Computer Science Journal of Moldova, vol.**33**, no.1(97), 2025

Linked Data Lineage as a foundation of continuous data integration and modern data governance

Georgiana Calancea, Lenuta Alboaie, Inga Titchiev

Abstract

Our paper focuses on Linked Data Lineage as the collection of modeled organizational data, enabling continuous integration and modern governance in Business-to-Business (B2B) sharing. We first explore approaches organizations adopt to enhance B2B data integration. Next, we highlight key B2B and Business-to-Government (B2G) linked data use cases and strategies for governance frameworks. We review software tools supporting data governance and propose a systemic method for defining organizational integration flows. This is demonstrated through a system enabling businesses to design linked data pipelines, leveraging Semantic Web technologies for continuous integration and governance. Insights and recommendations for improving data sharing processes are provided.

Keywords: Semantic Web technologies, B2B data sharing, enterprise data governance, continuous data integration.

MSC 2020: 68T30, 68P20, 68U35, 91B64, 68Q85.

1 Importance of data governance frameworks in the data sharing processes

Data sharing plays an important role in forming a company's ability to gain customers and satisfy the existing client base. However, as shown in [1], none of the current representative data sharing solutions offer a complete solution in terms of data governance [2], confidentiality, and

doi:10.56415/csjm.v33.05

real-time modeling of the data to be reusable, interoperable, accessible, and easily findable. Since the efficiency of collecting and processing data can impact business decisions, using traditional data integration techniques is not enough. “Modern data integration strategies” [3] are necessary to build a complete image of the customer’s behavior and to connect to more data sources as fast as they are made available.

According to two data strategy overview reports [4], [5] the future directions target a fully distributed data ecosystem, made available through services, meshes, query accelerators and knowledge graphs. Data as a service (DaaS) introduces the idea that integration should be done through an access layer composed of APIs and database/data warehouse communication protocols in order to take full advantage of iPaaS platforms [6], data virtualization tools [7], and data meshes [8]. On top of data meshes and Semantic Web technologies [9] and concepts, knowledge graphs [10] can be built to infer connections between data, both inter-domain and cross-domains. Their applicability in data analytics targets the design of software that can predict user behavior, automatically curate and integrate data in the knowledge model, and facilitate knowledge-based inferences and decision-making.

An organization’s data sources are no longer restricted to its own data harvesting processes. The sources are supplemented through data sharing pipelines with other businesses, which leads to an ever-growing amount of data, hard to organize and manage over its whole life cycle [11].

In this context, each organization adopts a data strategy in order to better take advantage of the available data assets to generate business value. There are two types of data strategies, namely defensive and offensive. The defensive data strategy focuses on minimizing risks in regard to fraud or theft, compliance with data privacy and financial reporting, and identifying authoritative data sources. The offensive data strategy is rather focused on gaining insights on what the customer needs and establishing future business goals, while also improving the efficiency of current organization processes. In order to continuously evolve, enterprises need to find a middle ground for the two strategies [11].

For this initiative to be successful, having a robust data governance framework should be part of the data strategy. Data governance is focused on data quality and accuracy metrics and allows enterprises to track the source of the data and its usage, while decreasing the risk of being non-compliant with data regulation policies, such as General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA). As such, data governance refers to the steps an organization must take to ensure its data has the key qualities described in the Findable, Accessible, Interoperable, Reusable (FAIR) principles. Data stewardship is a common use-case of applicability for data governance, considering that the entities using the data must be held accountable and responsible for the ways they use the data [12].

In the enterprise environment, two approaches to building a data governance framework have been standardized over time.

The top-down approach [13] is a centralized strategy, which considers data control the key to optimizing data quality in its core principles. Data modeling is the priority of this approach – only a small circle of people is entrusted with this process. Once they complete this assignment, data can be made available for the rest of the organization for analytics. The main issue with the top-down approach is scalability, since a single department handling data modeling is a bottleneck for the other business units, blocking the product roadmap.

The bottom-up approach [14] is a decentralized data governance strategy, which starts from the raw big data, on top of which it builds the metadata extraction processes, quality controls, and security procedures. With this approach, businesses have no control over the authorship and trustworthiness of data from the beginning, since governance is implemented after each business department models the data on its own terms.

Considering these aspects, organizations need to adopt a balanced data governance approach, which harmonizes the need for access with the need for control. Any part of an organization should be able to produce, own, and curate data, while also ensuring that rigorous control mechanisms are imposed early on to protect data quality. In this strategy, validating the business data is trustworthy is a common goal for all actors of an organization, through a larger set of appointed data stewards in each department. They ensure the framework rules are followed and the resulting data is ready to be shared without having any risks not mitigated beforehand [11]. Adopting the latter data governance framework allows an organization to focus on growing its data heritage by modeling and sharing the data it generates.

