

Towards a Digital Twin for Simulation of Organizational and Semantic Interoperability in IDS Ecosystems

Patrício de Alencar Silva^{1,2}, Reza Fadaie¹ and Marten van Sinderen¹

¹ University of Twente, Drienerlolaan 5, Enschede, 7522 NB, The Netherlands.

² Graduate Program in Computer Science UERN/UFERSA, Rio Grande do Norte State University (UERN), Federal University of the Semi-Arid Region (UFERSA), Mossoró, RN – Brazil.

Abstract

An International Data Space (IDS) aims to support sharing sensitive data among trusted actors, enabling data owners to control how other agents could use their data, a property commonly denoted as data sovereignty. Data sharing with autonomy is increasingly essential for modern businesses to form ecosystems providing complex services to demanding clients. An IDS ecosystem requires the formation of data-sharing agreements involving different business roles. A data usage contract constitutes a central artifact to formalize this type of agreement. It can also guide actors in implementing or selecting the software components required to enforce data sovereignty. However, there are at least two critical challenges to overcome before forming data-sharing agreements in IDS. First, actors may interpret or represent data usage contracts differently, resulting in a semantic interoperability problem. Second, even assuming semantic mismatches as resolved, contract formation, in this case, would require business process alignment, which leads to an organizational interoperability problem. To address these issues, we envision a digital twin to simulate the formation of data-sharing agreements in IDS, which could support companies exploring semantic and organizational mismatches in this kind of environment. It could also help them assess the risks of adopting and implementing IDS technology. The contribution of this paper is threefold: (1) a research design based on the problem-solving perspective of Design Science; (2) a preliminary architectural model of the digital twin; and (3) a capability assessment of tools for modeling the digital twins envisioned by this research.

Keywords

Data sovereignty, digital twin, enterprise interoperability, international data spaces.

1. Introduction

The notion of data sovereignty involves the control, property, or ownership over data claimed by different agents, ranging from individuals to countries [1]. An International Data Space (IDS), more specifically, is an environment to enforce companies' sovereignty over the exchange of competitive advantage data [2]. IDS may soon turn into a shortcut whereby companies will become trusted by competence, not only long-term business cooperation, to access and use data from one another to, among many possible goals, optimize internal operations and service delivery. In Europe, the International Data Spaces Association (IDSA) is the frontline organization promoting the organizational and technical guidelines to realize the IDS vision [3]. Those guidelines allow actors to assume multiple and overlapping business roles, eventually leading to different corporate alliances, supported by multi-sided data platforms and possibly constituting collective data-sharing agreements [2, 4]. Such contracts could formalize data flow restrictions in an IDS ecosystem.

Proceedings of the Workshop of I-ESA'22, March 23–24, 2022, Valencia, Spain

EMAIL: p.dealencarsilva@utwente.nl (P. de Alencar); m.r.fadaie@utwente.nl (R. Fadaie); m.j.vansinderen@utwente.nl (M. van Sinderen)

ORCID: 0000-0001-6827-1024 (P. de Alencar); 0000-0002-0536-1521 (R. Fadaie); 0000-0001-7118-1353 (M. van Sinderen)



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

[4] propose a set of constraints necessary to form data-sharing agreements in IDS, subdividing them into *contractual conditions* and *data usage contracts*. *Contractual conditions* include commercial clauses (e.g., costs of data usage), legal conditions (e.g., IPR constraints), and service level parameters (e.g., data accuracy). A *data usage contract* should include: (1) an access control policy defining role-based access permissions; (2) a data usage policy constraining data disclosure; and (3) a security profile of authentication and authorization requirements to which a data user should comply. These constraints constitute a starting point for discussing more specific realizations of data-sharing agreements for IDS. It should be possible to enforce those constraints at different managerial levels. While a business network model could prelude the contractual conditions, a data usage contract would constitute a document provided by a data owner to support software-based enforcement of data sovereignty claims over a particular data asset.

However, some practical barriers may arise in establishing a consensus over contractual conditions and data usage contracts in IDS ecosystems. First, those contractual conditions will push organizations to comply with IDS recommendations, such as the Reference Architecture Model provided by the International Data Spaces Association (IDSA) [5]. Such compliance demands business process alignment: an organizational interoperability problem. Second, organizations may interpret data usage contracts differently, leading to semantic interoperability issues. Despite efforts to propose data-sharing standards for specific business domains (e.g., the Open Trip Model (OTM) for Logistics), data owners can attempt to protect the same data assets with different data policies, depending on the targeted data users. These problems may even impact legal and technical aspects of Enterprise Interoperability, as prescribed by the European Interoperability Framework [6]. They can also hinder companies' adoption of the IDS vision and the implementation of its recommended technological infrastructure.

