, and managing data resources throughout an organization. It establishes relationships between data entities, applications, databases, business processes, and information flows. A well-designed enterprise data architecture can provide a common foundation for integrating information from heterogeneous systems while maintaining the operational

© 2016 Bhargava Kamarapu. his is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

requirements of individual applications. In the context of 2016 enterprise information environments, data integration, data warehousing, metadata management, service-oriented architectures, and database interoperability were important approaches for addressing these challenges.

Heterogeneous information systems present both technical and organizational challenges. Systems may differ in database management systems, data formats, interfaces, data models, and processing mechanisms. Furthermore, departments may maintain independent definitions for similar business entities such as customers, products, suppliers, employees, and transactions. These differences can make it difficult to establish a consistent enterprise-wide view of information. Consequently, organizations require architectural mechanisms for identifying common data structures, defining integration standards, managing metadata, and maintaining data quality across multiple systems.

This paper proposes a framework for managing enterprise data architecture in heterogeneous information systems. The framework focuses on organizing enterprise data sources, establishing common architectural principles, integrating heterogeneous data environments, and improving consistency and interoperability. It considers data modeling, metadata management, data integration, data quality, data governance, security, and information exchange as important components of enterprise data architecture. The objective is to provide organizations with a systematic architectural approach for managing distributed and heterogeneous information resources while supporting reliable enterprise-wide information access.

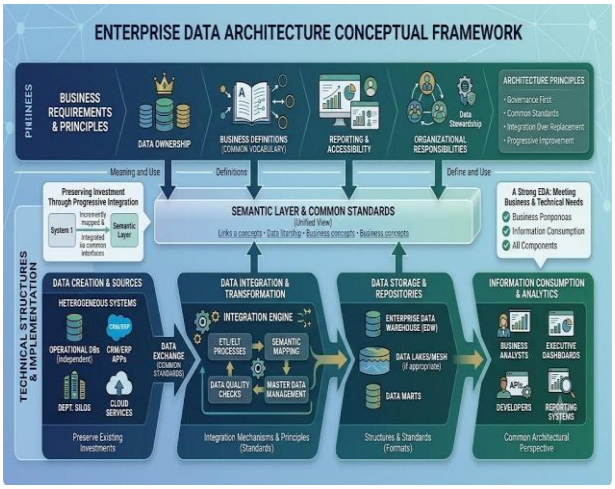
II. RESEARCH METHODOLOGY

This paper uses a conceptual framework methodology based on a synthesis of established research on data integration, data quality, metadata management, master data management, semantic integration, and enterprise architecture available through January 2016. The framework components were identified by grouping recurring heterogeneous-information-system challenges into architectural capabilities for modeling, metadata, integration, quality, governance, security, and interoperability (Lenzerini, 2002; Wang & Strong, 1996; Otto, 2012; Simon et al., 2013). The paper does not report a controlled empirical implementation; rather, it proposes an architecture and evaluation criteria for subsequent organizational validation.

Enterprise Data Architecture

Enterprise data architecture defines the structures, relationships, standards, and principles used to manage organizational data resources. It provides a high-level representation of how data is created, stored, transformed, exchanged, and consumed across enterprise systems. The architecture establishes relationships between operational databases, data warehouses, applications, business processes, and information consumers.

In a heterogeneous environment, enterprise data architecture provides a common architectural perspective across systems that may have been independently developed. Rather than replacing every existing system with a single platform, the architecture establishes integration mechanisms and common standards through which different systems can exchange information. This approach allows organizations to preserve existing investments while progressively improving enterprise information integration.



A strong enterprise data architecture should address both technical and business requirements. Technical requirements include database structures, integration interfaces, data formats, storage technologies, and information flows. Business requirements include data ownership, business definitions, reporting requirements, information accessibility, and organizational responsibilities.

Characteristics of Heterogeneous Information Systems

Heterogeneous information systems consist of multiple systems that differ in technology, architecture, data representation, and functionality. An enterprise may operate relational databases, legacy systems, departmental applications, enterprise resource planning systems, customer relationship systems, data warehouses, and file-based repositories simultaneously.