2 Linked data sharing – motivation and use-cases

Considering the opportunities that linked permissioned data brings in the data sharing processes, government and enterprise customers could benefit in terms of data sharing, integration, and governance policies definition if linked permissioned data were to become commonly used. There are multiple use-cases where linked permissioned data can be an important part of the data integration, as well as the data sharing process:

- *Inferencing for Money Laundering Detection (B2G)* – This use case targets the easier identification of fraudulent transactions through data integration pipelines, for example, in the case of Romania – between the National Agency for Fiscal Administration (ANAF) and banks;

- *Integration of public & private healthcare systems with the National Health Insurance House/Casa Națională de Asigurări de Sănătate (CNAS) (B2G)* – Private healthcare providers have appeared to complement the public healthcare system. Integration pipelines should be set up to reduce the paperwork for the provided medical services;

- *Adjustment of accounting related software systems to legislative changes (B2G)* – Software accounting systems must be continuously changed to adapt to legislative changes. The solution often relies on a financial analyst to translate to the engineers the changes to perform.

- *Natural resource dependency graph between countries (B2G)* – Accentuated by the 2022 Ukraine-Russia war, most European countries were taken by surprise by their dependency to survive on Russian fossil fuels and gas. This situation could have been avoided if more accurate data on supply and demand had been exchanged between governments and businesses;

- *Adjustment of accounting related software systems to legislative changes (B2G)* – Software accounting systems must be continuously changed to adapt to legislative changes. The solution often relies on a financial analyst to translate to the engineers the changes to perform;

- *Seamless supply/demand chain data integration and quantities estimation (B2B)* – Nowadays, in the European context, most products are widely available on the shelves for consumers to buy at any time. This is possible through complex distribution chains built between businesses. The integration struggles regarding quantity estimation and production in agriculture and retail became highly visible with the COVID-19 pandemic [15];

- *Employee & Business heritage (B2B)* – Similar to the "Software Heritage" solution [16], an "Employee Heritage" and a "Business Heritage" could be built. Past behaviors – social, legal, economic – of employees and businesses throughout their collaboration are documented through linked permissioned data. Based on these heritages, both potential employees and employers could make informed decisions in regard to starting a new collaboration.

- *Build the user profile & accounting reports within a Business Intelligence Team (B2B)* – In order for a business to detect the needs and wishes of its customers, aggregating data from payments, on-premise products, marketing, and cloud services is required. That data must often be continuously integrated with data from other business partners – this helps businesses understand the importance and productivity of their partnerships.

3 Software tools for building a data governance framework

Linked data sharing, in comparison to plain data sharing, can help organizations build mechanisms to control data quality, ownership, and access in a more straightforward manner. In order to show the way linked data sharing can simplify the existing enterprise data governance approaches, we compare the professional data governance software, based on a set of relevant criteria.

3.1 Comparison Criteria

Ideally, an organization should be able to customly transform a dataset with low-code or no-code involved in order to simplify the integration process with the existing data. An integrated data catalog should be available for the organization in order to search, manage, and share the existing data. A data governance platform should also offer permission-based access and management of resources, in order to ensure that each actor of the organization actually has the least privilege needed over the data. In order to ensure data ownership rights are preserved, provenance of newly integrated datasets should not be overlooked. The data governance framework of an organization is also sustained through data curation and obfuscation mechanisms – this partly contributes to another requirement referencing the need for data confidentiality. Since each organization may use its own governance tool, structured, non-structured, and partially structured data sources should be supported in the sharing process. Continuous data integration should be sustained through Extract, Transform, Load (ETL), or similar processes, which also contribute to building a data lineage [33]. A complementary business requirement to data integration is building a business glossary, which promotes understanding of business terms by ensuring everyone has access to their clear, unambiguous definitions. Constant data quality validations must be performed in the process of integration. Security standards must be put in place to ensure data is protected from leaks, loss, and unauthorized access. High visibility over a business's assets can be achieved through the ability to build reports and metrics.

Good compatibility with third-party tools for this purpose should exist. Software usability should be taken into consideration since it can improve the productivity of the data engineers using it.

3.2 Software Tools Overview

Considering several Magic Quadrant reports issued by Gartner in 2020 – 2022 period, covering the topics of data quality [15], data integration [16], master data management [17], and integration platform as a service [18], we chose five data governance solutions we believe to be relevant in this study:

1. *Axon Data Governance* [19] – developed by Informatica, offers a platform with quick access to curated data, where compliance with the organization's data governance framework can be easily supervised; in this marketplace, teams of engineers, data stewards, data owners, and business stakeholders can find and access data assets and are also able to collaboratively document and manage the data items that require governance.