Therefore, the main research question addressed in this paper is: *how to simulate organizational and semantic interoperability problems in the formation of data-sharing agreements in IDS ecosystems?* A digital twin could help address this problem in at least three ways: (1) represent the formal structure of a system of organizational roles underlying the IDS ecosystems; (2) clarify the internal structure of data policies and simulate semantic mismatches in their descriptions; and (3) mimic the discovery and selection of IDS data connectors to enforce data sovereignty in IDS ecosystems. This work takes a problem-solving approach based on Design Science [7]. The specific guidelines to build the digital twin grounds on a methodology for modeling digital twins in the context of Industry 4.0 [8].

The rest of this paper develops as follows. **Section 2** provides a research design with knowledge, technical and practical questions decomposing the main research question stated in the introduction. **Section 3** presents a preliminary architecture for a digital twin for simulating semantic and organizational interoperability issues in IDS ecosystems. **Section 4** brings an analysis of tools for the modeling and implementation of the digital twin. **Section 5** provides preliminary conclusions and proposes directions for future research.

2. Research design

Problem decomposition is an essential part of Design Science [7]. For research in Information Systems, Design Science recognizes at least three types of research questions: (1) *knowledge questions*, i.e., descriptive or explanatory inquiries about the system's phenomena of interest; (2) *technical questions* that depict the state-of-the-art technology for proof-of-concept prototyping; and (3) practical questions expressing stakeholders' demands. The questions found relevant in this research follow.

- **Knowledge questions:** What are the underlying organizational models of IDS ecosystems? What phenomena are relevant to describing or explaining (e.g., enactment of data usage contracts and forming data-sharing agreements)?
- **Technical questions:** What are the architectures, methodologies, requirements, and standards for building a digital twin for IDS-based Logistics ecosystems?
- **Practical questions:**

- *Trade-offs*: What are the risks and benefits of disclosing sensitive data and investing in IDS technology? How could a company mitigate these risks?
- *Sensitivity analysis*: How would changes in bilateral data usage contracts affect the organizational interoperability of an IDS ecosystem? Conversely, how would organizational changes in an IDS ecosystem affect the semantic interoperability of its internal data usage contracts?

The methodology to address these problems will comprehend a triangulation of research methods. A *literature review* will unveil organizational elements of an IDS ecosystem and the state-of-the-art technology for modeling them in a digital twin. *Simulation* with the digital twin will help answer the trade-off and sensitivity analysis questions. The construction of the digital twin will demand specific research methods to guide the modeling of the digital twin and a networked ontology to represent the internal state of the digital twin. Last, *Technical-Action Research* (TAR) will help evaluate the utility of the digital twin in promoting acceptance of the IDS vision by small and medium enterprises.

3. Preliminary architecture of a digital twin for simulation of organizational and semantic interoperability in IDS

[9] stated that *digital twins are cyber representations of objects in the real world that facilitate their development, analysis, simulation, monitoring, persistence, and management*. Kritzinger et al. classified digital twins by the type of communication between their physical system and digital counterparts as (1) *digital models*, where the transfer of information between the physical system and the digital object is manual; (2) *digital shadows*, in which the physical object pushes information automatically to the digital object, but the converse transfer occurs manually; or (3) *digital twins*, where the communication is bilateral and fully automated. Modeling a digital twin depends on the type of problem encountered in the physical system, e.g., innovation, optimization, or repair.

The IDS vision of trusted data exchange ecosystems is essentially innovative. The technical and organizational guidelines provided by IDSA are promising [5], but real-world implementations are yet to come. Companies still struggle to understand the benefits of the IDS vision, which will eventually arrive at the cost of data usage prices and certifications for organizational assets and software components. If companies do not understand such a trade-off, investing resources to pave the IDS vision will become problematic. A digital twin could help address this problem. In this research, we are mainly concerned with Enterprise Interoperability problems that may surge in different levels, such as:

- **Organizational interoperability**: *formation of data-sharing agreements* – a data-sharing agreement in IDS is a multi-sided contract of responsibilities and obligations in dealing with sensitive data from the IDS actors [4]. Organizational interoperability issues may occur in: (1) assigning IDS roles to actual actors; (2) ascribing business activities to roles, e.g., metadata discovery and publication, software, and organizational certification issuance; (3) discovering and selecting data connectors; (4) acquiring IDS-ready labels; (5) ranking organizational models of IDS ecosystems based on economic effectiveness and efficiency criteria.
- **Semantic interoperability**: *semantic reconciliation of data usage contracts* – a data owner may impose different restrictions on multiple data users. Semantic interoperability issues may arise in: (1) reconciling multiple vocabularies to express data policies; (2) aligning different interpretations of the same data access policy; (3) minimizing loss of information due to data inconsistency; and (4) semantic discovery and selection of IDS data connectors.

