



UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO
UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO



Towards Legal Interoperability in International Data
Spaces

MASTER'S CANDIDATE

VICTOR BENOISTON JALES DE OLIVEIRA

SUPERVISOR

PROF. DR. PATRÍCIO DE ALENCAR SILVA

CO-SUPERVISOR

PROF. DR. JOÃO LUIZ RABELO MOREIRA

Mossoró-RN

2025

VICTOR BENOISTON JALES DE OLIVEIRA

**Towards Legal Interoperability in International Data
Spaces**

Dissertation submitted to the Graduate Program in
Computer Science - Collaboration between the Federal
Rural University of Semiarid Region, and the the Rio
Grande do Norte State University as part of the re-
quirements for a Master's degree in Computer Science.

Supervisor: Prof. Dr. Patrício de Alencar Silva

Mossoró-RN

2025

© Todos os direitos estão reservados a Universidade Federal Rural do Semi-Árido. O conteúdo desta obra é de inteira responsabilidade do (a) autor (a), sendo o mesmo, passível de sanções administrativas ou penais, caso sejam infringidas as leis que regulamentam a Propriedade Intelectual, respectivamente, Patentes: Lei nº 9.279/1996 e Direitos Autorais: Lei nº 9.610/1998. O conteúdo desta obra tomar-se-á de domínio público após a data de defesa e homologação da sua respectiva ata. A mesma poderá servir de base literária para novas pesquisas, desde que a obra e seu (a) respectivo (a) autor (a) sejam devidamente citados e mencionados os seus créditos bibliográficos.

048t Oliveira, Victor Benoiston Jales de.
 Towards Legal Interoperability in
 International Data Spaces. / Victor Benoiston
 Jales de Oliveira. - 2025.
 108 f. : il.

 Orientador: Patrício de Alencar Silva.
 Coorientador: João Luiz Rebelo Moreira.
 Tese (Doutorado) - Universidade Federal Rural
 do Semi-árido, Programa de Pós-graduação em
 Ciência da Computação, 2025.

 1. Espaços de Dados Internacionais. 2.
 Interoperabilidade Legal. 3. Ontologia. 4. Web
 Semântica. 5. Revisão Sistemática de Literatura.
 I. Silva, Patrício de Alencar, orient. II.
 Moreira, João Luiz Rebelo, co-orient. III. Título.

Ficha catalográfica elaborada por sistema gerador automático em conformidade
com AACR2 e os dados fornecidos pelo(a) autor(a).
Biblioteca Campus Mossoró / Setor de Informação e Referência
Bibliotecária: Keina Cristina Santos Sousa e Silva
CRB: 15/120

O serviço de Geração Automática de Ficha Catalográfica para Trabalhos de Conclusão de Curso (TCC's) foi desenvolvido pelo Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (USP) e gentilmente cedido para o Sistema de Bibliotecas da Universidade Federal Rural do Semi-Árido (SISBI-UFERSA), sendo customizado pela Superintendência de Tecnologia da Informação e Comunicação (SUTIC) sob orientação dos bibliotecários da instituição para ser adaptado às necessidades dos alunos dos Cursos de Graduação e Programas de Pós-Graduação da Universidade.

VICTOR BENOISTON JALES DE OLIVEIRA

Towards Legal Interoperability in International Data Spaces

Dissertation submitted to the Graduate Program in Computer Science to obtain a Master's degree in Computer Science.

DATE OF APPROVAL: 25 / 06 / 2025.

Committee Members

Prof. Dr. Patrício de Alencar Silva
Supervisor and Committee Chair - UFERSA

Prof. Dr. João Luiz Rabelo Moreira
Co-Supervisor - University of Twente

Prof. Dr. Geiser Chalco Challco
Internal Reviewer - UFERSA

**Prof. Dr. Cristine Leonor Pereira
Griffo Beccalli**
External Reviewer - Eurac Research

1 Acknowledgments

After studying such an intriguing convergence of fields—philosophy and computer science—I have come to understand that one does not need to fully comprehend all processes and relationships to embrace life’s profound excitement. For me, the most remarkable phenomenon in its purest form is witnessing God’s work unfold in my life. Throughout these past two years, everything has transformed, yet the more I seek to understand completeness, causality, and relationships, the more I recognize the beauty of mystery. How everything interconnects, how circumstances evolve from periods of mourning and uncertainty into flourishing outcomes—this remains wonderfully beyond full comprehension. Faith transcends understanding; it embodies the profound feeling that everything has its ordained time and place, requiring only our trust in that divine timing. For this revelation, I thank God for guiding me through this transformative two-year journey.

Had someone predicted two years ago that my life would unfold as it has, I would not have believed them, not even in my most optimistic dreams. I cannot complete this reflection without acknowledging the extraordinary friends I have gained throughout this journey. I extend special recognition to my supervisor, mentor, and friend, Dr. Patrício Alencar. Beyond the technical development achieved, the networking and relationships cultivated represent the most valuable outcomes of this master’s program. Among these treasured friendships, I must highlight my dear friend Ryllari Marques, with whom I shared countless hours in the laboratory alongside a thriving friendship that extended well beyond our research endeavors.

Furthermore, I am deeply grateful to my close family circle: my mother, Mrs. Dalvani Dantas; my father, Mr. João Evangelista; and my girlfriend, Ms. Julia Sousa—individuals who maintained unwavering faith and belief in me, especially during moments when I doubted myself.

One of the most significant highlights of this journey was the remarkable collaboration we established with the University of Twente, particularly through my co-supervisor, Dr. João Moreira, who graciously volunteered to support my research. I must also acknowledge another treasured friend born from this partnership, Dr. Sebastian Piest, who encouraged and facilitated my attendance at FOIS 2024 in The Netherlands—my first international academic conference as a master’s student. This experience not only strengthened our professional relationship but also revealed opportunities and recognition I had never anticipated.

I conclude this acknowledgment by expressing sincere gratitude to CNPq for providing the master’s scholarship that supported me throughout these two years. This financial support enabled me to focus entirely on my research while developing as both a scholar and software engineer.

Abstract

The value of data exchange is indubitably a thriving approach, however, it must be conducted in a safe and sovereign space, avoiding the loss of control and unintended use of data. The International Data Spaces is supposed to be a trusted environment in which companies could share sensitive data upholding data sovereignty. Thus, mitigating the risk of losing industrial secrets and further threats to competition. Along with the mentioned two foundations of International Data Spaces, its architecture allows a free contract endorsement, on which, companies may negotiate their policies and governing laws. A service contract should be able to unambiguously represent all involved policies, leaving no breach for subjectivity. Another important aspect of International Data Spaces is to follow the Findable, Accessible, Interoperable, and Reusable (FAIR) principles. In particular, we focus on the Legal Interoperability. As one of the proposed layers of interoperability (intended by the European Interoperability Framework), Legal Interoperability is proposed as the capability of companies from different countries (under different governing laws) to cooperate. Adopting the lack of contractual representation, and common knowledge in a machine-readable way as a problem set, it takes a Design Science perspective for problem decomposition into specific issues, triangulation of research methods, and projection of a solution space. This dissertation presents the comprehensive process of designing, developing, and implementing a domain ontology that formally represents legal interoperability concepts within International Data Spaces environments, the so-called **Legal Interoperability Ontology** for International Data Spaces (*LegION-IDS*). The research foundation comprises a systematic literature review, thoroughly documented and analyzed to establish the current state of knowledge in the domain. Subsequently, the ontology underwent a twofold verification and validation process: initial assessment through mock instances to evaluate accuracy, completeness, and generalizability; and formal validation using three representative use cases extracted from International Data Spaces Association (IDSA) literature. Finally, this work identifies and delineates several promising future research directions, acknowledging the inherent scope limitations of a Master's thesis while establishing a solid foundation for continued investigation in this emerging field. The resulting ontology contributes to the advancement of legal interoperability in data spaces and provides a framework for automating legal compliance assessment in cross-jurisdictional data sharing scenarios.

Keywords: International Data Spaces, Legal Interoperability, Ontology, Systematic Literature Review.

Resumo

A troca de dados é, sem dúvidas um próspero campo de estudo, todavia, esta troca de dados deve ser conduzida em um espaço seguro e que permita a soberania, evitando a perda de controle e o uso inadequado dos dados. Os Espaços de Dados Internacionais propõem um espaço de confiança, no qual empresas possam trocar dados, prevalecendo a soberania de dados. Portanto, mitigando o risco de perda de segredos industriais, os quais ameaçariam à competição justa. Além dos dois pilares mencionados, a arquitetura permite a livre fomentação contratual, na qual, empresas podem negociar as políticas que o compõem, além da lei governante. Um contrato de serviço deve ser inequivocamente representar todas as políticas envolvidas, sem deixar brecha para a subjetividade. Outro aspecto importante dos espaços de dados internacionais, são os princípios de localizável, acessível, interoperável e reutilizável (FAIR). Em particular, focamos no princípio da interoperabilidade legal. Sendo uma das camadas de interoperabilidade propostas pelo framework de interoperabilidade europeu, a interoperabilidade legal é definida como a capacidade de empresas de diferentes países (consequentemente com diferentes legislações) cooperarem. Adotando a falta de representação contratual e falta de conhecimento comum em uma forma legível por computadores como o conjunto problema, empregamos uma perspectiva de design Science para a decomposição de problemas em problemas específicos, triangulação de métodos de pesquisa, e projeção de um espaço solução. Esta dissertação apresenta um processo compreensível de design, desenvolvimento e implementação de uma ontologia de domínio que formalmente representa conceitos de interoperabilidade legal em Espaços de Dados Internacionais, aqui chamada de **Legal Interoperability Ontology for International Data Spaces** (*LegION-IDS*). A fundamentação da pesquisa é delineada por uma revisão sistemática de literature, analiticamente documentada e analisada para estabelecer o atual estado de conhecimento sobre o domínio. Ademais, a ontologia passou por um processo de verificação e análise duplo: uma avaliação inicial com instancias geradas para avaliar a acurácia, completude e generabilidade; e uma validação formal usando três estudos de caso extraídos da literature da IDSA. Finalmente, este trabalho identifica e aponta para diversas direções de pesquisa, tendo em vista o escopo limitado pelo tempo de uma tese de mestrado, enquanto estabelecendo uma sólida fundação para a contínua investigação na área de estudo. A ontologia resultante contribue para o avanço da interoperabilidade legal em espaços de dados e provê uma framework para automação legal em cenários de compartilhamento de dados.

Palavras-chave: Espaços de Dados Internacionais, Interoperabilidade Legal, Ontologia, Revisão Sistemática de Literatura.

List of Figures

Figure 1 – Problem-Solution Set of the Proposed Work.	21
Figure 2 – Data Sharing Process and Involved Roles Summary	24
Figure 3 – Data Sharing Process and Involved Roles Summary	26
Figure 4 – Subset of the IM ontology topography, focusing on the class <i>Contract Agreement</i>	31
Figure 5 – Subset of the IM ontology topography, focusing on the class <i>Contract Offer</i>	32
Figure 6 – Main SABiO’s processes	35
Figure 7 – Systematic Literature Main Phases	39
Figure 8 – Word CCloud	42
Figure 9 – Paper Selection Iterative Lyfecycle	44
Figure 10 – Paper Selection Summary	45
Figure 11 – Number of Publications Over the Years	47
Figure 12 – Future Works Addressing Legal Aspects Distribution	49
Figure 13 – Intended Users Onion Graph	55
Figure 14 – Intended Uses Historical Dependency	56
Figure 15 – Design Foundation of <i>LegION-IDS</i>	57
Figure 16 – Reference Ontology (Main View)	61
Figure 17 – Reference Ontology (Service Contract View)	62
Figure 18 – Reference Ontology (Policy View)	63
Figure 19 – Reference Ontology (Legal Moments View)	64
Figure 20 – Reference Ontology (Legal Moments View)	65
Figure 21 – <i>LegION-IDS</i> development lifecycle and environment	67
Figure 22 – <i>LegION-IDS</i> Classes Hierarchy (Asserted)	68
Figure 23 – Example of <i>disjointness</i> in class <i>DutyRule</i>	69
Figure 24 – Axiomatic representation of the class <i>Service Cotract</i>	70
Figure 25 – Axiomatic representation of the class <i>Contract Negotiation</i>	70
Figure 26 – Axiomatic representation of the class <i>Contractual Agreement</i>	71
Figure 27 – Closure Axiom of class <i>PermissionRule</i>	72
Figure 28 – <i>LegION-IDS</i> ’ object properties hierarchy	73
Figure 29 – <i>LegION-IDS</i> ’ data properties hierarchy	74
Figure 30 – Instances Map (Service Customer)	74
Figure 31 – Instances Map (Service Provider)	75
Figure 32 – CFD Contract	78
Figure 33 – CCM Data Access Contract	78

Figure 34 – CCM Data Access Agreement	79
Figure 35 – SFWR Registration Agreement	79
Figure 36 – DSR treatment lifecycle and next steps	89

List of Tables

Table 1 – Search String per Database	42
Table 2 – Step 1 Inclusion and Exclusion Criteria	43
Table 3 – Step 2 Inclusion and Exclusion Criteria	43
Table 4 – Purpose and Scope of the Ontology	53
Table 5 – <i>LegION-IDS</i> Non-Functional Requirements	58
Table 6 – Competency Questions (SCO and EIF related)	59
Table 7 – Formal and Informal Questions and Equivalent Answers	80
Table 8 – Retrieved Legal Aspects and their Relevance Towards RQs	104

List of Abbreviations and Acronyms

IDS	International Data Spaces
EIF	European Interoperability Framework
IDSA	International Data Spaces Association
RAM	Reference Architecture Model
IM	Information Model
SLR	Systematic Literature Review
DSR	Design Science Research
UFO	Unified Foundational Ontology
GDPR	General Data Protection Regulation
FAIR	Findable, Accessible, Interoperable, and Reusable
ODRL	Open Digital Rights Language
SCO	Service Contract Ontology
SE	Software Engineering
RE	Requirements Engineering
RDF	Resource Description Framework
OWL	Ontology Web Language
ORSD	Ontology Requirements Specification Document
CQ	Competency Question
FOL	First-Order Logics
DL	Description Logics
OntoUML	Ontology Unified Model Language

Contents

1	ACKNOWLEDGMENTS	4
2	INTRODUCTION	14
2.1	Context Setting	14
2.2	Motivation and Supporting Arguments	15
2.3	Design Science Research	16
2.4	Research Questions	16
2.4.1	Research Goals	18
2.4.2	Research Methodology	19
2.5	Document Arrangement	21
3	THEORETICAL FOUNDATION	22
3.1	International Data Spaces	22
3.2	European Interoperability Framework	24
3.3	Ontologies	25
3.3.1	Types of Ontology	27
3.3.2	Unified Foundational Ontology	28
3.3.3	Service Contract Ontology	29
3.3.4	Information Model	30
3.4	Ontology Engineering	32
3.5	SABiO	34
3.5.1	Reference Ontology	35
3.5.2	Operational Ontology	36
4	SYSTEMATIC LITERATURE REVIEW	38
4.1	Protocol	38
4.2	Goals and Scope	39
4.3	Research Questions	40
4.4	Search Strategy	41
4.4.1	Databases	41
4.4.2	Search Strings	42
4.4.3	Study Selection	43
4.5	Quality Assessment and Data Extraction Fields	45
4.6	Results	47
4.6.1	Retrieved Legal Aspects	48
4.6.2	Future Works Addressing Legal Aspects	49

4.6.3	Discussion	49
5	LEGAL INTEROPERABILITY ONTOLOGY	51
5.1	Ontology Requirements Specification Document	51
5.1.1	Purpose and Scope	52
5.1.2	Type and Implementation Language	53
5.1.3	Intended Users	54
5.1.4	Intended Uses	55
5.1.5	Ontology Requirements	56
5.1.5.1	Non-Functional Requirements	56
5.1.5.2	Functional Requirements	58
5.2	Reference Ontology	59
5.2.1	Main View	59
5.2.2	Service Contract View	61
5.2.3	Policy View	62
5.2.4	Legal Moments View	63
5.2.5	Interoperability View	64
5.3	Operational Ontology	65
5.3.1	Environment	66
5.3.2	Class Hierarchy	67
5.3.3	Axioms	69
5.3.4	Object Properties	72
5.3.5	Data Properties	73
5.3.6	Instances Map	74
6	ONTOLOGY VERIFICATION AND VALIDATION	76
6.1	Ontology Validation through Manufacturing Use Cases	76
6.2	Ontology Completeness	77
6.3	Competency Questions	79
7	CONCLUSION	86
7.1	Research Questions	86
7.2	Research Summary and Future Works	88
	BIBLIOGRAPHY	92
	APPENDIX A – ONTOLOGY REQUIREMENTS SPECIFICATION DOCUMENT - ORSD	101

**APPENDIX B – SYSTEMATIC LITERATURE REVIEW PA-
PERS 104**

2 Introduction

In this section, we comprehensively present the context of our research, alongside the Design Science Research overview, i.e., context, motivation, goals, research questions, and methodology for design and validation.