2. *IBM Infosphere Information Governance Catalog* [20] – a web-based tool that allows organizations to explore and analyze their data through a common business language, defined by the organization's stakeholders; in order to leverage existing curated data sets and their metadata, the on-premise data repository can be extended with the Cloud information, provided by the IBM Watson Knowledge Catalogue.

3. *Collibra Data Governance* [21] – part of the Collibra Data Intelligence Cloud, a platform that aims to help organizations manage their data sources and maintain data quality and security through its lifecycle.

4. *SAP Data Governance* [22] – software solution that helps organizations manage and govern their data, including both master data and transactional data; it offers resources for collaboration across the organization and data quality management, with an emphasis on assuring data security, correctness, and consistency;

5. *Erwin by Quest* [23] – offers a unified platform with features that help enterprises in multiple aspects of their management, such as building and maintaining a data governance framework, while modeling their business architecture and processes. This solution is provided as a managed service that enables users to create and deploy data sources, locate and harvest data, and link physical metadata to particular business terms and definitions.

In Table 1, we synthesize the software solutions review we performed. According to the results shown, we can conclude that all solutions are more than suitable candidates to help enterprise users build a data governance framework suitable for their business. However, all of them are proprietary solutions, and in order to be able to automate the whole process of extraction, transformation, and load of the data in a unified data catalog with corresponding semantic information, additional mechanisms must be built and configured by software engineers over the base solutions provided by each tool. Considering the trend to take advantage of processing and extracting metadata from the datasets shared across enterprises, highlighted in Section 1, we can rely on the FAIR principles and the Semantic Web technologies in order to build and manage the integration and governance processes across many shared data sources. In Section 4, we propose an architecture for such a solution.

4 Enhancing Data Governance practices, resources, and processes for B2B sharing

In order to show the ways linked data lineage could positively contribute to the decision-making process within businesses and facilitate the data sharing process, an example of integration between a Subscription Service Provider (SSP) and an IoT Hardware Manufacturer (IHM) is presented. The SSP has the role of providing subscription management services to the IoT hardware manufacturer, which sells smart home equipment and needs subscriptions for their management. An inventory of the users, devices, and subscriptions is actively maintained and updated through the cloud infrastructure.

Table 1. Software solutions review for building a data governance framework

Criteria	**Solution Providers**				
	Infor matica Axon	**IBM**	**Colli bra**	**SAP**	**Erwin**
Low-code / No-code Customization for FAIR Data Transformation	o	-	+	o	-
Collaborative Data Management	+	+	+	+	+
Offers an Integrated Data Catalog	+	+	+	+	+
Permissioned Data Sharing	+	-	o	o	-
Data Provenance Information	+	+	+	-	-
Data Curation Capabilities	+	-	+	+	+
Data Accessibility Mechanisms	+	+	+	+	+
Business Glossary Management	+	+	+	o	+
FAIR Metadata Extraction and Management	o	-	o	o	+
Continuous Data Quality Assessment	+	o	+	+	+
Data Confidentiality Mechanisms	+	o	+	+	-
Compliance with Data Security Standards	+	+	+	+	-
High Visibility of Business Assets through Data Metrics and Reports	+	+	+	+	+
Data Lineage Visualisation	+	+	+	+	+
Ability to aggregate structured and unstructured data sources	+	+	+	+	+
Compatibility with third-party software	o	o	+	-	+
Trial version	+	+	+	+	+
Community Version/ Free Tier	-	-	-	-	-

Legend:
+ – feature available in the studied software solution;
o – feature partially available or could be improved – through third-party solutions;
- – feature not available in the studied software solution.

Figure 1 describes how the data integration is performed between the two organizations. The data produced by the SSP is sent to the Business Intelligence (BI) department through a publish/subscribe system, such as RabbitMQ [24]. The BI department has its own software jobs, in charge of consuming the data, curating it, and saving it into a data warehouse. The data is aggregated regularly in custom financial report scripts, which are sent to the accounting department. The reports are also sent to the IoT Hardware Manufacturer through a webhook in order for the two businesses to synchronize in terms of costs. Another custom integration pipeline is designed directly between the data warehouse of the SSP and the data warehouse of the IHM. A major downside to this approach is that every change in the data model (adding new fields, renaming existing ones) requires further development in the data transformation and curation processes.

Considering the current approach, each integration with a business partner requires custom ETL processes to be developed and data types and models to be established in agreement with all the involved parties within the B2B process. In the long run, this proves to be highly inefficient in terms of associated time and required maintenance.