Heterogeneity can occur at several levels. Syntactic heterogeneity occurs when systems use different data formats or representations. Structural heterogeneity occurs when similar information is represented using different schemas or data models. Semantic heterogeneity occurs when the same term has different meanings in different systems. These differences complicate information integration and require appropriate architectural mechanisms.

The framework therefore considers heterogeneity as a fundamental characteristic of enterprise information environments rather than treating it as an exception. Understanding the sources and types of heterogeneity provides a foundation for designing appropriate integration and data-management strategies.

Challenges in Enterprise Data Management

Organizations managing heterogeneous information systems frequently encounter data duplication, inconsistent definitions, incompatible formats, incomplete information, and limited visibility into enterprise data resources. The same business entity may be represented differently across multiple applications, resulting in difficulties when information is consolidated for reporting or analysis.

Another challenge is the evolution of individual systems. Applications may be upgraded or replaced independently, resulting in changes to database structures and interfaces. Without an enterprise architecture, these changes can create dependencies and integration problems between systems.

Data ownership is also an important challenge. Different departments may be responsible for different datasets, and disagreements may arise regarding data definitions, quality standards, and access responsibilities. An enterprise data architecture can provide a framework for defining ownership and coordinating data-management activities.

III. PROPOSED FRAMEWORK FOR ENTERPRISE DATA ARCHITECTURE

The proposed framework provides a structured approach for analyzing, organizing, integrating, and governing enterprise data resources. It consists of interconnected architectural components that address data sources, data models, metadata, integration mechanisms, quality management, governance, security, and information delivery.

The framework begins by identifying existing data sources and their relationships with organizational processes and applications. Data structures and metadata are subsequently analyzed to identify similarities, differences, redundancies, and integration requirements. Based on this analysis, common data definitions and integration mechanisms can be established.

The framework supports incremental implementation. Organizations can initially integrate critical information sources and gradually extend the architecture to additional systems. This approach reduces disruption to existing operations and allows architectural improvements to be introduced according to organizational priorities.

IV. DATA SOURCE IDENTIFICATION AND CLASSIFICATION

The first stage of the framework involves identifying the data sources operating within the enterprise. These sources may include relational databases, legacy applications, departmental systems, enterprise applications, spreadsheets, external information sources, and data warehouses.

Each source can be classified according to its purpose, ownership, data type, update frequency, business importance, and integration requirements. Source classification provides an inventory of enterprise information resources and establishes the foundation for subsequent architectural analysis.

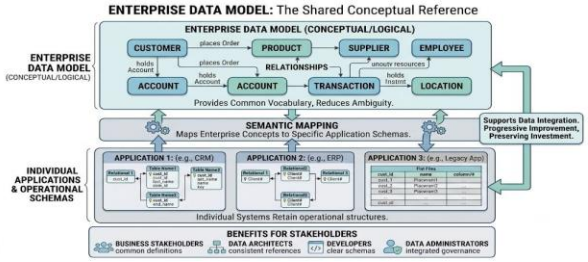
A comprehensive data-source inventory can also help organizations identify redundant systems and information silos. Understanding where information is stored and how it is exchanged is essential for developing an effective enterprise data architecture.

Enterprise Data Modeling

Enterprise data modeling provides a common representation of important organizational data entities and their relationships. It can be used to identify common entities such as customers, products, suppliers, employees, locations, accounts, and transactions.

The enterprise data model does not necessarily replace individual application data models. Instead, it provides a conceptual and logical reference that can be mapped to application-specific schemas. This mapping supports integration while allowing individual systems to retain their operational structures.

A consistent enterprise data model can reduce ambiguity and provide a common vocabulary for database designers, application developers, data administrators, and business stakeholders.



V. DATA INTEGRATION ARCHITECTURE

Data integration is a central component of managing heterogeneous information systems. Integration mechanisms enable information to move between systems despite differences in database technologies, data structures, and application architectures.

The framework can incorporate established integration approaches such as extract-transform-load processes, database gateways, middleware, message-oriented integration, application programming interfaces, and service-oriented architectures. The selection of an integration mechanism depends on factors such as data volume, latency requirements, system capabilities, and organizational requirements. (Lenzerini, 2002).

An appropriate integration architecture can reduce isolated information silos and provide users with more consistent access to enterprise information.