2.1 Context Setting

Companies nowadays try to keep the balance between sharing data among business partners to optimize operations and controlling it for competitiveness and integrity. This reality leads to a discussion about *data sovereignty* - individuals' and companies' ability or power to control who and how one could use their private data (OTTO, 2019). There have been initiatives to enforce data sovereignty in the corporate domain. For instance, an Industrial Data Space has been proposed as an environment where companies could share sensitive data based on mutual trust assumptions (OTTO; HOMPEL; WROBEL, 2019). As business ecosystems evolve, sharing corporate data may cross international boundaries, which motivated the proposition of International Data Spaces – an environment where companies could share data based on competence legitimized by certifications and explicit data usage policies – as defined by the International Data Spaces Association (IDSA)¹. The architecture and implementation guidelines proposed by IDSA are aligned with the European Interoperability Framework (EIF)². The EIF proposes the division of interoperability into six operational layers, the foundational ones (GANCK, 2017), i.e., Legal, Organizational, Semantical, and Technical interoperability, and the recently added, i.e., Interoperability Governance and Integrated public service governance.

As for the scope of our research, we are complying with the foundational layers, responsible for grounding the IDS RAM³. For instance, in an IDS-based business ecosystem, a data usage policy formalizes technical aspects of data exchange (e.g., data formats, standards, and transformations) (GANZHA et al., 2017); data brokers may rely on ontologies to describe, discover, and select data connectors suitable to the needs of data owners or data users (FIRDAUSY et al., 2022c), or enterprise architectures may guide the development of processes and services to leverage Enterprise Interoperability (FIRDAUSY et al., 2022b). However, little attention has been paid to enforcing or promoting legal interoperability in IDS. For instance, the

¹ <<https://internationaldataspaces.org/>>

² <https://ec.europa.eu/isa2/eif_en/>

³ <<https://internationaldataspaces.org/offers/reference-architecture/>>

IDSA Dataspace Protocol⁴ specifies schemas and protocols required from entities to publish data and negotiate data usage policy agreements. However, it lacks explicit guidance on enforcing legal restrictions and compliance in an IDS-based business ecosystem. Moreover, in a multi-organizational scenario different legal norms may impact how actors exchange data in the IDS-based business ecosystem, and data may face interoperability barriers, such as different schemas, different protocols, different descriptions, and subjective representations (CHITUC; AZEVEDO; TOSCANO, 2009).

2.2 Motivation and Supporting Arguments

Not only responsible for managing, maintaining, and certifying the IDS initiative, IDSA is also involved in several types of research, as seen in (OTTO; HOMPEL; WROBEL, 2019)(OTTO; WROBEL; HOMPEL, 2022), and the main architecture such as the Reference Architecture Model (RAM), and Information Model (IM). International Data Spaces facilitate data exchange and linkage in a trusted, protected, reliable, and standardized business ecosystem. The two main aspects of an IDS are data sovereignty and trust. The IDS initiative proposes a reference architecture model for data sovereignty and related aspects, including secure and trusted data exchange in business ecosystems. With numerous data spaces in Europe as well as in China, the Americas, and beyond, an authentic international phenomenon, and these spaces must be trusted to create value from data. Hence, there is a recurrent urge to develop a protocol with international validity.

Although Catena-X⁵ and European Health Data Space (STELLMACH; MUZOORA; THUN, 2022) are good examples of large data spaces, there are smaller data spaces that only exist for fewer days, and with a smaller number of actors. Similarly, data spaces may be centralized, based on an organization or government body, or decentralized, adhering to a common rulebook but not bound by a central association. The proposed protocol should encompass a minimal viable interoperability approach for all the different frameworks, products, or services. Even though IDSA is currently developing the Dataspace Protocol, the legal aspects subject are addressed by a so-called Task Force Legal (GRAS, 2023). Despite the legal approach being developed, there is still a lack of protocols and frameworks to work as a foundation, leading to repetitive and resource-intensive processes for each data exchange agreement.

Projects such as the Eclipse Dataspace Components⁶ aim at providing a framework that could act as a valid reference for third parties, enabling products and

⁴ <<https://docs.internationaldataspaces.org/dataspace-protocol/>>

⁵ <<https://catena-x.net/en/>>

⁶ <<https://projects.eclipse.org/projects/technology.edc>>

services built on top of the framework automatically to implement the Dataspace correctly, thus, being compatible with others using the same protocol. The IDS itself holds two contract samples, which represent (based on German law, a data purchase contract and a data rent contract. Those contracts work as an example for companies to provide their own contracts, referred to as ‘contract freedom’, by (DUISBERG, 2022a). However, this lack of standards leads to time and resource-intensive negotiations, due to contract subjectiveness and ambiguous interpretations. Further, (DUISBERG, 2022a) also states the development of the so-called legal test bed ⁷, which would be able to perform contract negotiation automation. However, it is still a future work. Finally, (MUNOZ-ARCENTALES et al., 2019a) and (WEICHHART; GUÉDRIA; NAUDET, 2016) propose as future work the specification of a policy specification language, which is theoretically addressed by IDSA in the Usage Policy Specification ⁸, but not implemented in the Information Model. Still, there is no current real application and validation regarding the legal interoperability layer.

2.3 Design Science Research

Design Science Research (DSR) consists of a practical research methodology that introduces a composition of a problem, and guided by questions and goals, leads to a set of solutions, through stated methodologies. (WIERINGA, 2014) proposes DSR as the design and investigation of artifacts in context. Furthermore, (BROCKE; HEVNER; MAEDCHE, 2020) defines DSR as a problem-solving paradigm that seeks the enhancement of human knowledge, by creating innovative artifacts. Those artifacts ought to solve specific problems and fulfill a set of requirements. The engineering lifecycle of DSR presets three steps towards the implementation, i.e., problem investigation, treatment design, and treatment validation. The process is meticulously presented in the next sections.

2.4 Research Questions

(WIERINGA, 2009) proposes decomposing a main research question into conceptual, technical, and practical questions for traceability and assessment. Conceptual questions seek knowledge about real-world phenomena without interfering with or changing their internal state, whereas technical questions concern state-of-the-art technology to solve a problem. Finally, practical questions relate to how a software

⁷ <<https://legaltestbed.org/en/start/>>

⁸ <<https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-g/UsageControl/Contract>>

artifact could impact stakeholders' needs. This work aims to treat the following general question:

- **General Research Question (GRQ):** How to Achieve Legal Interoperability in IDS?

Assumptions: We follow the guidelines of the European Interoperability Framework, which sets legal interoperability as a top-level layer of Enterprise Interoperability, i.e., above the organizational, semantical, and technical layers. Treating Enterprise Interoperability issues with a bottom-up approach is possible in this context (and mostly preferred by current research). Still, we plan to follow a top-down approach, starting with the legal interoperability layer. Further, for the scope of this paper, we are addressing the four fundamental legal layers, once the recently added two address public integration, which is not supported by the Reference Architecture Model for now.

Problem Decomposition: The main research question splits into three other major questions, which are explained as follows.

General Conceptual Question (GCQ): *What is legal interoperability in IDS?*

Assumptions: Data owners and users in an IDS-based ecosystem may operate under distinct governing laws for data sharing in different countries.

Problem Decomposition: *(CQ1) What is the current representation of service contracts in IDS? (CQ2) How do legal aspects interfere with Data Sovereignty? (CQ3) How are the legal norms and usage policies currently represented in IDS? (CQ4) What are the legal moments/positions of IDS participants? (CQ5) Is personal-data management approached in IDS architecture?*

General Technical Question (GTQ): *How to effectively enforce Legal Interoperability in IDS?*

Assumptions: Resolving legal interoperability issues in an environment such as an IDS-based business ecosystem will ultimately involve human negotiation, especially with legal aspects, due to its subjectivity. By semi-automatic enforcement, we mean to promote legal interoperability in design time (interoperable by design).

Problem decomposition: *(TQ1) Which machine-readable specification language could address policy representation? (TQ2) How to enforce a common understanding of contractual bindings (policies)? (TQ3) What constitutes the workflow of data exchange within IDS? (TQ4) What is the current state-of-the-art regarding contract automation? (TQ5) how could an application help resolve legal interoperability issues as a prelude to a contractual agreement?*

General Practical Question(GPQ): *How does Proper Legal Interoperability affect IDS-based ecosystems?*

Assumptions: Achieving a contractual agreement in IDS is a time-consuming and onerous process. Companies should be able to fulfill a proper negotiation and agreement, with a common understanding of the parties while endorsing data sovereignty and trust. The IDS architecture provides the possibility of dynamization in contract creation, the proposed dataspace may foster a mean understanding of the governing laws, or the instauration of a new one.

Problem Decomposition: *(PQ1) How could a semi-automated approach facilitate the contract/policies negotiation? (PQ2) What is the common understanding potential of an ontology? (PQ3) What implementations could reduce costs and time towards contract negotiation? (PQ4) What is the impact of establishing the legal availability of participants, to foster the choice or creation of governing law?*

2.4.1 Research Goals

In Design Science Research (DSR) methodology, (WIERINGA, 2009) defines the possible goals of research as social context goals (External stakeholders' goals and Context improvement goals) and the design science research goals, divided into Artifact Design Goal: to design or redesign an artifact, Instrument Design Goal: to design or redesign a research instrument, Knowledge Goal: to answer knowledge questions, and Prediction Goals: to predict future events. The stakeholders' goals and context improvement goals have been acquired by the further presented Systematic Literature Review as future works and gaps. Lastly, the present research encloses the following DSR goals:

Prediction goal: The prediction goal is not directly applicable to this research, however, the future implementation of proposed machine learning models may reach, at some level, the capability of inferring predictions.

Knowledge goal: Identify the ongoing gaps, challenges, and opportunities in the literature regarding the legal interoperability layer when applied to IDS. The knowledge goal is directly aligned with the SLR, answering knowledge (and conceptual) questions.

Artifact Design goal: Develop an expandable and acceptable legal interoperability protocol for data exchange negotiation among countries grounded by different data exchange policies, allowing the unambiguous representation of policies that compose a service contract. As a work in progress, we might move toward the implementation of the proposed ontology, in a Retrieval Augmented Generation (RAG) model, that encompasses the legal interoperability nuances of IDS and leverages the

automation of processes regarding contract automation.

Instrument Design goal: Develop a FAIR (GUIZZARDI, 2020) ontology, which works as a legal base for further application development (ontology-driven development (PAN et al., 2012)). The ontology should leverage legal interoperability, therefore unambiguously representing the legal aspects domain regarding contract formulation, negotiation, and agreement. To do so, we are grounded by the SABiO methodology, which defines a five-step (i.e., purpose identification and requirements elicitation, ontology capture and formalization, design, implementation, and testing) iterative guide for developing ontologies. The first and second steps are responsible for creating a so-called reference ontology, which graphically represents the ontology, and may be designed with languages such as OntoUML ⁹. The design and implementation steps foster the development of the operational ontology, which is the most common approach in ontology engineering (KEET, 2018), ending up with an OWL/RDF operational ontology. The last phase is summarized as testing, but it also comprises its verification, completeness, and validation with stakeholders. Another Important approach of DSR is the clear eliciting of requirements. Requirements may be addressed as functional requirements and non-functional requirements. (SUÁREZ-FIGUEROA; GOMEZ-PEREZ; TERRAZAS, 2009) defines the Ontology Requirements Specification Document (ORSO), as a clear statement of why the ontology is being built, for whom, what the intended uses are, and especially, listing the requirements.

2.4.2 Research Methodology

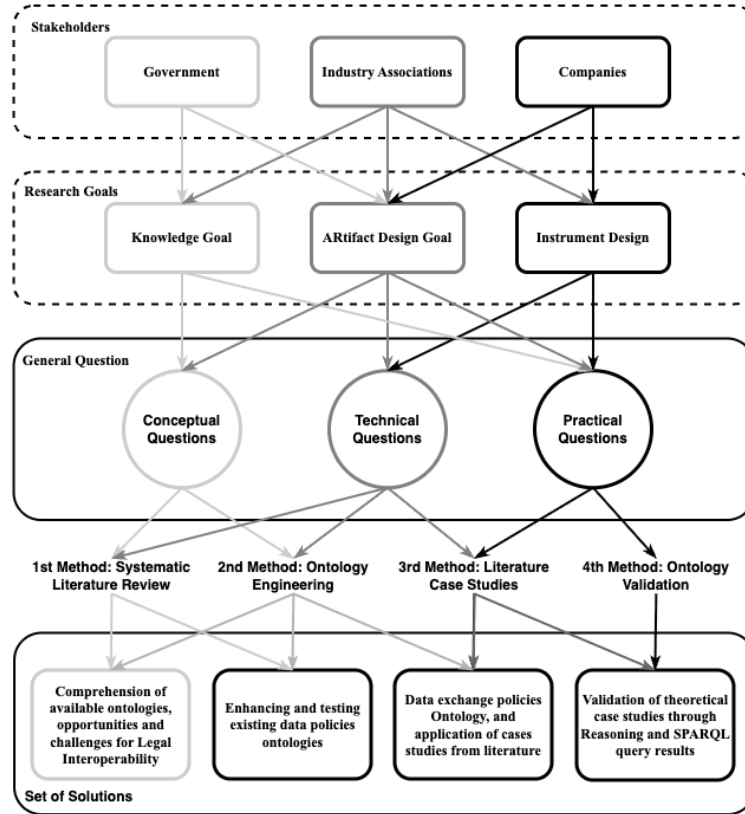
To answer the proposed RQs and foster the next step of DSR, i.e., treatment design, which comprises the specification of requirements, possible contributions to goals, available treatments, and design of new treatments, we focused on the following research methodologies: Review of literature, which, according to (SNYDER, 2019), is the ability to rely on existing valid work, being the foundation of all academic research activity. For this particular research methodology, we proposed a Systematic Literature Review, which is fairly described in 4; Ontology Engineering, in Agreement with (KEET, 2018), plays a critical role regarding the machine-understandable web, and domain representation. Several other methodologies are encompassed, such as Formal Conceptual Analysis (STUMME, 2002), Prototyping (LUQI; STEIGERWALD, 1992), Completeness Test (TAMBASSI, 2021), and others. As for the last step of DSR (i.e., treatment validation), we propose the following methodologies: Case Study, defined by (FEAGIN; ORUM; SJOBERG, 2016) as an in-depth, multifaceted investigation. It allows the resemblance of the theoretical

⁹ <<https://ontouml.org/>>

models applied to real-world scenarios; it is possible to summarize the following methodologies:

- **Literature Review:** *A systematic literature review for comprehending the legal aspects, gaps, challenges, opportunities, and future works encompassed in Legal Interoperability regarding IDS. The proposed SLR is based on the (KITCHENHAM, 2004) guidelines. Furthermore, the retrieved aspects and papers may be used as a database for future implementation of machine learning models, and RAGs.*
- **Ontology Engineering:** *As proposed by (FALBO, 2014), formulate a reference ontology, encompassing the domain of the legal aspects regarding legal interoperability, relying on foundational ontologies such as Unified Foundational Ontology (UFO)(GUIZZARDI et al., 2022), Information Model (IM) (BADER et al., 2020), and Service Contract Ontology (SCO) (GRIFFO et al., 2021a). The ontology should answer open questions spotted on the SLR and fulfill the requirements established by the stakeholders and literature as well. Furthermore, the ontology should be enhanced to an operational version in OWL/RDF, in order to allow further implementation in applications and machine learning models.*
- **Literature Case Studies:** *Instantiate the core principles of the ontology with data from cases retrieved from IDSA literature (LARRINAGA, 2022)*
- **Ontology Validation:** *In-depth understanding of the results from reasoning and queries, along with challenges elicitation.*

Therefore, we may encapsulate our research goals, questions, and solutions road map as follows in Figure 1. Please note that, although we establish a possible prediction goal, it should be further addressed as the work proceeds, and it is not addressed in our proposed road map.