We aim to offer a better solution based on linked permissioned data and templates of ETL processes. Our solution does not require engineers to write custom code or advanced knowledge of data integration. The proposed solution is designed as outlined in Figure 2.

The solution presented in Figure 2 is based on six essential components described as follows: the **'Vocabulary Marketplace'** component helps data engineers search for existing vocabularies, suitable for describing the data in the business sharing process. It starts from a predefined set of standard well-known vocabularies, retrieved from linked data vocabulary catalogs, such as Linked Open Vocabularies (LOV). If additional vocabularies need to be defined to adjust all the concepts, the **'Vocabulary Wizard'** component allows defining concepts and properties with their value types. This wizard also comes bundled with a **'Vocabulary Suggestions Webhook'**, which retrieves possible matches for concepts and properties that the engineers try to define in the modeling tool. The suggestions are retrieved based on Semantic Web engines and the permissioned vocabulary catalogs. Afterward,

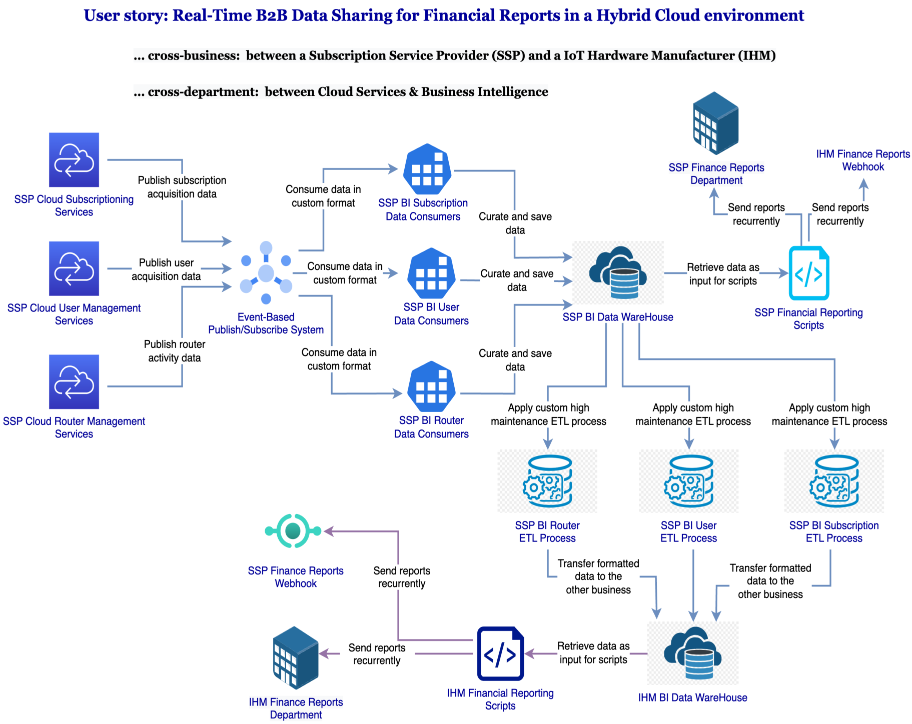

Figure 1. B2B Data Sharing Current Use-Case

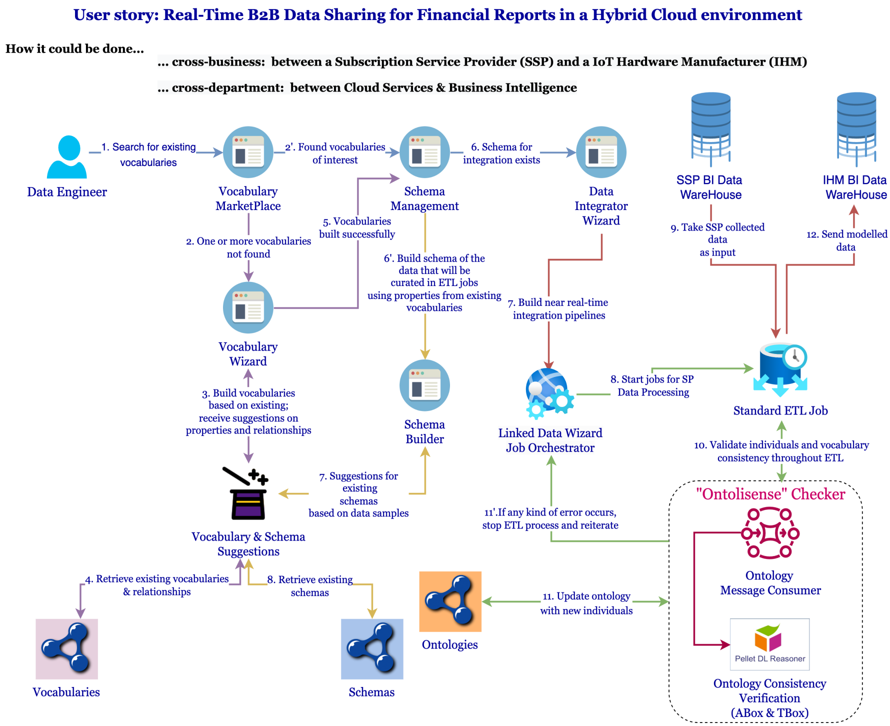

Figure 2. B2B Data Sharing using our solution

the engineers can use the **'Schema Builder'** tool, where properties from the chosen vocabularies are used to describe the final shape for the modeled data. The resulting schema should be further used as a parameter to an ETL process template to obtain custom modeling behavior within a standard ETL process. All custom ETL schemas can be visualized and managed in terms of access in order to allow open or permissioned sharing. The next step would be to access the **'Data Integration Wizard'** component, which allows both technical and non-technical personnel within an organization to configure near real-time pipelines of continuous integration. This request is handled by the **'Linked Data Wizard Job