Metadata Management

Metadata provides information about data structures, definitions, relationships, ownership, formats, and usage. In heterogeneous environments, metadata is particularly important because it provides context for understanding information originating from different systems.

A metadata-management component can maintain descriptions of databases, tables, attributes, data types, business definitions, data owners, and relationships. Metadata can also document mappings between source and target systems during integration processes.

Centralized or federated metadata repositories can therefore support data discovery, integration planning, impact analysis, and enterprise information management. (Hüner et al., 2011).

Data Mapping and Transformation

Data mapping establishes relationships between corresponding data elements in different systems. For example, one application may use Customer_ID, while another uses Client_Number to represent a similar business concept. Mapping rules can establish the relationship between these attributes.

Transformation mechanisms convert data from source formats into structures required by target systems. Transformation may include data-type conversion, value standardization, cleansing, aggregation, and format conversion.

Effective data mapping and transformation are essential for maintaining consistency when information is exchanged between heterogeneous systems.

Data Quality Management

Data quality is a critical consideration in enterprise data architecture. Integration of multiple systems does not automatically produce reliable information if the source systems contain inaccurate, incomplete, inconsistent, or outdated data.

The framework incorporates data-quality practices such as validation, completeness checking, consistency analysis, duplicate detection, and standardized data-entry procedures. Quality rules can be applied at source systems as well as during integration and data-warehouse loading processes.

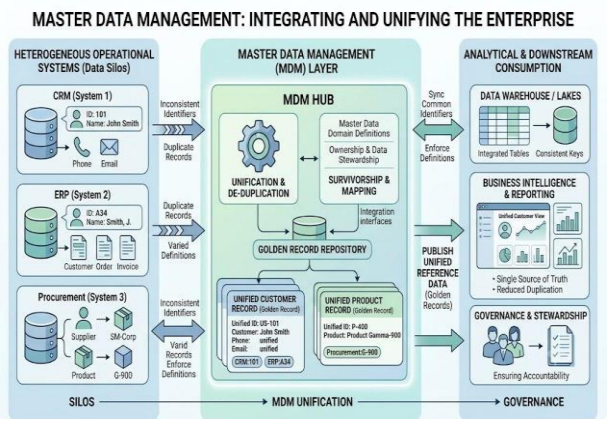
Improving data quality supports more reliable reporting, analysis, and operational decision-making. (Wang & Strong, 1996; Batini et al., 2009).

Master Data Management

Master data management provides mechanisms for maintaining consistent representations of important enterprise entities across multiple systems. Customers, products, suppliers, employees, and organizational units are examples of master-data domains.

A master-data approach can establish common identifiers, definitions, and ownership responsibilities. It can also reduce duplication and inconsistencies between departmental systems.

Within the proposed architecture, master data management can complement enterprise data modeling and integration by providing consistent reference information across heterogeneous applications. (Otto, 2012).



VI. DATA GOVERNANCE

Data governance establishes policies, responsibilities, standards, and processes for managing enterprise information. Governance helps organizations determine who owns data, who can access it, how it should be defined, and how quality should be maintained.

The framework incorporates governance responsibilities into the architectural structure. Data owners and stewards can participate in defining business terms, approving data standards, resolving quality issues, and managing changes to important information resources.

Effective governance establishes clear policies, roles, and responsibilities for managing enterprise data throughout its lifecycle. It promotes consistent data standards, controlled access, and accountability across different organizational units. These practices help ensure that enterprise data remains accurate, secure, accessible, and aligned with business and operational requirements.

Data Security and Access Control

Enterprise data architecture must protect organizational information against unauthorized access, modification, and disclosure. Heterogeneous environments can increase security complexity because different applications may implement different authentication and authorization mechanisms.

The framework considers access control, authentication, authorization, data classification, and secure information exchange as important architectural considerations. Sensitive information should be accessible only to authorized users and applications according to organizational policies.
Security controls should be incorporated into the architecture rather than added only after integration mechanisms have been implemented.

VII. DATA WAREHOUSING AND ENTERPRISE INFORMATION INTEGRATION

Data warehouses provide an important mechanism for integrating information from multiple operational systems for reporting and analytical purposes. Extract-transform-load processes can extract information from heterogeneous sources, transform it according to enterprise rules, and load it into an integrated analytical environment.