Figure 1 – Problem-Solution Set of the Proposed Work.

Source: “Original Authorship (2025)”

2.5 Document Arrangement

The rest of the paper is organized according to the following: **Section 2** presents the theoretical foundation, as an overview of our scope, a set of supporting arguments, and the foundation of ontology engineering. **Section 3** thoroughly presents the performed Systematic Literature Review, describing its protocol, methodology, and results. Furthermore, **Section 4** introduces the ontology engineering step, fostering the methodology, modeling, development, and showcase. Finally, **Section 5** presents the overall conclusions, along with the roadmap for future works.

3 Theoretical Foundation

The presented section introduces important concepts, for an unambiguous understanding of this document. We present concepts such as dataspace, focusing on the International Data Spaces (IDS), the European Interoperability Framework (EIF), which is a grounding concept for IDS, and the concept of ontologies. Further, we take an in-depth view of the ontology engineering process, which leads to the chosen methodology for development, and further division into referential and operational ontologies.

3.1 International Data Spaces

Over the years, seamless collaboration and information exchange have been the grounding assumption of business models. Massive internet-based platforms have emerged, connecting people worldwide, and exchanging information and data at an outstanding rate. As final users, convenient communication is aspired in a business domain. However, collaboration and data exchange among companies face a sensitive path, with an unprecedented amount of open issues (BADER et al., 2020). Companies realized that to remain competitive, internal and open-source data does not hold enough potential for value creation, leading to a paradox. Sharing valuable and sensitive data carries the risk of losing sovereignty and ownership, hence, the company's competitive advantage, whereas not exchanging data prevents value creation, and undermines upcoming revenue opportunities. There is currently a lack of standardized, interoperable, and trustworthy means to ensure data sovereignty and traceability. The International Data Spaces (IDS) initiative proposes a safe and data-sovereign environment, tackling the requirements above, and privacy concerns.

The International Data Spaces Association (IDSA) not only manages, but proposes specifications, requirements, reference architecture, and different research overviews regarding the IDS initiative. The proposed Reference Architecture Model (RAM) ¹ presents a broad set of implementation policies and guidelines, along with a complete overview of IDS architecture. The main goals of IDS are trust (as the base of the ecosystem, should provide the participants trust in the system itself, but also provide other participants with valuable data), data sovereignty (which is the capability of being self-determined concerning its data, defining who, for how long, and when third parties may use it), data ecosystems (by enabling new business models,

¹ <<https://internationaldataspaces.org/offers/reference-architecture/>>

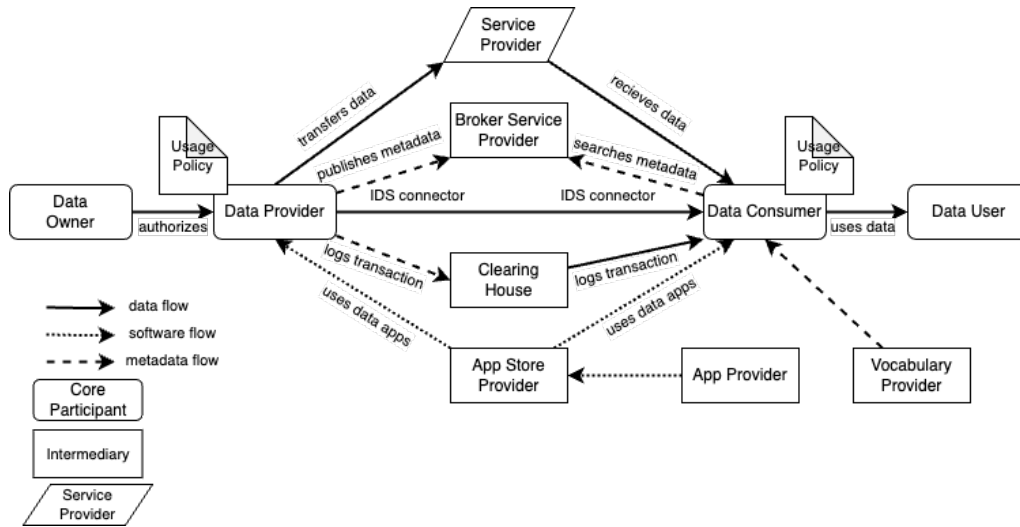
once it provides a broader source of data), and finally, standardized interoperability (allowing different companies to work together, due to different schemas, semantics, policies, and even governing laws (PETTENPOHL; SPIEKERMANN; BOTH, 2022)). The RAM is divided into five layers, i.e., business, functional, process, information, and system. The five layers uphold the perspectives of security, certification, and governance.

The business layer defines and categorizes the possible different roles that participants may assume. In summary, the core participants are involved and mandatory in every data exchange action. Those are:

- **Data Owner:** Legal entity of natural person executing control over data. The data owner must define the data usage policies and provide access to its data.
- **Data Provider:** Makes data available. The main activity performed by this actor is to provide a data consumer with the data owner's data.
- **Data Consumer:** Receives data from a data provider. The data consumer may be defined as the mirror entity of the data provider.
- **Data User:** The legal entity that holds legal control over data. The data user has the legal right to use the data owner's data, under a specified usage policy.
- **App Provider:** Responsible for developing data apps to be used.

Assumption: Usually, a participant acting as data owner also assumes the paper of the data provider. Whereas the data user is also identified as the data consumer. Furthermore, all participants must be certified, fostering trust in the ecosystem. Finally, data sovereignty is guaranteed to the data owner, once it allows the formulation of data usage policies.

Moreover, intermediary actors such as the Broker Service Provider (an intermediary that stores metadata about data sources available in IDS (FIRDAUSY et al., 2022c) and Clearing House (responsible for providing clearing and settlement services for financial transactions). Other possible actors are the Software/Service Providers and Governance Body. Finally, the IDS Connector is the main component of the IDS, used by the participants to establish the connection between the data sources and data providers, and the applications of the data consumers. Figure 2 presents a summary of the data exchange workflow in an IDS ecosystem.

Figure 2 – Data Sharing Process and Involved Roles Summary

Source: “Original Authorship(2023)”

3.2 European Interoperability Framework

The European Interoperability Framework (EIF) ² defines basic interoperability guidelines in the form of simple principles, recommendations, and models. The EIF is the reference standard ³ for interoperability in IDS. The EIF defines interoperability as *the capability among companies to cooperate towards mutually beneficial goals, by sharing information and knowledge, through the business process they support..* The underlying principles of EIF are grounded into **Principle setting the context for EU actions on interoperability** (Subsidiarity and proportionality), **Core Interoperability Principles** (Openness, Transparency, Reusability, and Technological neutrality and data portability), **Principles related to generic user needs and expectations** (User-centricity, Security and privacy, Inclusion and Accessibility, and Multilingualism), and **Foundation principles for cooperation among public administrations** (Administrative Simplification, Preservation of information, and Assessment of Effectiveness and Efficiency).

The EIF proposes the division of interoperability into four layers, a cross-cutting component, and a background layer. The four layers of interoperability are defined as:

- **Legal Interoperability:** Allowing companies with different legal frameworks, policies, and strategies to work together.

² <https://ec.europa.eu/isa2/eif_en/>

³ <<https://internationaldataspaces.org/dataspace-protocol-ensuring-data-space-interoperability/>>

- **Organizational Interoperability:** Alignment of business processes, responsibilities, and expectations in order to achieve a commonly agreed, and holistic goal. Practically, means the documentation and integration or alignment of business processes and relevant information exchanged.
- **Semantical Interoperability:** Preservation of the common understanding of information by its precise format and meaning. It covers the semantic (meaning and relationship of elements), and syntactic (describing the exact format of information to be exchanged)
- **Technical Interoperability:** Aspects regarding the infrastructure of applications linking systems. It involves interface specifications, interconnection services, data integration services, data presentation and exchange, and secure communication protocols.

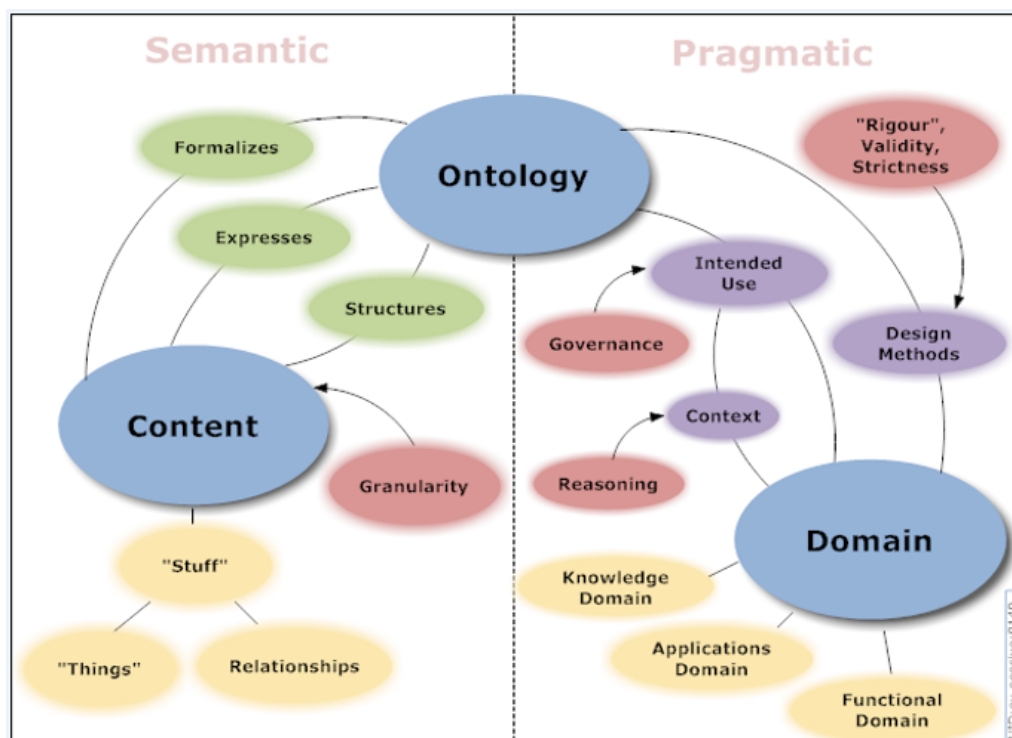
The cross-cutting component called *Integrated Public Service Governance*, and the background layer *Interoperability Governance* have been recently added (2024), and are not considered for our proposed research, once the Reference Architecture Model and Information Model have not recognized them yet.

3.3 Ontologies

The Etymology of the word ‘ontology’ marks as far as back the philosophers’ centuries. The concept of ontology was first introduced as a branch of philosophy that describes what is, how is structured, and what are the relations among the processes of reality (SMITH; ALLEE, 2009). As for computer science, (KEET, 2018) summarizes the definition of ontology to *a text file containing structure knowledge about a particular domain*. The first analogy is conducted to a relational database, however, an ontology holds an in-depth approach to computational systems, even from a practical engineering perspective. (KEET, 2018) also proposes ontologies as the foundation of so-called *intelligent information systems*. (GUARINO, 1998b) defines smart web, as the capability of web systems to retrieve information of the environment, and generate knowledge. For instance, if the user searches for hotels Amsterdam – NL, the engines may be capable of retrieving the concept, and performing inferences such as providing restaurants in the same location, and also the cheapest flight. Furthermore, (BORGIO et al., 2022) presents the concept of ontology as *the basic infrastructure for the Semantic Web. The main role of the ontologies is semantically describing ‘the intended meaning’ of a domain, by defining vocabulary terms. Furthermore, it must present an unambiguous and machine-readable presentation of the content*.

Ontologies have been largely used to solve data integration problems, once the ontologies have the intrinsic capacity to provide a common understanding of a proper domain. Throughout time, researchers figured out other applications for ontologies, being able to solve different problems, such as question-answering systems (i.e., *chatbots*), with an enhanced finding capability (if compared to string and keyword matching). (GUIZZARDI, 2005) emphasizes the difference between the term when used for data modeling, on which, the term conforms to its philosophical definition, obtaining well-founded systems of formal categories, able to properly represent specific domains by its relations and properties. Whereas, for artificial intelligence, software engineering, and semantic web communities, the term is generally used as a concrete engineering artifact, designed for a specific purpose, without considering foundational use, or even a representation of a singular domain expressed in knowledge representation languages (e.g., Resource Description Framework (RDF), Ontology Web Language (OWL)). As previously mentioned, ontologies are usually in OWL, on which, data must be identified by the Uniform Resource Identifier (URI), and the data model is provided in RDF. While the RDF specifies the syntax, the ontologies specify the semantics, adding meaning in a higher layer. Figure 3 presents a template for discourse regarding ontologies, and its connection to content and the proposed domain of representation.

Figure 3 – Data Sharing Process and Involved Roles Summary



Source: "OntologySummit2007's Dimension Map"

3.3.1 Types of Ontology

(GUARINO, 1998b) introduces the classification of different kinds of ontology, according to their level of granularity. As a hierarchy, at the top, we may define the top-level ontologies, which can be used to develop domain and task ontologies, these two can be implemented in an application ontology, as the lower level. The types of ontology are defined as:

- **Top-Level Ontology:** Those ontologies are employed to describe general concepts such as space, time, matter, etc. which are independent of a particular problem, or environment (further described as *endurants*. It is reasonable to unify such concepts in a top-level ontology for a large community of users.
- **Domain Ontology:** Describes a certain domain, by its relationships and hierarchy of classes. This is the most employed type of ontology, once it holds the potential to unambiguously represent a specific domain. A domain ontology may be used as a higher reference when further detailing concepts in the same domain.
- **Task Ontology:** Similar to domain ontologies, task ontologies can represent specific tasks or activities (e.g., selling, negotiations, etc.).
- **Application Ontology:** Conditionally to its domain and task, it seldom specializes the related ontologies. It showcases the concepts corresponding to roles played by *domain* entities while performing a certain activity described in a *task*.

(GUARINO, 1998a) also states the difference between application ontologies and generic knowledge bases. The ontology's purpose is aligned with a particular knowledge base, describing facts, that are often assumed as true by a community of users (i.e., state-independent). Whereas a generic knowledge base may describe facts and assertions related to a particular state of affairs, and status (i.e., state-dependent). (BORGIO et al., 2022), proposes the concept of *Foundational Ontology*. A foundational ontology intrinsically describes a high-level concept, ultimately devoting the mutual understanding and interoperability among people and machines. Therefore, we may define a foundational ontology, as a top-level ontology. In conclusion, the sense of foundational ontology (as a result of human labor, seldom based on philosophy linguistics, and cognitive science) is complementary to the sense of *lightweight ontology* (which can be developed by machine learning techniques).

3.3.2 Unified Foundational Ontology

The reuse and referencing of validated ontologies are advisable for a better foundation. The Unified Foundational Ontology (UFO) is an axiomatic theory (conceptual model) developed by joining several theories of formal ontologies i.e., philosophy, cognitive science, linguistics, and philosophical logic (GUIZZARDI, 2005). In this Foundational Ontology, a fundamental distinction between Types and Individuals is set. While individuals are entities that exist in reality possessing a unique identity, universals, on the other hand, are a pattern of features, that can be instantiated into individuals. Although each individual holds to their identity, they share the features of its universal type. Another important disclosure is the definition of *moment*. A moment relates to an individual that can only exist in other individuals (existentially dependent). Moments are twofold, an *intrinsic moment* is dependent on one single individual, its *bearer* (e.g., color, height, size), Whereas *relational moments*, which are referred to in UFO as *relators*, depend on a plurality of individuals (e.g., a marriage, a service contract, an agreement). Another important nuance regarding UFO is the principle of *identity*. Identity refers to functional characteristics (identity function) that define the essence and nature of individuals. For instance, assuming we consider the famous painting Monalisa, we may picture all the colors, the lines, and the expressions, in a fine frame. However, if at some point in time, the painting has no longer a frame, and all the colors are fading, does it still hold the identity of Monalisa? Further, in some period of time, the painting is reduced to a stained canvas and a broken frame. Is it still Monalisa? Identity refers to conductive thinking, based on a vector of characteristics, it must be a particular individual.