Orchestrator'**, whose purpose is to communicate with an external open-source job orchestration solution and to create from the ETL Job Template a custom processing flow, which transforms and moves the security service provider's data towards the router manufacturer's warehouse. At the same time, the vocabularies, data schema, and the individuals resulting from the dataset form an ontology. As a result, the **'Ontolisense Checker'** was created to validate the consistency of the individuals with the initial modeling, using known ontology validators. In case of issues within the modeling or individuals, it notifies engineers about the problems and offers reconciliation mechanisms. In Figure 2, we can see how all components in the proposed solution interact with each other to meet the needs of our use case.

In Figure 3, we outline the necessary infrastructure of nodes and open-source technologies to host the solution. The aim is to offer the product as an easily accessible open platform, where organizations can create and transform their own datasets. We also offer the solution as a Docker file [25] with a Linux-based OS, which can be used by organizations with their own infrastructure. In this way, the proposed solution can be used both as an industrial data platform – only trusted organizations can join it – and as a public data marketplace, in terms of B2B data sharing tools classification [26]. Some of the data and metadata stored on the nodes are synchronized using the RabbitMQ publish/subscribe technology [24]. The blockchain solution [27] is used for auditing purposes, while the smart contracts are the ones controlling the performed operations with the data. Apache Airflow [28],

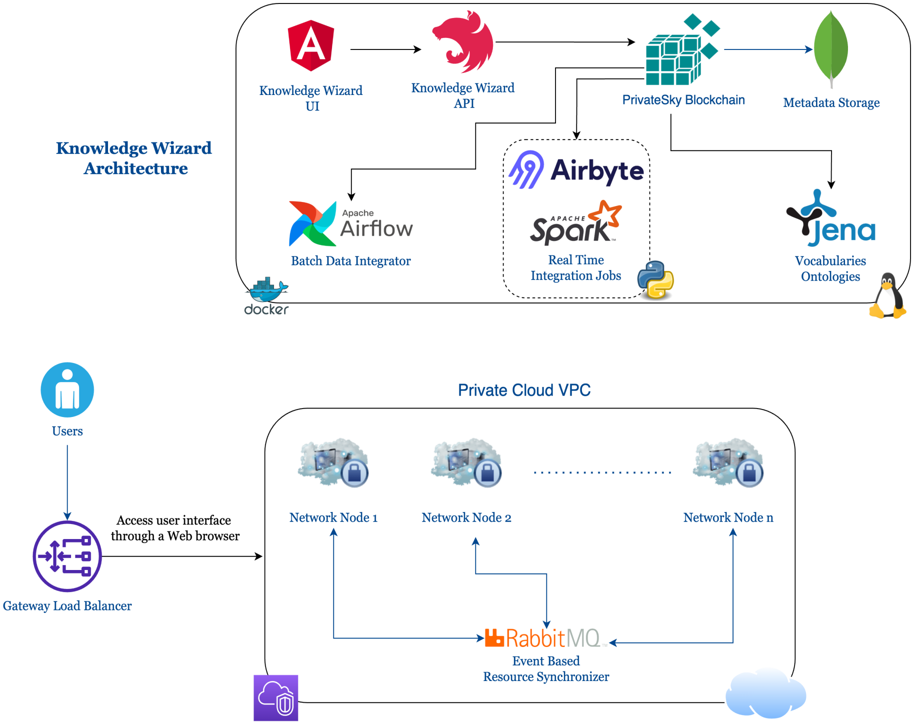

Figure 3. "Linked Data Governor" Technological Approach

Airbyte [29], and Spark [30] are used for the near real-time and batch integrations of data sharing. Apache Jena [31] has the purpose of storing the data and metadata related to custom vocabularies, schemes, and the resulting ontologies.