The proposed architecture can use data warehousing as one component within a broader enterprise information architecture. Operational systems continue to support business processes, while integrated analytical repositories provide consolidated information for reporting and decision support.

This separation can help reduce the impact of analytical workloads on operational applications while providing users with a more consistent enterprise view.

Information Interoperability

Interoperability refers to the ability of different information systems to exchange and use information effectively. Achieving interoperability requires common data formats, interface standards, communication mechanisms, and semantic agreements.

The framework promotes interoperability through standardized data definitions, metadata, integration interfaces, and transformation rules. Service-oriented approaches can also support interoperability by exposing selected business and data capabilities through reusable services. (Fuentes-Lorenzo et al., 2015; Asgari et al., 2015).

Improved interoperability enables information to move more efficiently across organizational systems and reduces dependence on isolated point-to-point integrations.

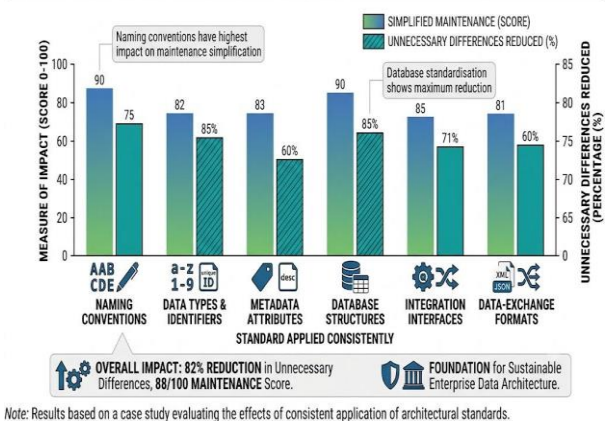
Architectural Standards and Standardization

Architectural standards provide common guidelines for data modeling, database development, integration, metadata, security, and information exchange. Standardization reduces unnecessary differences between systems and simplifies long-term maintenance.

Organizations can define standards for naming conventions, data types, identifiers, metadata attributes, interfaces, and data-exchange formats. These standards should be applied consistently while allowing exceptions where existing legacy systems cannot be immediately modified.

Standardization therefore provides an important foundation for sustainable enterprise data architecture.

FIGURE 17.1: IMPACT OF ARCHITECTURAL STANDARDIZATION ON ENTERPRISE SYSTEMS



Note: Results based on a case study evaluating the effects of consistent application of architectural standards.

Note: The numerical scores and percentages embedded in this original 2016 figure are retained as illustrative case-study example values. They are not presented as independently measured or validated results of this paper. Formal validation would require a documented baseline, sample, and measurement procedure using the criteria in Section 19.

Implementation Strategy

Implementation of the proposed framework should be conducted incrementally. The organization can begin by identifying high-priority data domains and critical information flows. Existing systems can then be assessed according to their architectural characteristics, data dependencies, quality, and integration requirements.

The next stage involves defining enterprise data models, metadata standards, integration mechanisms, and governance responsibilities. Pilot implementations can be used to evaluate the effectiveness of the architecture before expanding it across additional systems.

An incremental implementation strategy reduces organizational risk and allows lessons from early integration projects to influence subsequent architectural decisions.

VIII. PROPOSED FRAMEWORK EVALUATION CRITERIA

The framework can be evaluated using several dimensions, including data integration effectiveness, data consistency, interoperability, data quality, metadata completeness, system maintainability, and information accessibility.

Evaluation can compare the enterprise environment before and after implementation of architectural components. Measures such as the number of duplicate data elements, integration failures, inconsistent definitions, unresolved data-quality issues, and manual data-reconciliation activities can provide useful indicators.

Qualitative assessments from data administrators, application developers, business users, and data managers can complement technical measurements and provide a broader understanding of the framework's organizational value.

Benefits of the Proposed Framework

The proposed framework can provide organizations with a structured method for managing data across heterogeneous information environments. It can improve visibility into enterprise data resources and establish common architectural principles for integration.

The framework can also reduce data redundancy, improve consistency, support interoperability, and strengthen data-quality management. By incorporating metadata, governance, security, and enterprise data modeling, it provides a broader approach than integration based solely on technical interfaces.