Essentially, it was firstly organized into three main fragments, UFO-A, which is the main fragment, is called the ontology of *endurants* i.e., individuals that exist in time with all their parts, having accidental and essential properties that may qualitatively change while maintaining their numerical identity through time (e.g. a cat might be a kitten in a given time, but fully-grown in another, but still holds onto the identity of cat). Along with the primordial concepts such as particular, universals, and moments, UFO-A also approaches the definition of concepts as relators and relations, all important concepts used as foundation. UFO-B, the ontology of *perdurants* i.e., individuals that accumulate temporal parts, only existing partially in the present, hence, at different time instants, their current properties may vary (e.g., a marriage, which is composed of a temporal series, the wedding ceremony, shared experiences, etc.). The main concept retrieved from UFO-B is the *event*. An event is essentially a transformation from a pre-state situation to a post-state situation, ontologically dependent on its participants. Finally, UFO-C, which is the ontology of social entities, is based on the latter two fragments and sets important definitions such

as the distinction between **Agents** and **Objects**. Agents are twofold, physical (e.g., a human being) or social (a company, a governmental body). Whereas Objects may also be categorized into physical (e.g., a table), and social (e.g., currency). Another example of *social object* is a **Normative Description**. A normative description introduces or more rules/norms intrinsically recognized by at least one social agent, and that can define nominal universals, such as a social moment universal, social objects, and social roles. Examples of Normative Descriptions are the Brazilian Constitution, an internship program regulation, or a service contract).

Ranging from economics (PORELLO et al., 2020) to biology (GUIZZARDI et al., 2021), it has been proven that UFO is a solid foundational ontology, leading to a better understanding of the proposed domain. Furthermore, it has been enhanced with the development of important fragments such as UFO-S, which is the ontology of services (GRIFFO et al., 2021b), and UFO-L (GRIFFO et al., 2021b), which is a legal core ontology. The latter was used as a reference for the construction of the Service Contract Ontology (SCO) (GRIFFO; GUIZZARDI; ALMEIDA, 2018), which along with UFO (i.e., its fragments), will lead the contextual foundation of the proposed work.

3.3.3 Service Contract Ontology

From a legal viewpoint, the analysis of contracts has shown that it is relevant to understand and explain the policies of an endeavor (raising the awareness and compliance of the involved contractual parties), but also to explain the legal positions of each party in a service relation, clarifying their roles, actions, and responsibilities (GRIFFO; GUIZZARDI; ALMEIDA, 2018). The proposed Service Contract Ontology (SCO) stands on those assumptions, and it is based on the concepts and relations of UFO-S and UFO-L. Built upon the notion of Service Agreement, SCO proposes the high-level concept of *Legal Service Agreement*. A legal service agreement can be composed of legal moments, i.e., Hired Service Provider Entitlement, and Service Provider Burden/Lack, Service Customer Entitlement, and Service Consumer Burden/Lack. Further, it specializes in the social notions of claim and commitment for the service provider and customer. The notion of *moment* in UFO is defined by some individual that is existentially dependent from its *bearer*, but also inheres characteristics (GUIZZARDI, 2005). The term holds no relation to the notion of time, acting as an objectified property. The *legal moments* introduced by the SCO are:

- **Legal Entitlement:** *Right*(one-way demand from another the performance of a certain conduct), *Permission* (subject A has permission towards subject

B to act, or refrain from acting), *Power* (subject A has the power to create, exchange, or extinguish legal positions towards subject B), and *Immunity* (subject A is immune to changes in its legal position performed by subject B).

- **Legal Burden/Lack:** *Duty* (subject A must omit, or perform an action), *No-Right* (subject A has the absence of the right to perform action C), *Subjection* (or ‘liability’, if subject A has power against subject B, subject B is in a subjection position), and *Disability* (subject A has no power to create, change, or extinguish a legal position for subject B).

Each legal moment must correlate to another, building the concept of a *legal-relator*. A *legal-relator* bridges the idea of a *relator*, as an individual responsible for changing the status of its parties, and the idea of legal moments. The ontology is formulated onto Alexy’s concept of triadic terms (i.e., holder of the position, holder of the correlative position, and object-action) (ALEXY, 2002). Correlative positions are those with a reflexive position in the same relation, and the correlations proposed are *right-duty*, *permission-no right*, *power-subjection*, *immunity-disability*.

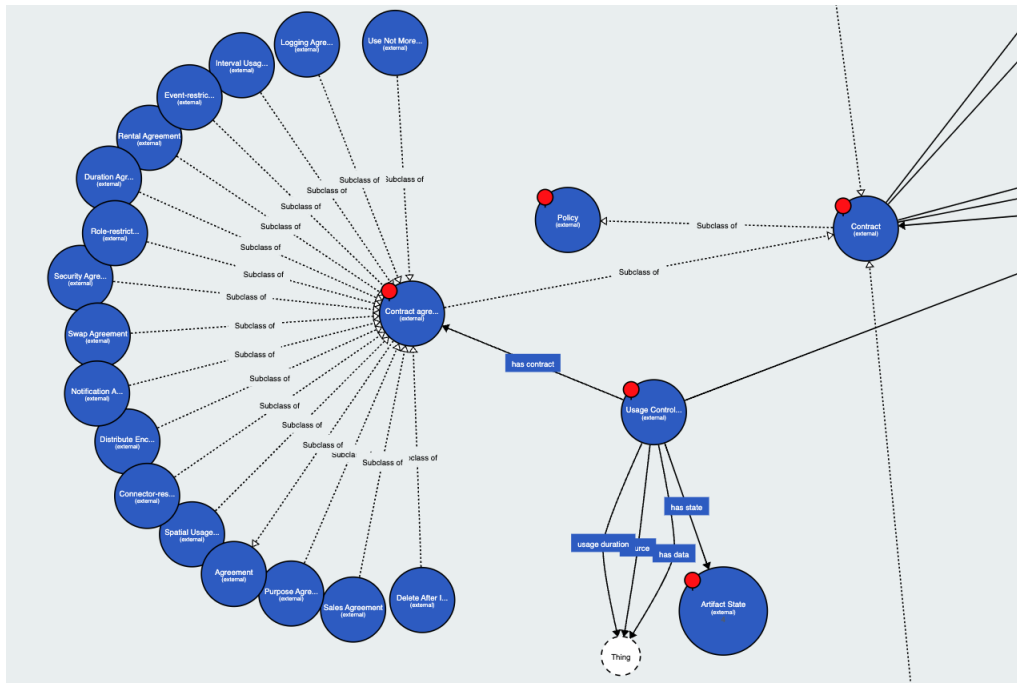
Moreover, a service contract is defined as the arrangement of two or more involved parties (usually a service provider and service consumer), whose purpose is to produce juridical effects, i.e., to create, maintain, extinguish, or modify legal positions. Hence, a service contract is a set of policies, and since it holds an inherently relational notion, Alexy’s relational theory acts as a foundation for an ontological analysis of contracts. As a top-level ontology, SCO will guide the development of the further proposed ontology, leading to perspectives, and future works, applying to the International Data Spaces domain.

3.3.4 Information Model

The IDS Information Model is an RDFS/OWL ontology that encompasses the 6 partitions of the Reference Architecture Model (Content, context, Concept, Community of Trust, Commodity, and communication) (BADER et al., 2020). The IM proposes an overview of the fundamental concepts, describing the actors and their interactions, the resources exchanged, and data usage restrictions. The ontology is further implemented in APIs and holds a community-driven identity (from industry and research organizations) for continuous development. The design principles of the IM are the *reuse* (the foundation of IM is based on refined terms from standard vocabularies, seldom W3C recommendations), *linked data* (IM is published under a stable namespace, in common RDF along with its documentation), *FAIR* (findable, accessible, interoperable, and reusable), and *separation of concerns* (the ontology is divided into 6 modules, addressing each concern that applies to a digital resource).

The Reference Architecture Model is aligned with the concept of usage control and a formal description of permissions and obligations. The Information Model slightly tackles this approach, by implementing the Open Digital Rights Language (ODRL), which provides the terms and concepts for these statements. The author proposes further detailing of these constructs and defining their implication with a keen goal of enhancing the solely descriptive vocabulary to legally bind and enforce statements. Nevertheless, Figure 4 and Figure 5 showcase the topography of the ontology (through the software WebVOWL⁴, and as we may verify, important classes, such as *Policy*, *Contract Agreement*, and *Contract Offer* present several subclasses, but no definition, relationships or in-depth detailing.

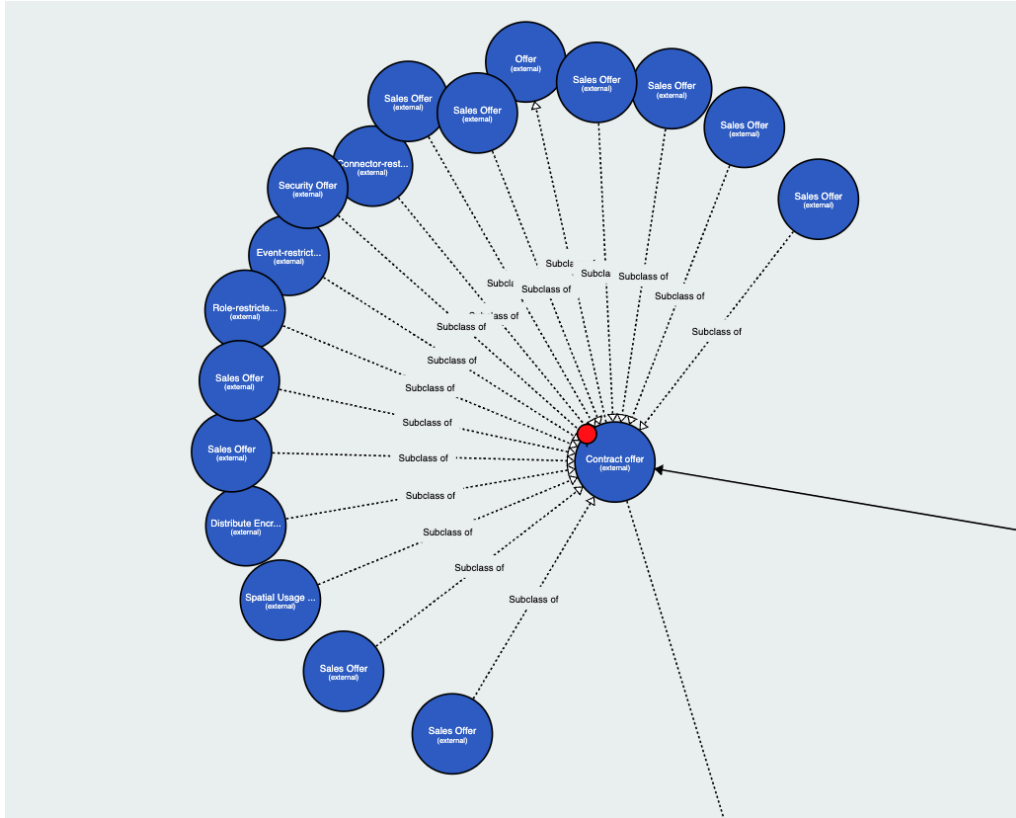
Figure 4 – Subset of the IM ontology topography, focusing on the class *Contract Agreement*



Source: “Original Authorship (2023)”

⁴ <<https://service.tib.eu/webvowl>>

Figure 5 – Subset of the IM ontology topography, focusing on the class *Contract Offer*



Source: “Original Authorship (2023)”

Finally, the author proposes the IM as insightful Semantic Web standards, that present the topics of IDS, and point towards future works, the further enhancement of each layer of the IM. Therefore, we may refer to the Information Model, as a foundation for developing the intrinsic concepts that compose IDS. The final proposed future work is the development of automated extraction tools of IDS IM metadata from the content of data sources, which serve as a functional requirement of the proposed work.

3.4 Ontology Engineering

As observed, ontologies are a subjective field, due to its philosophical and linguistic foundation. The process of developing an ontology is no trivial task as well. (KEET, 2018) proposes that, although there is a lack of ontology development methodologies standards, whichever chosen methodology may result in a good output ontology. (SIMPERL, 2009) points out the main tasks involved in several methodologies, and summarizes in:

- **Ontology Management:** scheduling, controlling, quality assurance, and

feasibility study (problems, opportunities, potential solutions, economic feasibility).

- **Ontology Development and Support:** It tackles the ontology reuse, knowledge acquisition, evaluation, and documentation steps. The methodologies adopted to perform such steps are: *domain analysis* (motivating scenarios, competency questions, existing solutions), *conceptualization* (integration, extension of existing solutions, and model development), and *implementation* (ontology authoring in a logic-based representation language).
- **Ontology Use:** Maintenance (adapting the ontology to new requirements) and Use (ontology-based search, integration, and negotiation).

Several methodologies for ontology development address similar steps, such as:

- **NeOn Methodology:** (SUÁREZ-FIGUEROA; GÓMEZ-PÉREZ; VILLAZÓN-TERRAZAS, 2009) propose a glossary of processes and activities involved in the development of ontologies, 9 different scenarios for developing ontologies, and a set of methodological guidelines for different processes and activities. Those combinations are essentially reduced to (i) the initiation phase, (ii) the design phase, (iii) the implementation phase, and (iv) the maintenance phase.
- **Methontology:** (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; JUZGADO, 1997) proposes a bottom-up approach for building ontologies, which starts with the planning, identification of scope and usage, conceptualization, integration, implementation, and maintenance. The author proposes the following lifecycle: (i) acquire knowledge, (2) build an Ontology Requirements Specification Document (ORSO), (iii) conceptualize the ontology, (iv) implement the ontology, (v) evaluation during each phase, and (6) documentation after each phase.
- **Ontology Development 101:** As one of the most used methodologies, this methodology was the precursor of several others, establishing itself as a pioneering model. (NOY; MCGUINNESS et al., 2001) introduces seven steps towards the engineering lifecycle of an ontology, which are: (i) determine the domain and scope of the ontology, (ii) consider reusing existing ontologies, (iii) enumerate important terms in the ontology, (iv) define the classes and the class hierarchy, (v) define the properties of classes – slots, (vi) define the facets of the slots, and (vii) create instances.

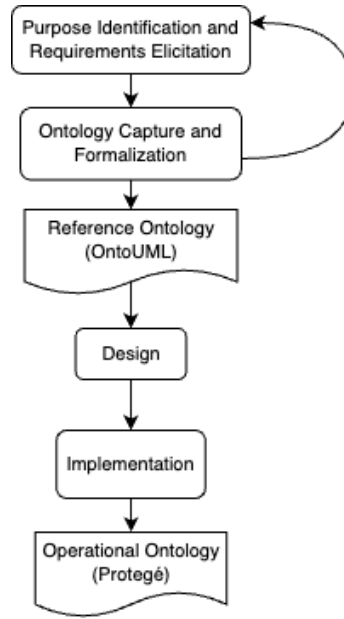
Although all of those methodologies are well-known and well-accepted among developers, they lack systematization. That is the main strength of the Systematic

Approach for Building Ontologies (SABiO). (FALBO, 2014) proposes the development of ontologies based on foundational ontologies (top-level), analyzing the new concepts and relations in the light of a foundational ontology. The proposed methodology guided the process of ontology engineering in our research, due to its systematic approach, and its keen capability of representing knowledge with the implementation of a reference ontology, with further operationalization. This methodology is thoroughly presented in the following section.

3.5 SABiO

This methodology proposes a few benefits by adopting it, such as (i) the rigorosity when defining models based on real-world semantics, (ii) the foundation of the problems within its definition, interpretation, or usage of concepts, and (iii) the possibility of formality improvement (FALBO, 2014). We may encapsulate the main processes approached by SABiO methodology in 5, i.e., the identification of the purpose, and requirements elicitation, the formalization and capture of the ontology, the design itself, implementation, and finally, testing. The first and second phases are iterative and may self-update throughout the process, and after performing them, we reach the first milestone of the ontology development process, the reference ontology. The reference ontology should be specified in a foundational ontology, in our case, the Unified Foundational Ontology (UFO), with further description in the next section. The second milestone of this development methodology is the operational ontology. The operational is the RDF/OWL model, with a machine-readable approach, and in our case, developed in Protegé⁵. To achieve the operational model, we may evolve our reference model through the design and implementation stage, adding functionalities, and nuances available in Protegé, such as data properties, and object properties. Finally, we achieve the last hierarchic process, the testing phase. The main processes are represented in Figure 6.