With the proposed architectural solution from Figure 2 and the technologies chosen for development from Figure 3, we aim to achieve every criteria outlined in Table 1, as follows:

- *"Low-code / No-code Customization for FAIR Data Transformation"* and *"FAIR Metadata Extraction and Management"* – This criterion will be achieved through the general-purpose transformation jobs, which are customized using the vocabularies and always result in customized FAIR-compliant linked datasets. The resulting datasets will be managed in a dedicated data catalog;

- *"Collaborative Data Management", "Offers an Integrated Data Catalog"*, and *"Data Accessibility Mechanisms"* – Users will be able to manage the datasets and their metadata in a data catalog section. The owner can provide role-based access to other platform users to collaborate in the data management process;

- *"Permissioned Data Sharing"* – A data sharing mechanism will be provided for the users inside the platform and also to external data destinations, where the data should be available post-transformation. Data sharing is possible according to the dataset distribution license and the permissions of the user initiating the sharing process;

- *"Data Provenance Information"* – Data provenance will be ensured by extracting provenance-specific metadata in the data transformation process. This is possible using the PROV-O (Provenance Ontology): The PROV Ontology [32] as a reference in the metadata modeling for the final linked data; this ontology provides a formal framework for representing and interchanging provenance information about entities, activities, and agents, enabling the tracing of data origins and processes in a standardized, machine-readable format;

- *"Data Curation Capabilities"* – Anonymization and data masking techniques will be developed in the data transformation mechanism. A mechanism to add additional components for data curation as plugins will be configurable;

- *"Business Glossary Management"* – The business glossary will be automatically computed based on the vocabularies selected in the modeling process. They will be visible, along with their descriptions and properties, to better help users understand the business terms;

- *"Continuous Data Quality Assessment"* – Linked data and metadata obtained in the ETL process will be validated to match the chosen data schema description – incomplete data will be highlighted; supplemental data quality rules regarding accuracy, deduplication, and relevance can be configured and assessed real-time during the transformation process;

- *"Data Confidentiality Mechanism"* and *"Compliance with Data Security Standards"* – Data confidentiality is ensured through the Privacy-By-Design approach of the system and the curation mechanisms designed in accordance with GDPR and Health Insurance Portability and Accountability Act (HIPAA). In terms of data security, we will use the PrivateSky platform and its encrypted data storage vaults for linked data assets and metadata.

- *"High Visibility of Business Assets through Data Metrics and Reports"* – The Knowledge Hub module will allow different linked datasets to be aggregated and visualized;

- *"Data Lineage Visualisation"* – Data lineage will be visible in the Data Integrator Wizard, where the user builds the transformation process step by step, starting from the template;

- *"Ability to aggregate structured and unstructured data sources"* – Multiple data sources will be configurable with the help of the Airbyte technology, which includes a complex catalog of adapters;

- *"Compatibility with third-party software"* – We plan to provide webhook-based integration mechanisms for notifications about ETL jobs progress, data access, and management changes and data sharing actions;
- *"Trial version"* and *"Community Version/ Free Tier"* – We aim to provide our solution as open source software, to be adapted according to every client's needs.

5 Conclusion

In this paper, five solutions concerned with the data governance processes are reviewed in terms of technical business requirements. Even if most solutions bring a useful tool set, there are concerns about the financial impact of a technical oversight caused by an incomplete data governance framework. Various B2B and B2G use cases, which could benefit from this Semantic Web integration and sharing solution were identified. The current industry approach to the integration of shareable data was documented and compared, in terms of resources invested, with the solution we proposed in this paper. Our proposed architecture and the chosen technology stack were presented in terms of offered features. We also outline the contribution of our solution to the improvement of current B2B and B2G data integration and sharing processes.

The goal of this research is to propose a systemic approach that businesses could follow to define specific organizational data integration flows. The prototype we aim to build allows businesses to design continuous linked data integration pipelines to improve their data governance in the B2B sharing processes. The proposed system has the potential to evolve into an organizational data integration and sharing platform, which validates automatically the compliance of the processes with the established data governance rules. In addition, we wish to obtain an ever-growing knowledge management system, enriched constantly with relevant FAIR data and metadata from multiple data sources, which will help organizational actors to take decisions and follow the appropriate growth policies.

Acknowledgments. This research is co-funded by the European Union, Marie Sklodowska-Curie Actions (MSCA), the HORIZON-MSCA-2021-SE-01 program, as part of the "Elevating Higher Education Public Policies: An empowering SPRIngboard" (HESPRI) project. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. Institutional Project 011301, Information systems based on Artificial Intelligence has supported part of the research for this paper.