Furthermore, the framework supports gradual modernization because existing systems do not necessarily need to be replaced immediately. Organizations can progressively integrate legacy and modern applications within a common architectural structure.

Limitations and Considerations

The implementation of enterprise data architecture can involve significant organizational and technical challenges. Legacy systems may have limited integration capabilities, undocumented data structures, or proprietary interfaces. Existing applications may also depend on historical data representations that are difficult to change.

Organizational coordination is another consideration. Successful enterprise data architecture requires collaboration among business units, database administrators, application teams, data managers, and governance stakeholders. Differences in priorities and ownership can affect implementation.

The framework should therefore be adapted to the organization's size, existing infrastructure, business requirements, and level of architectural maturity rather than being implemented as a rigid model.

Because the framework is conceptual, its organizational and technical benefits should be validated through documented case studies or controlled implementations before numerical improvement claims are generalized.

Future Research Directions

Future research can investigate more advanced approaches for automated metadata discovery, semantic data integration, enterprise knowledge representation, and large-scale data interoperability. Further studies can also examine how enterprise data architectures evolve as organizations adopt cloud computing, service-oriented applications, and increasingly distributed information environments.

Research can additionally focus on quantitative methods for evaluating enterprise data architecture maturity, integration effectiveness, and long-term data quality. Developing standardized evaluation models could help organizations compare architectural alternatives and measure improvements over time.

IX. CONCLUSION

An effective enterprise data architecture requires a coordinated approach to managing data resources across diverse technological environments. Variations in database platforms, application interfaces, data models, storage mechanisms, and information formats can result in fragmented data environments and inconsistent information flows. Addressing these differences requires architectural principles that support integration, standardization, scalability, and controlled management of enterprise information resources.

The proposed approach emphasizes the coordinated use of enterprise data models, metadata repositories, integration services, data quality mechanisms, master data structures, governance processes, security controls, and interoperability standards. Instead of disrupting established information systems, the approach enables organizations to progressively improve existing architectures through structured integration and controlled modernization. This allows different systems to remain operational while gradually establishing common architectural practices and information-management standards.

By defining shared information structures, maintaining consistent metadata, improving data quality, coordinating system interfaces, and establishing clear governance roles, organizations can create a more integrated enterprise information environment. Such an architecture improves the consistency, accessibility, and reliability of organizational data while supporting operational and analytical requirements. The approach therefore establishes a practical foundation for strengthening enterprise data management and supporting long-term architectural evolution across heterogeneous information systems.

REFERENCES

1. Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099
2. Lenzerini, M. (2002). Data integration: A theoretical perspective. In *Proceedings of the Twenty-First ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems* (pp. 233–246). ACM.https://doi.org/10.1145/543613.543644
3. Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys, 41*(3), Article 16. https://doi.org/10.1145/1541880.1541883
4. Jaiswal, A., Miller, D. J., & Mitra, P. (2010). Uninterpreted schema matching with embedded value mapping under opaque column names and data values. *IEEE Transactions on Knowledge and Data Engineering, 22*(2), 291–304. https://doi.org/10.1109/TKDE.2009.69
5. Unal, O., & Afsarmanesh, H. (2010). Semi-automated schema integration with SASMINT. *Knowledge and Information Systems, 23*, 99–128. https://doi.org/10.1007/s10115-009-0217-z