⁵ <<https://protege.stanford.edu/>>

Figure 6 – Main SABiO’s processes

Source: “Original Authorship (2023)”

Along with the main steps, we might include what is called the support processes. The support processes help enhance the quality and trustworthiness of the developed ontology by addressing 5 optional steps, i.e., the reuse of ontologies, evaluation, configuration management, documentation, and knowledge acquisition. One of the main strengths and differences of this methodology is the modularization of the steps, and the self-enhancement throughout the iterative main and support steps.

3.5.1 Reference Ontology

As mentioned in the previous section, the Reference ontology is the product of the first and second steps of the methodology. The steps are:

- **Purpose Identification and Requirements Elicitation:** This step is usually performed by the ontology users, ontology engineers, and domain experts (stakeholders), and is divided into four activities:
 - *Purpose and Intended Uses Identification*
 - *Requirements Elicitation:* The requirements may be divided into Functional Requirements (FR), which are translated into Conceptual Questions (CQs) that the ontology ought to answer. i.e., defining the scope. Whereas the Non-Functional Requirements (NFR), refer to qualities and general aspects not related to the ontology content. E.g., ontology quality attributes, project requirements, intended uses-related requirements, and

interoperability requirements. While the FRs define the scope and final goals, the NFRs define how fast, or how detailed the information should be retrieved.

- *Competency Questions Identification*: The CQs may be developed in a *top-down* approach –starting at high-level questions, and decomposing them into simpler ones, *bottom-up*– starting with simpler questions, and composing them into complex ones, and *middle-out* approach – the ontology engineer elicits the questions with no specific order, and then compose or decompose them into more or less abstract questions.
- *Ontology Modularization*: consists of the identification of modules (or sub-ontologies) that can be separated while upholding its concept representation. The benefits of modularizing ontologies include easier implementation and maintenance, swifter reuse, and improved performance.
- **Ontology and Formalization**: This phase is performed by the ontology engineers and domain experts (stakeholders), and consists of conceptual modeling, dictionary of terms definition, informal axioms definition, and formal axioms definition.

The relevant concepts and relations should be identified and organized in a graphical model, supporting the communication, meaning negotiation, and consensus establishment with domain experts. This graphical representation is henceforth defined as *reference ontology*. A reference ontology is a conceptual model, grounded by highly-expressive languages, the focus of this model should be the representation adequacy, once the results intend to be used by humans. (FALBO, 2014) propose OntoUML ⁶ as adopted language. OntoUML is a graphical language that is anchored by the Unified Foundational Ontology. As a UML-like notation, the language allows the classification of the *universals* into UFO stereotypes, such as relators, kinds, subkinds, etc.

3.5.2 Operational Ontology

The second artifact proposed by SABiO methodology is the *operational ontology*. The operational ontology is based on the referential ontology and overcomes two phases:

- **Design**: This phase is performed by the Ontology Engineers and Ontology Designers, and it is composed of the technical non-functional requirements

⁶ <<https://ontouml.org/>>

elicitation, implementation environment definition, architectural design, and detailed design. In this phase, the environment of development and implementation language must be chosen.

- **Ontology Implementation:** After choosing the implementation environment and language, this phase regards the action of implementing the ontology.

An operational ontology is defined by the usual representation, in languages such as Ontology Web Language (OWL), in a machine-readable way. Finally, the methodology proposes the testing phase.

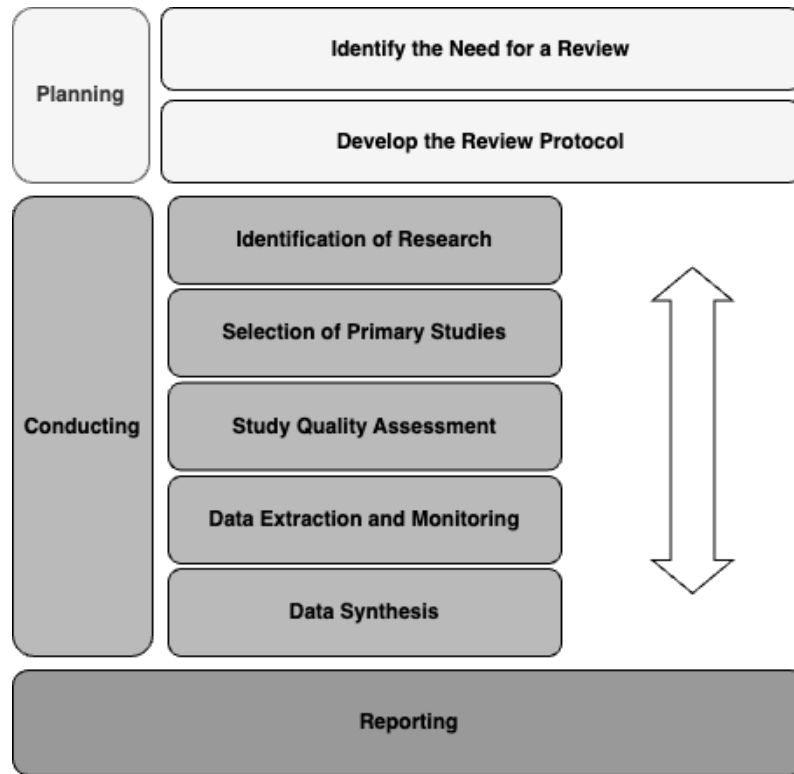
4 Systematic Literature Review

A Systematic Literature Review (SLR) is a specific kind of literature review. The main goal of an SLR is to synthesize research findings, gaps, and challenges in the literature in a systematic, transparent, and reproducible way (SNYDER, 2019). Although SLR has been developed within medical sciences, it is currently used in most fields due to its trustworthiness. As for Software Engineering (SE), (KITCHENHAM, 2004) has proposed a complete guideline for performing SLR for Software Engineering endeavors.

(KITCHENHAM, 2004) defines an SLR as a means to identify, evaluate, and interpret all available relevant research to a specific domain, topic, area, or particular research. An SLR is a secondary study that has different reasons to be performed, such as: to summarize existing evidence and research regarding a singular technology or treatment; to spot gaps in current research, thus, suggesting new approaches and breakthroughs in the field; and to provide a foundation/framework to develop new research activities.

4.1 Protocol

As a primordial step for performing an SLR, defining a systematic and reproducible protocol leverages the potential quality of the review. The guideline summarizes the SLR procedures into three steps, i.e.: planning the Review, Conducting the Review, and Reporting the Review. The Planning and Conducting procedures are associated with self-improving and enhancing steps, and the Report is a single-stage procedure. The lifecycle is presented in Figure 7. Although the proposed lifecycle may appear to be sequential, the process of conducting a Systematic Literature Review (Guided by (KITCHENHAM, 2004)) proposes an iterative approach, being refined on each phase, and is open for improvement when the proper review takes place (represented in Figure 7 by the bidirectional arrow).

Figure 7 – Systematic Literature Main Phases

Source: “Original Authorship (2023)”

The conducted SLR was proposed as a research methodology, targeting the discovery of legal aspects that foster legal interoperability in IDS and similar data ecosystems. The protocol is composed of the definition of goals and scope, research questions, search strings definition, and inclusion and exclusion criteria. The presented SLR was performed through 2023, and its protocol is thoroughly exploited in the next sections.

4.2 Goals and Scope

The presented SLR has as its main goal, to retrieve legal aspects and the state-of-the-art of how policies (hence contracts) are currently formally and informally represented. The review must be capable of answering the further proposed research questions, by leveraging the concept of legal interoperability for the particular domain of International Data Spaces. As mentioned next, the inclusion of data spaces that use IDS technology have also been included, once the scope is limited.

Even though the chosen guideline proposes a population, intervention, and outcome framework, we expanded it to a population, intervention, comparison, outcome, and context (PICOC) (SILVA et al., 2023), aiming toward better readability and reproducibility. The scope is specified as follows:

- **Population:** IDS Participants, Data Governance Specialists, and Business Representants
- **Intervention:** Ontology, Knowledge Graph, and Knowledge Representation
- **Comparison:** Reference Models, Reference Models Architecture, and IDS Information Model
- **Outcome:** Legal Ontology, Legal Interoperability, Data Sovereignty, and Data Exchange Policies
- **Context:** International Data Spaces, Industrial Data Spaces, and Business Ecosystem

4.3 Research Questions

To guide and delimit the Review, founded by (KITCHENHAM, 2004), we may set the following Research Questions (RQ):

- *RQ1. What is Legal Interoperability in IDS?*
- *RQ2. What is the Conceptual relationship between Data Sovereignty and Legal Interoperability regarding Dataspaces?*
- *RQ3. What is the current representation of legal moments in Data/Usage policies within IDS and related data ecosystems?*
- *RQ4. What are the legal challenges within the IDS domain?*
- *RQ5. Are there gaps and opportunities for research and development regarding the legal aspect of IDS?*

With the afore-arranged set of questions, we accomplish the following assessment for checking the question structure quality:

- Meaningful and important to practitioners and researchers: Checking the current version of the so-called Dataspace Protocol ¹, it is possible to infer that it is still a new technology, and there is ongoing research (once it is on its first version);

¹ <<https://internationaldataspaces.org/dataspace-protocol-ensuring-data-space-interoperability/>>

- Will lead to changes in current software engineering practice: At the end of the Review, we expect to map the current approach on how Data exchange agreement ², compliance, and legal interoperability, this way, being able to develop a road map of enhancement.
- Identify Discrepancies between commonly held beliefs and reality: Once IDS is a considerably new initiative, it is growing, and tools such as IDS RAM and Dataspace Protocol are developed in locally available assumptions (It has been performed only in Europe), it may lead to rigid and irreproducible cross country rules, causing ambiguous with the FAIR principle (WILKINSON et al., 2016), and EIF

4.4 Search Strategy

The search strategy is aligned with the goals and research questions. We defined the databases to perform the initial research and elicited the research strings to query through the bases.

4.4.1 Databases

Due to the complexity and uniqueness of the proposed domain of research, we explored the main available databases (LEAL; GUÉDRIA; PANETTO, 2019) such as ScienceDirect, Taylor & Francis Online, and SpringerLink ³, and even expanded the search in order to retrieve the most related works. Other sources were ACM Digital Library ⁴, EI Compendex ⁵, IEEE Digital Library ⁶, and Scopus ⁷. After several trial searches with different combinations of search terms (combinations of the PICOC elements), we randomly selected 10 papers from the database outputs. Then, data analysis has been performed, extracting the text, and populating a word cloud, available in Figure 8, allowing comparison and assessment of the retrieved papers.

² <<https://internationaldataspaces.org/lets-talk-about-idsas-task-force-legal/>>

³ <<https://link.springer.com/>>

⁴ <<https://dl.acm.org/>>

⁵ <<https://www.engineeringvillage.com/>>

⁶ <<https://www.ieee.org/>>

⁷ <<https://www.scopus.com/>>

Figure 8 – Word Cloud



Source: “Original Authorship (2023)”

4.4.2 Search Strings

Finally, we reduced the main string twofold, a specific string for the IEEE Digital Library had to be done, once the first option only retrieved one paper. Even though using the second string we have a broader aspect of papers, the results were satisfactory. Table 1 summarizes the search strings per database.

Table 1 – Search String per Database

Database	Search String
ACM Digital Library, EI Compendex, ScienceDirect, Scopus, Springer Link	("Ontology" OR "Conceptual Model" OR "Taxonomy" OR "Knowledge Graph") AND ("Sovereignty" OR "Enterprise Interoperability") AND ("International Data Space" OR "Industrial Data Space" OR "Business Ecosystem")
IEEE Digital Library	((("Ontology" OR "Conceptual Model" OR "Taxonomy" OR "Knowledge Graph") AND ("Sovereignty" OR "Enterprise Interoperability"))

Source: “Original Authorship (2023)”

4.4.3 Study Selection

Upholding prior knowledge in the field, and initial discussion with specialists, we may establish the first round of inclusion and exclusion criteria, henceforth, referred to as step 1, as follows in Table 2. The first step is performed by reading the metadata available, i.e., title, abstract, and keywords (adaptive reading approach (ALI; PETERSEN, 2014)).

Table 2 – Step 1 Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Papers Written in English	10 Years Time Span
Primary Studies	Duplicated Studies
Studies About Interoperability	Other Literature Reviews
Studies Regarding Data Exchange	Studies Outside Business Ecosystems

Source: “Original Authorship (2023)”

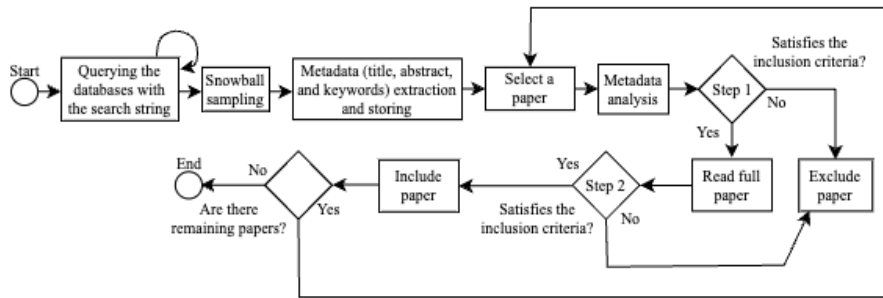
Criteria such as "studies about interoperability" might sound ambiguous, but as the scarcity of the theme requires, as verified in the first step, some papers address legal interoperability intrinsically to the concept of interoperability. IDS is a relatively new field of study (approximately 2016), nevertheless, a similar initiative, the Industrial Data Spaces (AHMADIAN; JÜRJENS; STRÜBER, 2018), are active for a longer period. Hence, to better encompass parallel studies, we may expand the period from the standard of 5 years to 10 years. Finally, it is important to encompass other kinds of business ecosystems in order to extract valid conduct already performed. The second set of inclusion and exclusion criteria, referred to as step 2, includes more specific constraints, as follows in Table 3.

Table 3 – Step 2 Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
It Must Resemble Data Sovereignty (e.g., digital sovereignty)	Studies With no Resemblance With Data Sovereignty or Parallel Concepts
It Must Address Legal Aspects Within A Business Ecosystem	Studies Outside Data Negotiation or Data Exchange Negotiation Scope

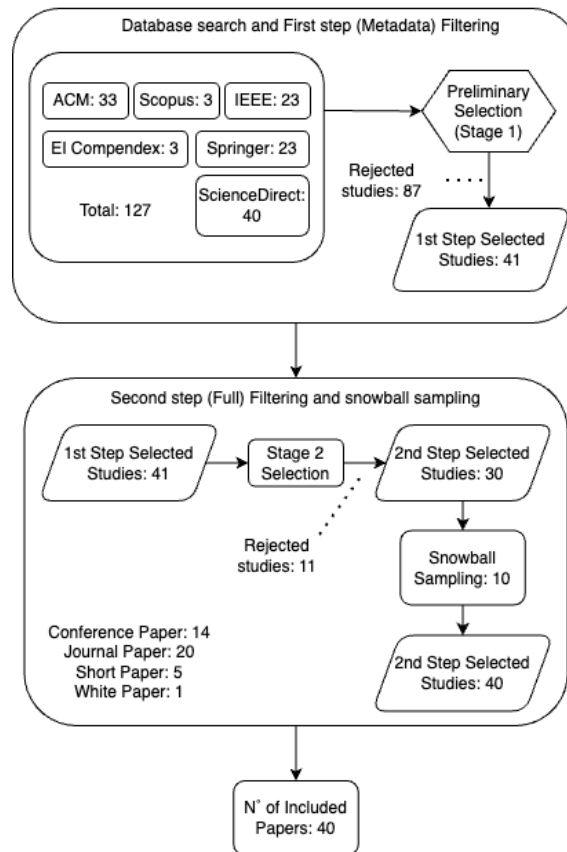
Source: “Original Authorship (2023)”

Finally, the study selection iterative lifecycle is synthesized in Figure 9.