There are alternative frameworks to guide the modeling of digital twins. This research adopts the Digital Twin Framework for Manufacturing guidelines based on the ISO/DIS 23247 standard [10], which provides a reference architecture to build digital twins with four logical layers supported by reusable software components. Figure 1 depicts a preliminary architecture of a digital twin to simulate the organizational and semantic interoperability issues to be addressed in this research. The description of its layers and components follows in a bottom-up sequence:

- **Observable element**: Unlike the Manufacturing cases described in the ISO/DIS 23247 standard (where an observed part is a physical machine), in this research, this element comprehends a model representing the current state of an IDS-based Logistics ecosystem.

- **Data Entity:** This layer has two sub-entities:
 - *Data Collection Sub-Entity:* A crucial element is the sensitive data asset to exchange. This sub-entity will provide a logical container for Logistics Operational Data described in an industry standard, e.g., the Open Trip Model (OTM).
 - *Control Sub-Entity:* This logical container will provide a process coordination model describing the control flow and the temporal restrictions on how the actors participating in an IDS-based Logistics ecosystem will exchange Logistics Operational Data. These models will support the derivation of Key-Performance Indicators (KPIs) for operational optimization in the ecosystem.

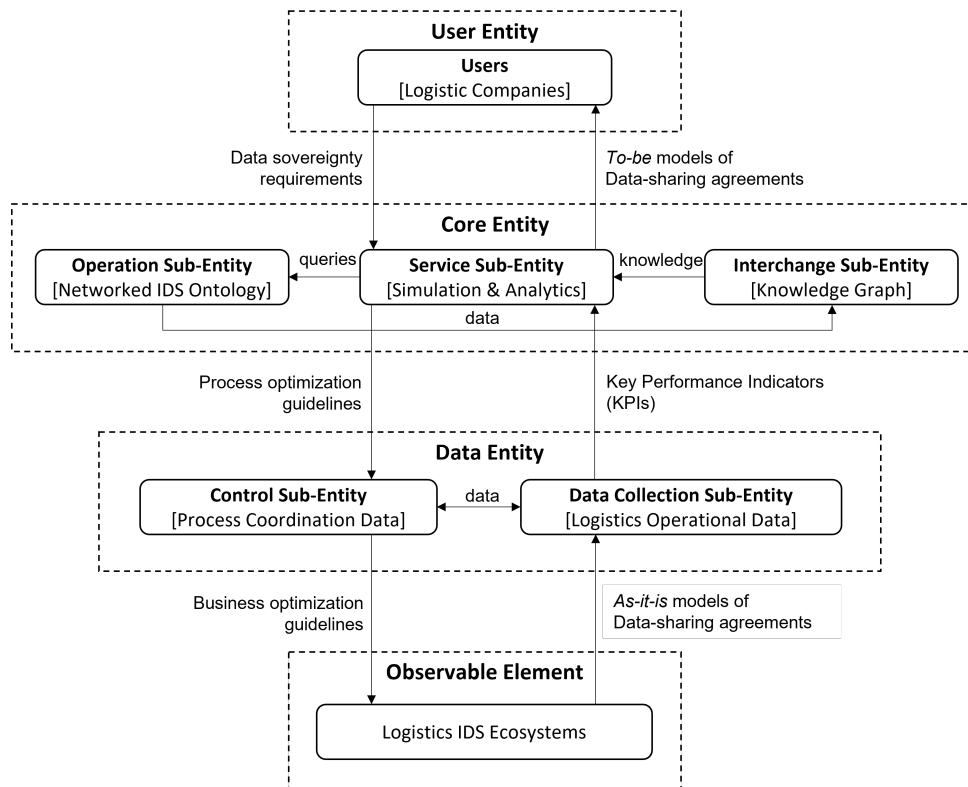


Figure 1. The architecture of a digital twin for simulation of semantic and organizational interoperability issues in Logistics IDS ecosystems