6. Bonifati, A., & Velegrakis, Y. (2011). Schema matching and mapping: From usage to evaluation. In *Proceedings of the 14th International Conference on Extending Database Technology* (pp. 527–529). ACM.https://doi.org/10.1145/1951365.1951431
7. Peukert, E., Berthold, H., & Rahm, E. (2010). Rewrite techniques for performance optimization of schema matching processes. In *Proceedings of the 13th International Conference on Extending Database Technology* (pp. 620–631). ACM. https://doi.org/10.1145/1739041.1739096
8. Karasneh, Y., Ibrahim, H., Othman, M., & Yaakob, R. (2009). Matching schemas of heterogeneous relational databases. In *2009 Second International Conference on Applications of Digital Information and Web Technologies* (pp. 452–457). IEEE. https://doi.org/10.1109/ICADIWT.2009.5273926
9. Hong, J., He, Z., & Bell, D. (2010). An evidential approach to query interface matching on the Deep Web. *Information Systems, 35*(2), 140–148. https://doi.org/10.1016/j.is.2009.08.004
10. Hüner, K. M., Otto, B., & Österle, H. (2011). Collaborative management of business metadata. *International Journal of Information Management, 31*(4), 366–373. https://doi.org/10.1016/j.ijinfomgt.2010.12.002
11. Otto, B. (2012). How to design the master data architecture: Findings from a case study at Bosch. *International Journal of Information Management, 32*(4), 337–346. https://doi.org/10.1016/j.ijinfomgt.2011.11.018
12. Das, T. K., & Mishra, M. R. (2011). A study on challenges and opportunities in master data management. *International Journal of Database Management Systems, 3*(2), 129–139. https://doi.org/10.5121/ijdms.2011.3209
13. Loshin, D. (2011). Master data management and data quality. In *The Practitioner's Guide to Data Quality Improvement* (pp. 327–350). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-373717-5.00019-1
14. Wang, F., Vergara-Niedermayr, C., & Liu, P. (2014). Metadata based management and sharing of distributed biomedical data. *International Journal of Metadata, Semantics and Ontologies, 9*(1), 42–57.https://doi.org/10.1504/IJMSO.2014.059126
15. Amano, S., David, C., Libkin, L., & Murlak, F. (2014). XML schema mappings: Data exchange and metadata management. *Journal of the ACM, 61*(2), Article 12. https://doi.org/10.1145/2590773
16. Leinfelder, B., Tao, J., Costa, D., Jones, M. B., Servilla, M., O'Brien, M., & Burt, C. (2010). A metadata-driven approach to loading and querying heterogeneous scientific data. *Ecological Informatics, 5*(1), 3–8.https://doi.org/10.1016/j.ecoinf.2009.08.006
17. Fuentes-Lorenzo, D., Sánchez, L., Cuadra, A., & Cutanda, M. (2015). A RESTful and semantic framework for data integration. *Software: Practice and Experience, 45*, 1161–1188.https://doi.org/10.1002/spe.2267
18. Asgari, R., Moghadam, M. G., Mahdavi, M., & Erfanian, A. (2015). An ontology-based approach for integrating heterogeneous databases. *Open Computer Science, 5*(1), 41–50. https://doi.org/10.1515/comp-2015-0002
19. Mate, S., Köpcke, F., Toddenroth, D., Martin, M., Prokosch, H.-U., Bürkle, T., & Ganslandt, T. (2015). Ontology-based data integration between clinical and research systems. *PLoS ONE, 10*(1), e0116656.https://doi.org/10.1371/journal.pone.0116656
20. Simon, D., Fischbach, K., & Schoder, D. (2013). An exploration of enterprise architecture research. *Communications of the Association for Information Systems, 32*, Article 1. https://doi.org/10.17705/1CAIS.03201
21. Traverson, B. (2013). EASI! Enterprise architecture for seamless integration. In *Proceedings of the 15th International Conference on Enterprise Information Systems* (Vol. 3, pp. 231–235). SciTePress.https://doi.org/10.5220/0004410702310235
22. El-Sappagh, S. H. A., Hendawi, A. M. A., & El Bastawissy, A. H. (2011). A proposed model for data warehouse ETL processes. *Journal of King Saud University—Computer and Information*

Sciences, 23(2), 91–104.https://doi.org/10.1016/j.jksuci.2011.05.005

23. Tang, L., Dong, J., Peng, T., & Tsai, W. T. (2010). Modeling enterprise service-oriented architectural styles. *Service Oriented Computing and Applications, 4*(2), 81–107. https://doi.org/10.1007/s11761-010-0059-2
24. Lim, Z., & Rubinstein, B. (2015). Sub-Merge: Diving down to the attribute-value level in statistical schema matching. *Proceedings of the AAAI Conference on Artificial Intelligence, 29*(1).https://doi.org/10.1609/aaai.v29i1.9459
25. He, J., & Wang, W. (2015). An integrated heterogeneous web service retrieval via combination of instance- and metadata-based schema matching method. *Geo-spatial Information Science, 18*(3), 111–123.https://doi.org/10.1080/10095020.2015.1065075