Figure 9 – Paper Selection Iterative Lifecycle

Source: “Original Authorship (2023)”

Performing the developed systematic protocol, the initial search in the digital databases, specifically: ACM Digital Library (33), EI Compindex (3), IEEE Digital Library (23), Science@Direct (40), Scopus (3), and SpringerLink (23). The first set of papers for analysis is composed of 127 papers. Filtering the first set of papers by its metadata (i.e., title, abstract, and keywords), we may assign the first step inclusion and exclusion criteria. The initial result ended up with 41 papers included, hence, 87 papers were excluded. Furthermore, thoroughly reading in full the first set of selected papers, aiming towards answering the aforementioned research questions, resulted in the exclusion of 11 papers, guided by the step 2 set of inclusion and exclusion criteria. Finally, we also performed the addition of papers manually, this technique may be addressed as forward snowball sampling, proposed by (WOHLIN et al., 2022) (identifying new papers based on the works citing the current paper), and backward snowball sampling (identifying new papers based on the references used in the current paper). appending the 10 references, retrieved from prior studies, references, and research registers, the final set of included papers has 40 selected studies. A summary of the previously mentioned workflow is available in Figure 10.

Figure 10 – Paper Selection Summary

Source: “Original Authorship (2023)”

4.5 Quality Assessment and Data Extraction Fields

Strictly following the approached guideline (KITCHENHAM, 2004), and supported by parallel methodologies, such as (??), we may assess the quality of the papers, by rating the studies according to the domain in which they meet different standards of quality. Moreover, favoring quality assessment and data extraction providing a form model. The quality assessment intrinsically encompasses different questions retrieved from viewpoints presented in the research questions, enforcing the need to appraise the quality of each particular study in the same way, minimizing bias. One of the main concerns regarding the quality assessment development in an RSL is the difficulty of reproducing, it. Reproducibility is one of the 4 main aspects of any RSL (i.e., systematic, explicit, comprehensive, and reproducible (MENGIST; SOROMESSA; LEGESE, 2020)), and the subjectivity of the quality assessment must be thoroughly concise and documented. The quality assessment fields are proposed as follows, with its particular founding assumption:

- Does the model refer to any data regulation act? (e.g., GDPR, LGPD) - Encompassing all the papers that refer to some kind of data regulation act,

hence, leveraging its credibility.

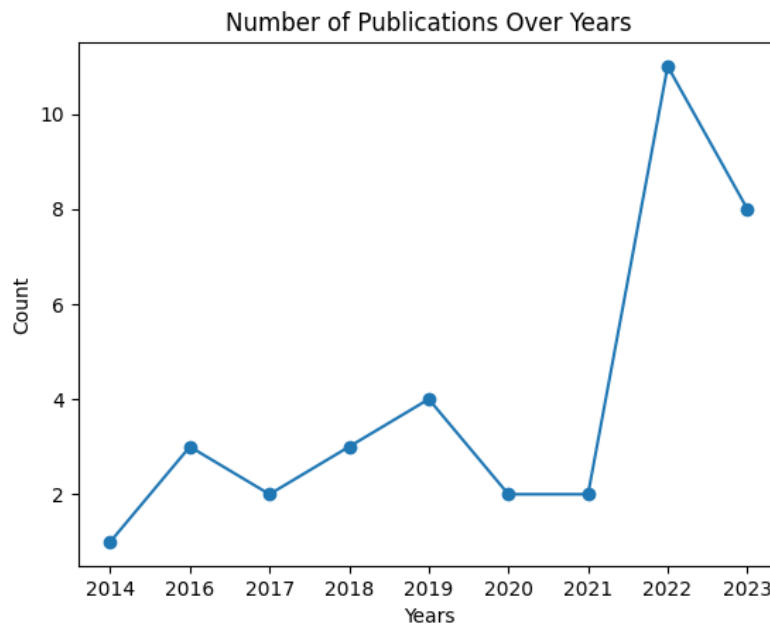
- Does the paper address IDS specifically? Does it cite Industrial Data Spaces or other Industrial ecosystems? - Expanding the context of IDS itself, once other industrial ecosystems conduct might be a good lead.
- Does it mention any legal aspect of the data usage? - Towards the legal aspects of IDS, characterizing one of the main aspects of data sovereignty.
- Does the paper consider or discuss the applicability of the concept in an international context? - Comparing Industrial Data Spaces and other local business ecosystems may constrain the cross-country capability.
- Does the paper address the specific topic of data sovereignty or scattered concepts i.e., Internet sovereignty, digital sovereignty, or cyber sovereignty? - We may include different aspects of sovereignty, once they are all reflected in the definition of power to constrain access.
- Does the paper propose or utilize an ontological approach? - Comparative ontologies are one of the main references targeted in this RSL, in order to spot current gaps and future works.
- Does the paper provide empirical evidence? (e.g., case studies, or practical examples) - If the paper provides empirical evidence, especially focus discussion groups or case studies, the validation of it raises, hence, its trustworthiness.
- Does the paper cover any aspect of legal interoperability? - One of the outcomes of our RSL aims towards the legal interoperability aspect within IDS.
- Does the first author have at least one publication about data sovereignty? - Engagement with the proposed research field enhances reliability.
- Are potential biases or limitations addressed and discussed? - Limitations and difficulties are primordial in a valid work, intrinsically, its credibility.

Each field is qualified with three possible answers, each one with its respective weight, i.e., yes (weight = 1.0), partial (weight = 0.5), and no (weight = 0.0). Some of the qualities are subjective, hence, the evaluation process has been systematically documented and is available in Table 6. Further, the scores for each paper range from 0.0 (minimum) to 10.0 (maximum). Although a few guidelines present the exclusion cutoff (e.g., papers under 4.0 are excluded), the sparseness of the domain requires an inclusive approach, and no cutoff was adopted. In order to avoid bias, on the pilot test mentioned above, at least two authors have read the complete papers, and assessed their quality, to compare such scores and fine-tune the criteria.

Moreover, along with the quality assessment, each paper must undergo a data extraction. The data extraction fields should extract data from the papers that can lead to answering the research questions (KITCHENHAM, 2004). A few fields in the form are standard, such as the authors, publication venue, publication year, etc. We proposed unique 18 fields for extracting data from the papers, i.e.: authors, publication date, venue, paper citations, publication type, mention of multi-country scenarios or datasets, involvement of law experts, the main goal of the paper, whether it represents a formal approach, its relation to governmental initiatives on private data usage, references to interoperability frameworks, additional notes (e.g., regarding the EIF), focus on specific layers, business ecosystem type, associated domain, future works, referred nationality, and adherence to fair principles.

A detailed version of the protocol (such as all the data fields extracted) and all the code involved in the data analysis is available in the GitHub Repository ⁸. Finally, Figure 11 represents the number of publications over the past 10 years (time constraint of our SLR).

Figure 11 – Number of Publications Over the Years



Source: “Original Authorship (2023)”

4.6 Results

As cited in section 4.5, each work undergoes a quality assessment, and the scores ranged from 3.0 (minimum) to 9.5 (maximum), with an average of 5.5. After reading the 40 selected papers, we may classify the kinds of dataspace they refer to.

⁸ <https://github.com/VictorBenoiston/towards_legal_interoperability_SLR>

It is noticed by analyzing the literature that different dataspace use IDS technology, e.g., Open Data Spaces (KIRSTEIN; BOHLEN, 2022) and Gaia-X (??). The results have shown that 22 (39.3%) of the analyzed literature, refers to IDS itself, 8 (14.3%) cite Gaia-X, 7 (12.5%) refer to Industrial Data Spaces, 6 (10.7) mention generic dataspace and 9 (16.1%) allude other kinds of dataspace or business ecosystem.

The overall results come from 15 journal papers, 16 conference papers, 6 book chapters, and 3 white literature papers. The research is mostly conducted in Germany with 25 mentions, China, Finland, Austria, and The Netherlands with 3 mentions each. Along with RAM, and the Information Model ⁹, IDSA also provides the Data Spaces Radar ¹⁰, which embraces the currently active IDS worldwide. In December of 2023, 61,54% of the data spaces were located in Europe, especially in Germany (8), Italy (7), France (7), The Netherlands (7), and Spain (7). Hence, the outcome indeed reflects these numbers, and although a few countries (e.g., Brazil) do not possess an active IDS, it is still on the research track.

4.6.1 Retrieved Legal Aspects

After thoroughly analyzing the 40 selected papers, and being grounded by (KITCHENHAM, 2004), it was possible to observe that some of the papers addressed the same aspects/approaches. Henceforth, we selected aspects that relate to our proposed research questions, and legal aspects within data spaces. We summarize into 10 Retrieved Legal Aspects (RLA) in the proposed SLR. The final list of retrieved aspects was introduced by the authors, as a possible summary with a possible domain capacity of properly answering the RQs. The recovered RLAs were:

- Domains of Business (9 aligned papers).
- Personal/Non-Personal Data (7 aligned papers).
- Usage and Data Policies (7 aligned papers).
- Legal Interoperability Constraints (9 aligned papers).
- Smart Contracts/Contract Automation (7 aligned papers).
- Semantic Appeal (10 aligned papers).
- AI Usage in IDS (9 aligned papers).
- Cloud (8 aligned papers).

⁹ <<https://github.com/International-Data-Spaces-Association/InformationModel>>

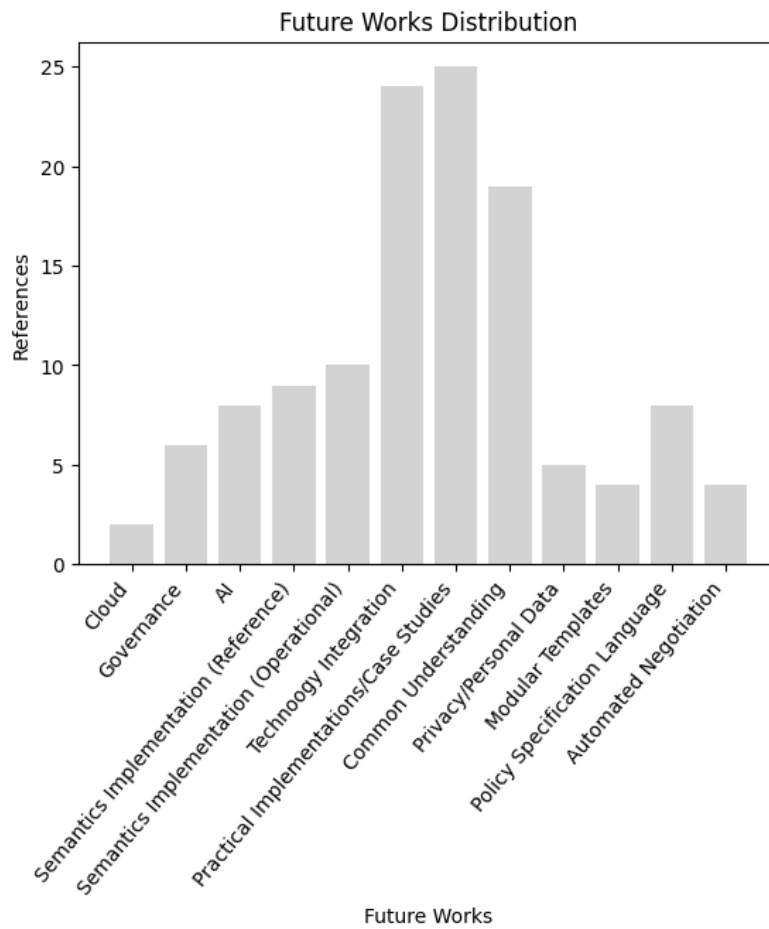
¹⁰ <<https://www.dataspaces-radar.org/radar/>>

- IDS Usage in Open Spaces (4 aligned papers).
- Future Works Addressing Legal Aspects (15 aligned papers).

4.6.2 Future Works Addressing Legal Aspects

Analyzing the future works established by the proposed 40 papers is a clear methodology to find out what the current gaps in the literature are. In order to summarize our findings, Figure 12 presents the distribution of future works proposed by the analyzed literature. An in-depth presentation is available in the previously mentioned GitHub repository.

Figure 12 – Future Works Addressing Legal Aspects Distribution



Source: “Original Authorship (2023)”

4.6.3 Discussion

As a scattered topic, we included journal papers, workshop papers, deep screening of white literature, and even book chapters. Nonetheless, maintaining a good level of fit, quality, and relevance. The overall results indicate that the legal aspects regarding data spaces, and IDS specifically are limited. Although the

EIF claims legal interoperability as the fundamental layer, there is a scarcity of research being performed. If we return to the protocol of this SLR, the first strings of search have returned no more than 50 papers, and after shortly reading them, fewer are framed on the actual scope. The results suggest that we may answer RQ1 as the semantic capability of unambiguously representing policies and rules among companies while respecting the so-called state sovereignty. Overall, the results show that RQ2 refers to the concept of sovereignty over data, i.e., the data provider should be able to determine who, by how long, and to what end the data consumer may use its data, and data pricing (with financial, or data exchange ends). Furthermore, a common and compliant set of rules must be set in a contract agreement, and this contract is usually classified as data rent or data purchase. The answer for RQ3 is twofold, the formal approach proposes a set of policy representation languages, especially ODRL, which is represented in a machine-readable way by the IM itself, even though, it carries a low detailing power. Whereas the informal approach tends towards the so-called legal contract, which encompasses the usage policies, constraints, and legal moments, in natural language.

For RQ4, the legal challenges of IDS cannot be summarized in a straightforward answer, but in a set of nuances such as the lack of formal representation for contracts, personal data exchange and usage, and especially, the absence of an unambiguous domain definition, conditioning the understanding to subjective matters, such as human interpretation. Finally, for RQ5, the main gaps spotted through the SLR were the development of an unambiguous language model for policies, knowledge representation through a semantical approach, smart contracts, and usage of AI and Cloud within IDS, with the main goal of promoting legal interoperability. It is important to recognize the superficiality of research being conducted within IDS itself, despite that, we may point out several other data spaces using IDS technologies while fostering the legal domain, and we may trace those commitments to IDS. For the scope of this document, we are summing the results to foster our research. For an in-depth approach, please refer to the proposed complete SLR paper.

5 Legal Interoperability Ontology

This chapter describes the ontology engineering steps. Firstly, by introducing the actions performed in each step of the SABiO methodology (FALBO, 2014). Providing the foundation, goals, intended uses, intended users, and more, the first step towards the development of a domain engineering is the identification of purpose and requirements elicitation. Several methodologies for developing ontologies propose, at some level, the elicitation of requirements; however, not in a systematic approach. The proposed ontology is hereby referred as **Legal Interoperability Ontology** for International Data Spaces (*LegION-IDS*)

5.1 Ontology Requirements Specification Document

Suárez-Figueroa, Gómez-Pérez e Villazón-Terrazas proposes a systematic approach to document and set goals, granularity, and vocabulary for the proposed ontology (SUÁREZ-FIGUEROA; GÓMEZ-PÉREZ; VILLAZÓN-TERRAZAS, 2009). The lack of direction and well-documented goals and foundations for ontologies may cause *construct deficit* or *construct overload* (GRIFFO et al., 2021b). Both scenarios are prejudicial, as deeply-axiomatized ontologies are not computationally efficient, and shallow ontologies have no completeness potential, leaving open competency questions. The main goal of domain ontologies is the *adequate description of a domain by its relationships*. As previously stated, the term *ontology* etymology is grounded by elder philosophical approaches, aiming at the description of high-level concepts, such as space and time (BORGIO et al., 2022). Such a description may cause the aggregation of several distinct domains, e.g., for the purpose of describing the human body, we might pragmatically represent this domain by dividing it into internal organs and external organs. This could be enough to answer competency questions such as *what are the inner organs of the human body?*, however, this approach becomes deficient for questions such as *what is the composition of human cells?*. This is called *granularity* (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; JURISTO, 1997). An ontology might grow with infinite layers, depending on the granularity. A domain ontology should be able to answer its competency questions with a proper description, which should be documented in the Ontology Requirements Specification Document (ORSO). The goal of these checks is to guarantee the conciseness and completeness of the ontology, aligning with the proper requirements. Finally, (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; JURISTO, 1997) proposes a nested evaluation of completeness, once its achievement is subjective (someone at a given time, may propose a new

relevant item to be included). The ORSD must have the following properties:

- **Concision:** The relevance of each item, and the nullity of duplicated or irrelevant items.
- **Partial Completeness:** Related to the coverage of the terms and the stopover problem (avoiding construct overload or deficit), being guided by the conceptual questions.
- **Consistency:** Refers to all terms and their meanings providing sense to the domain.