- **Core Entity:** Contains the reasoning mechanisms of the digital twin, aimed to simulate semantic interoperability issues that could impact data-sharing contracts for IDS ecosystems. It has the following sub-entities:
 - *Operation Sub-Entity (Networked IDS ontology):* Three interrelated domain ontologies will describe the operational state of an IDS ecosystem – an ontology to describe the organizational roles of an IDS ecosystem, one to formalize data usage contracts, and yet another one to classify IDS data connectors;
 - *Interchange Sub-Entity (Knowledge graph):* Will aggregate ontology instances and external data sources (e.g., ERP or IoT stream data and events), enabling knowledge inference about the state of an IDS ecosystem;
 - *Service Sub-Entity (Simulation & Analytics):* Will query the ontologies based on the data sovereignty requirements provided by the users and provide feedback in terms of trade-offs and sensitivity analysis of loss of information and sovereignty. It will also provide process optimization guidelines to the components of the data entity.
- **User Entity:** Logistics companies will assess the risks of joining an IDS ecosystem. They could provide requirements for data usage contracts as an input for the digital twin. These requirements constitute their data sovereignty constraints.

The ISO/DIS 23247 standard also recommends cross-layer components to support data assurance, security, and translation as part of the digital twin. In those regards, we make two assumptions. On a first moment, the digital twin envisioned here will not receive input from real-world ERP systems but only experimental data. Secondly, we will treat data translation and security only as part of the data sovereignty requirements expressed in the networked IDS ontology. The following section assesses state-of-the-art tools for implementing and deploying the IDS ecosystem digital twin.

4. Application requirements and tools for digital twin implementation

[11] provide an extensive list of requirements to help assess tools for the deployment and implementation of digital twins. The following criteria are the most relevant for the digital twin project of this research:

- **Continuous integration and deployment:** easy integration of changes from the observable element into the digital twin, possibly avoiding inconsistencies;
- **Domain expert involvement:** easy to use by business analysts and IT architects operating the digital twin without advanced knowledge about its technicalities;
- **Modifiability:** inclusion of new components, such as cross-layer security mechanisms, graphical user interface, or enterprise data repositories;
- **Platform interoperability:** extension of the digital twin platform using value-adding services, e.g., machine learning, simulation, or visualization;
- **Provisioning:** deployment of the digital twin in the cloud/edge computing for external scrutiny;
- **Reusability:** easy inclusion of external software components into the digital twin, e.g., ontology reasoners or machine learning algorithms;
- **System interoperability:** interaction between the digital twin and physical devices. In our case, the observable elements will comprehend Logistics Operational Data coming from an IDS-based Logistics ecosystem. Therefore, the tool should support integrating data from external systems and applications, such as ERP systems and IoT-streamed data.

Table 1 summarizes an assessment of state-of-the-art tools for implementing the digital twin envisioned by this research. We extended the evaluation framework proposed by [11] to analyze Amazon Web Services, Arena, Microsoft Azure, Eclipse, LeanIX, Matlab, and Stardog regarding complete, partial, or total absence of coverage of the requirements.

Table 1

Digital twin modeling tool assessment in conformity with essential requirements extracted from the ISO/DIS 23247 and metrics proposed by [11]

| Requirement | Metric | AWS | Arena | Azure | Eclipse | LeanIX | Matlab | Stardog |
|---------------------------|---|-----|-------|-------|---------|--------|--------|---------|
| Continuous Integration | Offers a complete CI/CD pipeline or integration. | ● | ○ | ● | ○ | ○ | ○ | ○ |
| Domain Expert Involvement | Enables domain experts to model physical devices, their types, and data structured and provides a graphical interface. | ○ | ○ | ● | ● | ● | ○ | ● |
| Modifiability | Allows modification of the DT during runtime. | ◐ | ◐ | ● | ● | ○ | ● | ◐ |
| Platform Interoperability | Provides standardized interfaces to retrieve data and structural information of DTs or make changes to the DTs for external services. | ● | ◐ | ● | ● | ◐ | ● | ◐ |
| Provisioning | The platform provides on-premise and cloud-native solutions for the | ● | ○ | ● | ◐ | ○ | ○ | ● |

| | | | | | | | |
|-------------------------|---|---|---|---|---|---|---|
| | deployment of the digital twin. | | | | | | |
| Reusability | Offers custom and pre-built components that projects can reuse. | ● | ● | ● | ● | ○ | ● |
| System interoperability | Provides support for defining connections between devices via their DTs and automatically implementing interactions based on this connection. | ● | ○ | ● | ○ | ○ | ○ |

[11] referred to Amazon Web Service, Microsoft Azure, and Eclipse as the state-of-the-art tools for digital twin implementation. Recent developments from Arena, LeanIX, Matlab, and Stardog indicate an effort to adapt these tools to become options for digital twin modeling. However, Microsoft Azure currently offers the best coverage for this research project's requirements, specifically continuous integration, user involvement, and system interoperability.