References

[1] C.-G. Calancea, "A review on Business-to-Business Data Sharing technical solutions," *International Journal On Optimization and Applications*, vol. 2, no. 1, pp. 33–39, 2022.

[2] K. Wende, "Model for Data Governance – Organizing Accountabilities for Data Quality Management," in *ACIS 2007 Proceedings*, pp. 417–425, Article ID. 80, 2007. [Online]. Available: https://aisel.aisnet.org/acis2007/80.

[3] M. Stonebraker and I. F. Ilyas, "Data Integration: The Current Status and the Way Forward," *IEEE Data Eng. Bull.*, vol. 41, pp. 3–9, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:49407081.

[4] "2025 Outlook: Data Integrity Trends and Insights," Precisely. [Online]. Available: https://www.precisely.com/resource-center/analystreports/lebow-report-2024. Accessed Jan. 06, 2025.

[5] M. Knight, "Data Strategy Trends in 2025: From Silos to Unified Enterprise Value," DATAVERSITY, December 3, 2024. [Online]. Available: https://www.dataversity.net/data-strategy-trends-in-2025-from-silos-to-unified-enterprise-value/. Accessed Jan. 07, 2025.

[6] S. Hyrynsalmi, "Definition of the enterprise integration platforms asa service—towards a common understanding," in *Software Business: 13th International Conference, ICSOB 2022, November*

8–11, 2022, Proceedings, (Bolzano, Italy), Springer, 2022, pp. 167-–181. DOI: 10.1007/978-3-031-20706-8_12.

[7] M. Gottlieb, M. Shraideh, I. Fuhrmann, M. Böhm, and H. Krcmar, "Critical Success Factors for Data Virtualization: A Literature Review," *The ISC International Journal of Information Security*, vol. 11, no. 3, pp. 131-–137, 2019. [Online]. DOI: https://doi.org/10.22042/isecure.2019.11.0.17. Accessed Jan. 07, 2025.

[8] I. A. Machado, C. Costa, and M. Y. Santos, "Data Mesh: Concepts and Principles of a Paradigm Shift in Data Architectures," *Procedia Computer Science*, vol. 196, pp. 263—271, 2022. DOI: 10.1016/j.procs.2021.12.013.

[9] P. Hitzler, "A Review of the Semantic Web Field," *Communications of the ACM*, vol. 64, no. 2, pp. 76—83, 2021. DOI: 10.1145/3397512.

[10] K. Cheng and Y. Sun, *Knowledge Graph Reasoning* (Synthesis Lectures on Data, Semantics, and Knowledge), Cham: Springer, 2025, 196 p. DOI: 10.1007/978-3-031-72008-6. [Online] `https://www.springer.com/series/16928`. Accessed Jan. 07, 2025.

[11] P. P. Tallon, "Corporate Governance of Big Data: Perspectives on Value, Risk, and Cost," *Computer*, vol. 46, no. 6, pp. 32–38, 2013. DOI: 10.1109/MC.2013.155.

[12] D. H.-T. Wong, N. Maarop, and G. N. Samy, "Data Governance and Data Stewardship: A Success Procedure," in *8th International Conference on Information Technology and Multimedia (ICIMU)*, (Selangor, Malaysia), 2020, pp. 54-–61. DOI: 10.1109/ICIMU49871.2020.9243574.

[13] A. Wirsch, "Analysis of a top-down bottom-up data analysis framework and software architecture design," M.S. thesis, Massachusetts Institute of Technology, 2014. [Online]. Available: `https://dspace.mit.edu/handle/1721.1/107346`. Accessed Jan. 07, 2025.

[14] S. Delacroix and N. D. Lawrence, "Bottom-up data Trusts: disturbing the "one size fits all" approach to data governance," *International Data Privacy Law*, vol. 9, no. 4, pp. 236–252, 2019. DOI: 10.1093/idpl/ipz014.

[15] "Complimentary 2024 Gartner® Magic QuadrantTM for Augmented Data Quality Solutions," Informatica. [Online]. Available: https://www.informatica.com/data-quality-magic-quadrant.html. Accessed Jan. 03, 2025.

[16] "Complimentary 2024 Gartner® Magic QuadrantTM for Data Integration Tools," Informatica. [Online]. Available: https://www.informatica.com/data-integration-magic-quadrant.html. Accessed Jan. 03, 2025.

[17] "Gartner Market Guide for MDM Solutions," Ataccama. [Online]. Available: https://www.ataccama.com/analyst-research/gartner-market-guide-for-mdm-solutions. Accessed Jan. 07, 2025.