Furthermore, a framework for evaluating the ontology is also provided by the authors. The evaluation refers to the technical judgment of the ontology and is composed of the *verification* and *validation* steps. Despite the semantic similarity, the verification refers to the technical process that guarantees the correctness of the ontology, associating the software environments and documentation, whereas the validation refers to the correspondence of the ontology and its represented domain. Moreover, aligned with the Requirements Engineering (RE) (LI et al., 2014), the ORSD is a fundamental step towards developing an ontology from scratch. Those specifications are set aiming the identification of what particular knowledge should be represented in the developed ontology, thus, proposing the ORSD to set boundaries and avoid the aforementioned issues. This subsection will provide a detailed description of the ORSD, and its sections, i.e., purpose, scope, implementation language, intended end-users, intended uses, ontology requirements (functional and non-functional), and a pre-glossary of terms. The complete version of the presented ORSD is available in Appendix A A.

5.1.1 Purpose and Scope

The purpose of developing the present ontology is to capture the policies, compliance, and assumptions of the legal agreement regarding the contractual encapsulation of the digital service in international data spaces while upholding the principle of Data Sovereignty. Such an ontology might be proposed as a treatment for a few open questions regarding the legal approach in IDS exploited by the International Data Spaces Association (IDSA) itself, through the Information Model (IM) ontology and the IDSA Reference Architecture Model (RAM). The scope of an ontology should tackle the degree of formality (i.e., depending on the development technique), the granularity i.e., the degree of details, and a glossary of terms. Further, based on a *middle-out* approach, we may gather the set of terms to be included with no historical order, then rearrange them after a better understanding and pilot

development of the ontology. Performing this technique allows the ontology engineer to spot gaps in important terms, and to avoid the misuse or overuse of synonyms.

Considering the above, the formality of our ontology ranges from semi-informal to semi-formal (LI et al., 2014), once we have, respectively, a restricted and structured form of natural language, and a formally defined language. Further, the granularity follows the *middle-out* approach by gathering pilot data and definitions, then being derived, or scaled to more complex, or less complex concepts. (LI et al., 2014) presents the concept of *granule*, as a basic element, or a *leaf*. Each *granule* is partitioned by an object and set together by similarity, proximity, or other metrics. The *identity* of an entity may reflect on partitions of concepts that they describe, which is called *granular topology of the ontology*. Each topology describes a partition of the concept. In our domain, we might exemplify the concept of *governing law*, which is composed of a *data protection law* and a *competition law*. Each represents a topology, which might be further described in policies, which have rules, and so on. Lastly, the scope of the ontology is bounded by the description of legal concepts, legal agreements, legal lacks/burdens, and entitlements (GRIFFO; GUIZZARDI; ALMEIDA, 2018), data sovereignty, contractual compliance, and upholding legal norms. A summary is proposed in Table 4 (Subset of the ORSD).

Table 4 – Purpose and Scope of the Ontology

1. Purpose
To provide a legal interoperability domain description in order to foster an unambiguous representation of policies, i.e., service contracts, within International Data Spaces architecture.
2. Scope
A knowledge-based approach ranging from a semi-informal to a semi-formal degree of formality. Based on a middle-out architecture. The level of granularity is defined by a continuous integration with the NFRs and FRs. The scope of the ontology is bounded by the description of legal concepts, legal agreements, legal lacks/burdens and entitlements, data sovereignty, contractual compliance, and grounded by legal norms. The proposed ontology is strictly a domain ontology, with foundational integration of top-level ontologies, such as UFO.
Source: “Original Authorship (2025)”

5.1.2 Type and Implementation Language

As stated in Section 3.3.1, there are four types of ontology, i.e., top-level, domain, task, and application ontologies. Our ontology is strictly a domain ontology, once we propose a granular description of the legal interoperability domain in the IDS context. Furthermore, one may infer, correctly, that by using a foundational ontology such as UFO, a top-level ontology, we are supported by both approaches.

The selected framework for ontology development SABiO describes the Ontology development process in five development phases, and five support processes, that are made to refine the phases without a sequential overflow. Further, the ontology is retrieved in two different artifacts: the reference Ontology, and the operational ontology. For this ontology, we propose the reference ontology implemented in OntoUML¹, which is a graphical extension of the Unified Foundational Ontology (UFO)², with further translation for the operational ontology in Web Ontology Language (OWL)³.

5.1.3 Intended Users

The ontology aggregates different aspects regarding legal agreements encompassed in the data negotiation/exchange process among actors under different legal jurisdictions. At first, the primary glimpse of end-users would be the IDS participants, and potential participants, which would rely on an easier way to assess the digital service contract composed of the legal burdens and legal entitlements (GRIFFO; GUIZZARDI; ALMEIDA, 2018) before the service commitment (compliance), hence, facilitating the legal terms negotiation. However, after a deeper understanding of the field, it makes sense that data governance specialists, lawyers, and regulation bodies could also count on the ontology, once it refers to legal aspects, covering legal norms from different countries and setting up contractual aspects. Finally, enthusiasts and researchers of the field might also lean on the ontology for the maintenance and continuity of future works. Therefore, we may encompass the intended uses as follows:

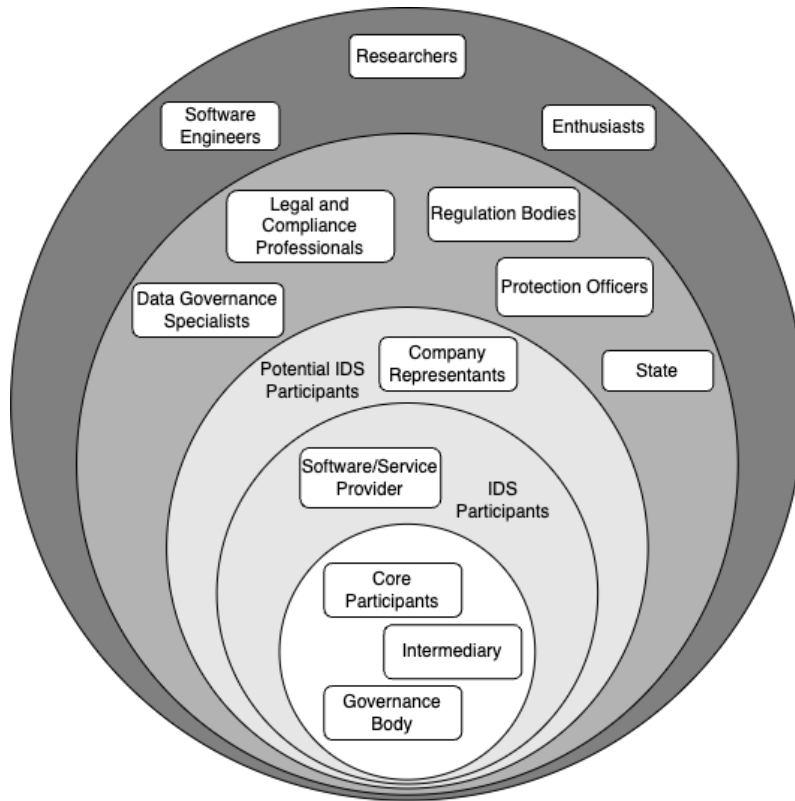
- User 1. IDS participants
- User 2. Potential IDS participants
- User 3. Legal and Compliance Professionals
- User 4. Data Governance Specialists and Protection Officers
- User 5. Software engineers and researchers who wish to acknowledge or expand the domain of legal interoperability in IDS

Figure 13 provides an overview of intended users distribution.

¹ <<https://ontouml.readthedocs.io/en/latest/classes/index.html>>

² <<https://ontouml.readthedocs.io/en/latest/intro/ontouml.html>>

³ <<https://www.w3.org/OWL/>>

Figure 13 – Intended Users Onion Graph

Source: “Original Authorship (2025)”

5.1.4 Intended Uses

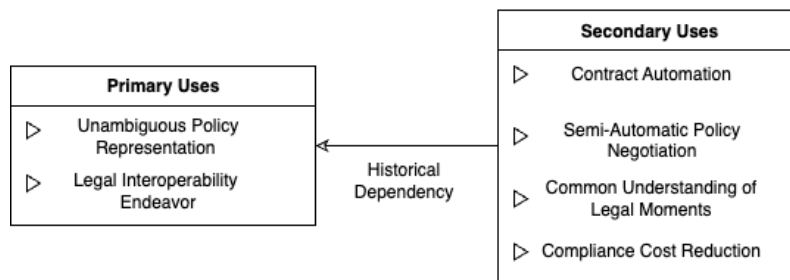
The intended uses play a key role in defining the boundaries of the ontology. As a goal-oriented software artifact, the ontology development process is primarily driven by the situations in which the ontology will be applied—that is, its intended uses. These usage scenarios are subsequently used to inform the following sections of the ORSD, namely the ontology requirements and the competency questions. For this step, intended uses were identified through a review of the literature, where future work and open challenges were proposed by stakeholders in the respective papers. Furthermore, the proposed SLR (Section 4) aimed to extract open future works, challenges, and opportunities, which among others, define the intended uses.

As retrieved in our SLR, the current research lacks the *unambiguous policy representation*. The SLR also introduces the so-called *IDS Usage Control Language*⁴, which is a machine-readable language based in the ODRL. However, there is no implementation of the Information Model. This was pointed out as a first gap and will represent a primary use of our ontology. Another important aspect retrieved by the SLR was the lack of capabilities to foster legal interoperability among actors

⁴ <<https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-g/UsageControl/Contract>>

under different jurisdictions. The EIF introduces the concept of *legal check*, which consists of the screening of interoperability barriers, such as different data models, geographical or social restrictions, etc. The implementation of such concepts has guided the development process of *LegION-IDS*. Figure 14 illustrates two sets of uses, *primary uses* and *secondary uses*. The primary uses were the starting point of the development process, as previously listed. From those two main intended uses, the secondary uses are logical consequences, such as contract automation, compliance cost reduction, common understanding of legal moments, and semi-automatic policy negotiation.

Figure 14 – Intended Uses Historical Dependency



Source: “Original Authorship (2025)”

5.1.5 Ontology Requirements

The elicitation of requirements, similar to Requirements Engineering (RE), has as its sole goal, the clear and concise modeling and elicitation of the stakeholders’ needs and wants (IMMONEN et al., 2016). The goal of this task is to acquire the set of requirements that the ontology should satisfy, and similarly to RE, we might as well divide the Requirements into Non-Functional Requirements (NFR) (LI et al., 2014), and Functional Requirements (FR) (IMMONEN et al., 2016).

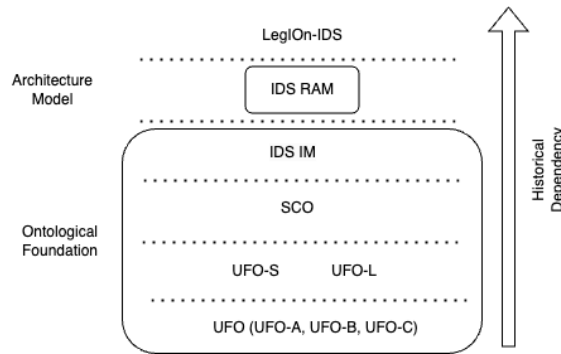
5.1.5.1 Non-Functional Requirements

Non-functional requirements (NFR) refer to reliability, maintenance, and performance (and many others) approaches (e.g.: how well does it perform, or how accurate is the inferred result), those requirements are not fundamentally necessary for the functionality of the software, e.g.: a software might run and perform its complied goals but at high latency, cost and hard to maintain (LI et al., 2014). Avoiding or misreading those requirements is often acknowledged to be one of the most expensive and time-consuming errors to fix, whenever the software has been implemented. Moreover, we may classify those NFRs into Performance, Design, and Adaptation (IMMONEN et al., 2016). Upholding the NFR Framework (LI et al., 2014), it elicits the two main approaches for characterization (i.e.: product-oriented

and process-oriented), but it also points out an alternative approach, which is the capability for developing the NFR *on demand*, therefore, refining them in an iteration-based evaluation, not as a final product. Based on that approach and aligned with our development methodology (SABiO), the NFRs might as well suffer alterations during the development/implementation steps. NFRs are a vast field of study in RE, hence, the concepts presented in this paper address superficial, yet sufficient aspects (for the presented problematization).

In our particular case, the ontology shall address the European Interoperability Framework and must support the English natural language. Furthermore, the complete documentation of the present ontology is available in an open-access GitHub repository ⁵, on which, principles of Findable, Accessible, Interoperable, and Reusable (FAIR), maintainability, and design were tackled. Furthermore, the ontology should follow the FAIR principles, especially focused on Interoperability, leading to the next NFR, the Legal Interoperability capability. An important NFR is the foundation of *LegION-IDS*. It must respect the UFO (and sub-ontologies) foundation, but should also comprise concepts available in the Information Model, and Reference Architecture Model. Figure 15 showcases the historical foundation used as a design requirement for the development of *LegION-IDS*.

Figure 15 – Design Foundation of *LegION-IDS*



Source: “Original Authorship (2025)”

The remaining are about maintainability and verifiability, with the purpose of better understanding and evaluation of the latter two, we may rely on the concept of Gradable NFRs (LI et al., 2014). A summary of the NFRs is available in Table 5.

⁵ <https://github.com/VictorBenoiston/legal_aspects_IDS_ontology>

Table 5 – *LegION-IDS* Non-Functional Requirements

Design	Support for Natural Language (English)
Adaptation	Addresses the EIF
Design/ Performance	Follow the FAIR Principles: F indable: It must have an open access. A ccessible: It shall possess unique URIs. I nteroperable: Use a formal and broadly applicable set of concepts and languages for representation. R eusable: Data meet domain-relevant accepted standards (ISO) and holds a clear and accessible usage license (MIT).
Design	UFO and SCO as foundational ontologies.
Design	Available at an open GitHub repository.

Source: “Original Authorship (2025)”

5.1.5.2 Functional Requirements

Functional Requirements (FR) are the clear specification of what the software is supposed to do, in our case, the knowledge presented in the ontology, including assertions, assumptions, and inferences (IMMONEN et al., 2016). For a clear understanding, and relying on (SUÁREZ-FIGUEROA; GÓMEZ-PÉREZ; VILLAZÓN-TERRAZAS, 2009) and (FALBO, 2014), we draw the FRs into Competency Questions (CQ), which are further partitioned into Informal Competency Questions, i.e. queries in the form of questions expressed in natural language, and Formal Competency Questions, i.e. queries in the formal language of ontology (in our case, SPARQL⁶, which is able to perform consults and retrieve answers in RDF).

CQs are questions that the proposed ontology must be capable of answering. The CQs should be thoroughly developed by clear communication among the stakeholders and ontology engineers. Another reliable way of extracting and founding the CQs is by retrieving key concepts from literature, e.g., research agendas, reference architectures, reference models, or similar works. As we earlier proposed, the current CQs of this ongoing work have been retrieved from our SLR. All CQs were acquired based on open future works in the analyzed literature (derived from their respective stakeholders). In this paper, we retrieved and verified our CQs from a few distinct sources, present in our SLR, which were classified as the most relevant works. To provide the CQs, we adopted the *middle-out* approach. Our final outcome set contains five questions that can encompass the dynamic aspect of legal interoperability: how agreements are formed through negotiation. After establishing all the static components (ownership, entitlements, rules, and purposes), those questions will further lead to queries that show how these elements come together in a process that results in contractual agreements. It provides a capstone that demonstrates how the

⁶ <<https://www.w3.org/TR/rdf-sparql-query/>>

ontology models not just static legal states but also the processes that create them and are displayed in Table 6.