5. Conclusions and future work

This paper provided a research plan to address semantic and organizational interoperability problems that hinder the formation of data-sharing agreements in IDS ecosystems. Solving it demands exploring and understanding Enterprise Interoperability problems, such as inter-organizational process alignment and semantic reconciliation of data usage contracts. We proposed a preliminary architecture of a digital twin to help business and IT architects assess the risks of joining IDS ecosystems for their companies. We have also analyzed state-of-the-art tools for implementing the digital twin, which partially responds to the technical questions derived from the main research question. This research project has three immediate steps: (1) the design of the networked business ontologies to describe an IDS ecosystem, data usage contracts, and (virtual) data connectors; (2) consolidation of value models to describe organizational options to configure the IDS ecosystems (in alignment with process coordination models for specifying the flow of sensitive data); and (3) technical action research for promoting user involvement on validating the relevance of the interoperability problems to simulate.

6. Acknowledgments

This research is financially supported by the Dutch Ministry of Economic Affairs and co-financed via TKI DINALOG and NWO, under the grant N° 439.19.633 (CLICKS project).

7. References

- [1] P. Hummel, M. Braun, M. Tretter, P. Dabrock, Data sovereignty: A review, *Big Data & Society* 8 (2021) 1-17. doi: 10.1177/2053951720982012.
- [2] B. Otto, M. Jarke, Designing a multi-sided data platform: findings from the International Data Spaces case, *Electronic Markets* 29 (2019) 561-580. doi: 10.1007/s12525-019-00362-x.
- [3] A. Braud, G. Fromentoux, B. Radier, O. Le Grand, The road to European digital sovereignty with Gaia-X and IDSA, *IEEE Network* 35 (2021) 4-5. doi: 10.1109/MNET.2021.9387709.
- [4] H. Bastiaansen, M. Kollenstart, S. Dalmolen, T. van Engers, User-centric network-model for data control with interoperable legal data sharing artefacts: Improved data sovereignty, trust and security for enhanced adoption in interorganizational and supply chain is applications, in: *Proceedings of the 24th Pacific Asia Conference on Information Systems: Information Systems (IS) for the Future, PACIS 2020, Dubai, United Arab Emirates*, pp. 1-15.
- [5] B. Otto, S. Steinbuß, A. Teuscher, S. Lohmann, *International Data Spaces: Reference Architecture Model Version 3*, 2019. URL: <https://www.internationaldataspaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf>.

- [6] V. Kalogirou, Y. Charalabidis, The European union landscape on interoperability standardisation: status of European and national interoperability frameworks, in: K. Popplewell, K. D. Thoben, T. Knothe, R. Poler (Eds.), *Enterprise Interoperability VIII, Proceedings of the I-ESA Conferences*, vol 9, Springer, Cham, 2019, pp. 359-368. doi:10.1007/978-3-030-13693-2_30.
- [7] R. J. Wieringa, *Design science methodology for information systems and software engineering*, Springer, Cham, 2014.
- [8] G. N. Schroeder, C. Steinmetz, R. N. Rodrigues, R. V. B. Henriques, A. Rettberg, C. E. Pereira, A Methodology for Digital Twin Modeling and Deployment for Industry 4.0, *Proceedings of the IEEE* 109 (2020) 556-567. doi:10.1109/JPROC.2020.3032444.
- [9] C. Atkinson, T. Kuhne, Taming the Complexity of Digital Twins, *IEEE Software* 39 (2021) 27-32. doi:10.1109/MS.2021.3129174.
- [10] G. Shao, Use Case Scenarios for Digital Twin Implementation Based on ISO 23247, 2021. URL: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=932269.
- [11] D. Lehner, J. Pfeiffer, E.-F. Tinsel, M. M. Strljic, S. Sint, M. Vierhauser, A. Wortmann, M. Wimmer, Digital Twin Platforms: Requirements, Capabilities, and Future Prospects, *IEEE Software* 39 (2021) 53-61. doi:10.1109/MS.2021.3133795.