[18] "2024 Gartner Magic Quadrant for iPaaS: Informatica recognized as a Leader," Informatica. [Online]. Available: https://www.informatica.com/ipaas-magic-quadrant.html. Accessed Jan. 03, 2025.

[19] "Informatica Axon Data governance – Data Governance Use Cases/Challenges/Framework." [Online]. Available: https://success.informatica.com/success-accelerators/informatica-axon-data-governance---data-governance-use-cases-cha.html. Accessed Jan. 07, 2025.

[20] "IBM InfoSphere Information Governance Catalog," IBM. [Online]. Available: https://www.ibm.com/products/information-governance-catalog. Accessed Jan. 03, 2025.

[21] "Collibra Data Governance," Collibra. [Online]. Available: https://www.collibra.com/us/en/products/data-governance. Accessed Jan. 07, 2025.

[22] "Describing the Reference Solution Architecture," SAP Community. [Online]. Available: https://learning.sap.com/learning-journeys/exploring-the-sap-master-data-governance-deployments/describing-the-reference-solution-architecture. Accessed: Dec. 16, 2024.

[23] "Data Intelligence & Data Modeling Solutions," erwin. [Online]. Available: https://www.erwin.com/. Accessed: Dec. 16, 2024.

[24] A. Videla and J. J. W. Williams, *RabbitMQ in Action: Distributed Messaging for Everyone*, New York: Manning Publications Co. LLC, 2012, 312 p.

[25] D. Merkel, "Docker: Lightweight Linux Containers for Consistent Development and Deployment," *Linux Journal*, May 19, 2014. [Online]. Available: https://www.linuxjournal.com/content/docker-lightweight-linux-containers-consistent-development-and-deployment. Accessed: Dec. 12, 2024.

[26] C. Arnaut, M. Pont, E. Scaria, A. Berghmans, and S. Leconte, "Study on data sharing between companies in Europe: final report," Publications Office of the European Union, 2018. [Online]. Available: https://data.europa.eu/doi/10.2759/354943. Accessed: Dec. 12, 2024.

[27] M. Sammeth, N.-C. Ursache, and S. Alboaie, "OpenDSU: digital sovereignty in PharmaLedger," *Front. Blockchain*, vol. 6, 2023. DOI: 10.3389/fbloc.2023.1126978.

[28] B. Harenslak and J. de Ruiter, *Data Pipelines with Apache Airflow*, Manning, 2021. eISBN: 9781617296901. [Online]. Available: https://ieeexplore.ieee.org/document/10280471. Accessed: Dec. 12, 2024.

[29] "Open-Source Data Movement for LLMs," Airbyte. AI Platform. [Online]. Available: https://airbyte.com/. Accessed: Dec. 12, 2024.

[30] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: cluster computing with working sets," in *Proceedings of the 2nd USENIX conference on Hot topics in cloud computing, in HotCloud'10*, USA: USENIX Association, 2010. [Online]. Available: https://www.usenix.org/legacy/event/hotcloud10/tech/full_papers/Zaharia.pdf. Accessed: Dec. 12, 2024.

[31] J. Carroll, I. Dickinson, A. Seaborne, and K. Wilkinson, "Jena: Implementing the Semantic Web Recommendations," in *Proceedings of the 13th International World Wide Web Conference (WWW 2004)*, ACM, 2004, pp. 74–83. DOI: 10.1145/1013367.1013381.

[32] T. Lebo, et al., "PROV-O: The PROV Ontology," World Wide Web Consortium, April 30, 2013, Technical report, [Online]. Available: `http://www.w3.org/TR/prov-o/`, `https://research.manchester.ac.uk/en/publications/prov-o-the-prov-ontology`. Accessed: Dec. 12, 2024.

[33] "What is data lineage?," IBM. [Online]. Available: `https://www.ibm.com/think/topics/data-lineage`. Accessed: Jan. 07, 2025.

Georgiana Calancea[1,3], Lenuta Alboaie[1,4],
Inga Titchiev[2,5]

Received November 19, 2024
Revised January 24, 2025
Accepted March 05, 2025

[1] Faculty of Computer Science, Alexandru Ioan Cuza University of Iasi

[2] Vladimir Andrunachievici Institute of Mathematics and Computer Science, Moldova State University

[3] ORCID: https://orcid.org/0000-0003-0592-2736
E–mail: `gcalancea@info.uaic.ro`

[4] ORCID: https://orcid.org/0000-0003-0969-8102
E–mail: `lalboaie@gmail.com`

[5] ORCID: https://orcid.org/0000-0002-0819-0414
E–mail: `inga.titchiev@math.usm.md`