Table 6 – Competency Questions (SCO and EIF related)

Competency Questions: SCO and EIF Related
CQ1: Who has primary control over specific <u>data assets</u> ?
CQ2: What is one legally <u>entitled</u> to do with this data?
CQ3: What specific <u>conditions</u> limit how data can be used?
CQ4: Why is this data being <u>processed</u> and is that <u>processing authorized</u> ?
CQ5: How do <u>parties</u> come to <u>legally binding arrangements</u> about <u>data sharing</u> ?
Source: “Original Authorship (2025)”

5.2 Reference Ontology

As the first artifact proposed by SABiO (FALBO, 2014), the Reference Ontology should encompass the visual representation of the types referred to in our ontology. After identifying the purpose, elicit the requirements, and capture and formalize the reference ontology (1st and 2nd steps). The support process of knowledge acquisition ought to be employed as well. For reference ontology stage of *LegION-IDS*, we follow the design assumption proposed in the NFRs and ground the ontology with the presented foundational ontologies (available in Figure 15). Furthermore, the reference ontology has been developed using the OntoUML language, which is provided through a plugin for the modeling tool Visual Paradigm⁷. Using such a tool aligns with the first and second steps of our grounding methodology, which allows the informal axiom definition (classes, referred to on UFO as *types*) (e.g., kinds, subkinds, categories, mixins) - and formal axioms (e.g., characterization, mediation) were modeled using UFO stereotypes. The reference ontology has been elaborated upon concepts and terms retrieved on the SLR, and post-performed verification of ISO standards for terms (general and specific legal ISO), the complete ISO vocabulary of our ontology is also available on the GitHub Repository). The reference ontology of *LegION-IDS* is composed of five views (granular topologies), and are thoroughly present in the next sections.

5.2.1 Main View

The main view encapsulates the main concepts tracing a parallel between the IDS core participants, and a service contract. For this view, we ought to consider the assumption that *in most cases, Data Owners acts as Data Provider and Data Consumer acts as Data User*, nonetheless, this view is henceforth documented. **IDS**

⁷ <<https://www.visual-paradigm.com/>>

Core Participant, as a *kind*, holds the identity of the fundamental participants every time a data exchange takes place. The IDS Core Participant may play three different *roles* (based on our assumption), i.e., *Data Owner*, *Data Consumer*, and *App Provider*. In ontoUML, generalization means correspondence, e.g., every *Data Owner* is an *IDS Core Participant*, but not every *Data Owner* is an *IDS Core Participant*. In first-order logic, we might represent as:

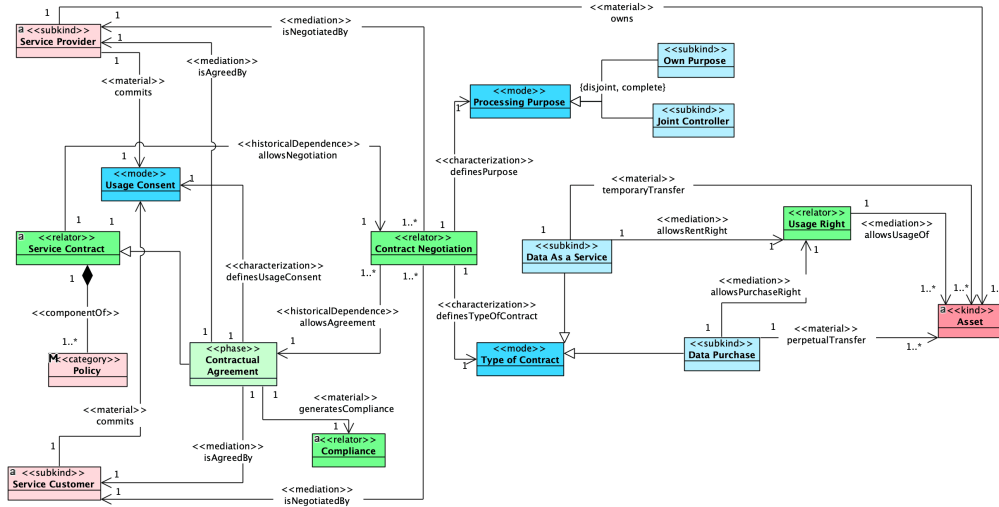
$$\begin{aligned} & \forall i (\text{DataConsumer}(i) \rightarrow \text{IDSCoreParticipant}(i)) \\ & \wedge \exists i (\text{IDSCoreParticipant}(i) \wedge \neg \text{DataConsumer}(i)) \end{aligned}$$

Furthermore, a *Data Owner* acts as a *Data Provider*, and *Data Consumer* as *Data User*. Each *Data User* is equivalent to a *Service Customer*, whereas *Data Provider* is equivalent to *Service Provider*. A **Contractual Party** holds the stereotype of *kind*, once it holds the identity of actors involved in a *Service Contract* (GRIFFO et al., 2021b). In first-order logic:

$$\begin{aligned} & \forall i ((\text{ServiceConsumer}(i) \vee \text{ServiceProvider}(i)) \rightarrow \text{ContractualParty}(i)) \\ & \wedge \exists i (\text{ContractualParty}(i) \wedge \neg (\text{ServiceConsumer}(i) \vee \text{ServiceProvider}(i))) \end{aligned}$$

Moreover, each *Service Provider* and *Service Customer* has its own *Governing Law* (DUISBERG, 2022b), and it is composed by a *Data Protection Law*, and a *Competition Law*. When two companies under different jurisdictions (own *Governing Law*) shall collaborate, the EIF proposes the comparison among them, allowing its *compliance*. The *Compliance* on its own, will be guaranteed after policies negotiation. The process has two possible beginnings, by a *Contract Request* performed by the *Service Customer*, or a *Contract Offer*, performed by the *Service Provider*, regardless the beginning, the policies must be represented in the *Service Contract*. The Main View is available in Figure 16.

Figure 17 – Reference Ontology (Service Contract View)



Source: “Original Authorship (2025)”

5.2.3 Policy View

As mentioned, *LegION-IDS* is grounded in a few perspectives, among which, is the Open Digital Rights Language (ODRL)⁸. This view emphasizes the concept of *Policy*, which is composed of a non-empty set of *Rules*, fostering the unambiguous representation of policies in a contract, leveraging its (semi)automatic negotiation. A *Rule* is an abstract concept that characterizes the permission, duty, or prohibition by some party to perform an *Action* over some *Asset*. It must be abstract, and hence, not able to characterize (defined). The ODRL proposes the classification of *Rule* into *Permission Rule*, which allows some *Action*, *Duty Rule*, which obligates some *Action*, and *Prohibition Rule*, which denies some *Action*. The Permission may also have the duty property that expresses an agreed *Action* that must be exercised (as a pre-condition to be granted the Permission) (DUISBERG, 2022b). In first-order logic:

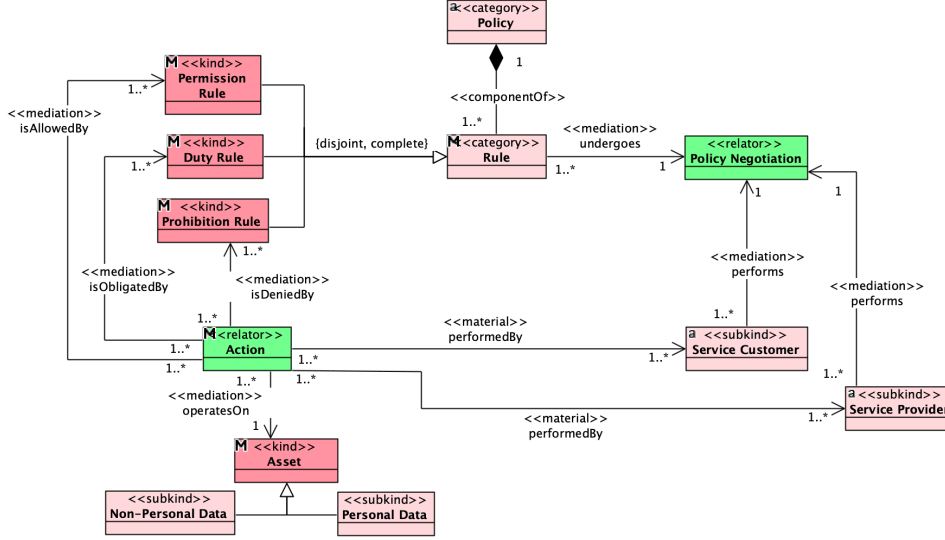
$$\begin{aligned}
 & \forall i ((\text{Permission Rule}(i) \vee \text{Duty Rule}(i) \vee \text{Prohibition Rule}(i)) \rightarrow \text{Rule}(i)) \\
 & \wedge \exists i (\text{Rule}(i) \wedge \neg(\text{Duty Rule}(i) \vee \text{Prohibition Rule}(i) \vee \text{Permission Rule}(i))) \\
 & \wedge \exists i (\text{Duty Rule}(i) \vee \text{Prohibition Rule}(i) \vee \text{Permission Rule}(i))
 \end{aligned}$$

Similar to a *Contract Negotiation*, a *Policy Negotiation* is the capability of dynamization of specific policies, and as a relator, mediates (performed by) the *Service Provider* and *Service Customer*. Moreover, an *Action* represents an operation on an *Asset*, which is a resource or a collection of resources that are the subject of a *Rule*, and may be classified as *Personal* and *Non-Personal Data*. A *Rule* is seldom characterized by a

⁸ <<https://www.w3.org/TR/odrl-model/>>

type of *constraint*, i.e., a boolean/logical expression that refines an *Action*, and *Asset* collection or the conditions applicable to a *Rule*. This view is available in Figure 18

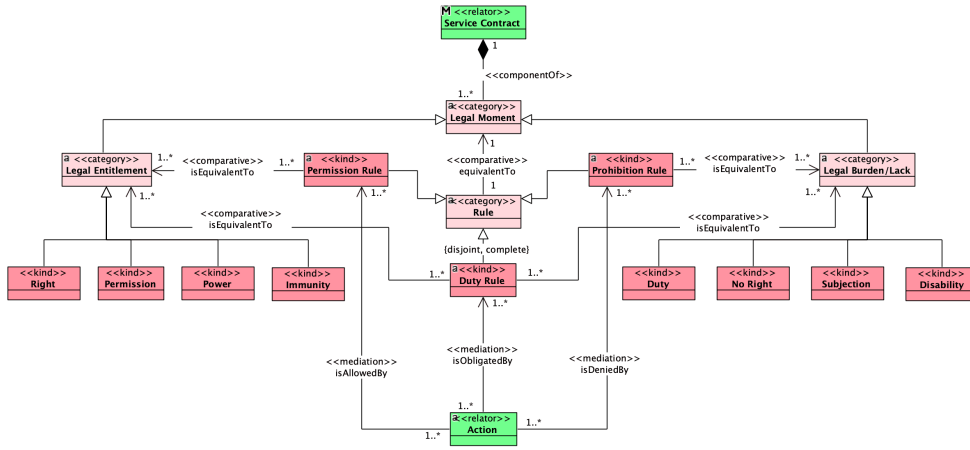
Figure 18 – Reference Ontology (Policy View)



Source: “Original Authorship (2025)”

5.2.4 Legal Moments View

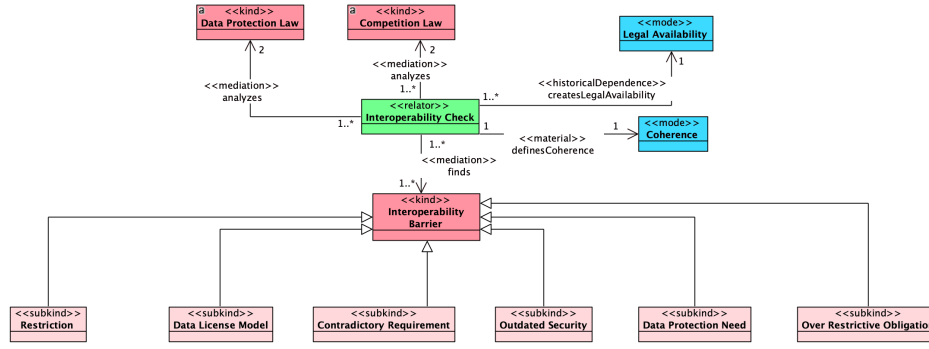
This view entails the concepts available in the Service Contract Ontology (SCO) (GRIFFO et al., 2021b). A *Service Contract* is composed of a set of *Legal Moments*, which recovers the concept of *moment* defined by UFO and employs it to the legal positions described by (ALEXY, 2002). As *category*, it holds a general concept, with different identity principles, which may be a *Legal Entitlement*, which implies positivity, or *Legal Lack/Burden*, which implies negativity (GRIFFO et al., 2021b). The *Legal Entitlements* may be derived into *Power*, *Permission*, *Right*, and *Immunity*, as *kinds*. Whereas the *Legal Burdens/Lacks* divides into *Duty*, *NoRight*, *Subjection*, *Disability*. A *Legal Moment* is compared to a *Rule*, which as seen before, may split into prohibition, duty, and permission. By tracing the compatibilities of SCO and ODRL, we might compare the *Legal Entitlements* to *Permission Rules* (once both rely on a positive endeavor), *Legal Burdens/Lacks* may be compared to *Prohibition Rules* (by carrying a negative payload), and finally, the *Duty Rule* may be a placeholder for both *Legal Moments*. Those equivalences will be further explored in the operational ontology, allowing several inferences based on their ability to describe a *service contract*. This view is represented in Figure 19

Figure 19 – Reference Ontology (Legal Moments View)

Source: “Original Authorship (2025)”

5.2.5 Interoperability View

The last view of *LegION-IDS* (reference ontology) consists of the *Interoperability Checks* proposed by the EIF. The EIF defines Legal Interoperability as *the capability of organizations operating under different legal frameworks, they manage to work together*. This is accomplished by aligning policies and strategies, requiring that current legislation does not block the proposed policies, generating clear agreements on how to deal with those differences across borders, and even allowing the inclusion of a new legislation. Our proposed SLR, however, defines Legal Interoperability as *the semantic capability of unambiguously representing policies and rules among companies while respecting the so-called state sovereignty*. This view treats Legal Interoperability as a consequence, achieved when aligning the proposed checks and alignments compliance. The *Interoperability Check* is a *relator* (event), which mediates (analyzes) the *Data Protection Law*, and *Data Competition Law* of the parties involved on the *Service Contract*. Furthermore, the *Interoperability Check* may find *Interoperability Barriers*, which are classified by the EIF, as *Restriction*, *Data License Model*, *Contradictory Requirement*, *Outdated Security*, *Data Protection Need*, and *Over Restrictive Obligation*. The application of *Interoperability Checks* leads to the definition of *Coherence*, which is the alignment of the analyzed governing laws, and further creates the so-called *Legal Availability*. *Legal Availability* refers to the dynamization potential of aligning the analyzed governing laws, based on their interoperability constraints. Figure 20 describes the proposed view.

Figure 20 – Reference Ontology (Legal Moments View)

Source: “Original Authorship (2025)”

5.3 Operational Ontology

The operational ontology should take an in-depth vision of the proposed domain, adding details to the ontology, such as disjointness, closing axioms, relationships, object properties, and data properties. Once we are grounded by UFO, the class hierarchy ought to provide the stereotypes as the foundation. De Souza et al. (2014) proposes that once we have a reference model ontology, we might translate it into an operational version, which could be used by computer applications (FALBO, 2014). To achieve such an operational ontology, we need to design and implement it in a particular machine-readable ontology language, such as OWL. As stated before, we use Protégé desktop to develop our operational ontology. While the reference ontology focuses on representation adequacy, the operational ontology focuses on guaranteeing desirable computational properties.

A special nuance of ontologies (yet as a philosophical domain) is the ability to empower *deduction*. Deduction is *a way to ascertain if a theory represented in a logic language entails an axiom that is not explicitly asserted in the theory* (KEET, 2018). Therefore, whether the axioms can be derived from the premises through *repeated* application of *deduction rules* - which for computer science, means axiomatic rules. An *axiom* is a problem’s assumption in the proposed logical theory (intrinsic to the ontology), which will lead to the *problem’s conclusion* that is computed by the automated *reasoner* as either *yes* or *no*.

Reasoning is performed on a daily basis by human agents, simply by taking a set of premises, and somehow, forming a conclusion. The agent is entailed by the premises (deductive process), which either leads to a hypothesis (abduction), or generalization to an assumption (induction) (KEET, 2018). Logic languages appeared to simulate this human behavior, in order to represent the knowledge and rules that may be applied to *axioms*, leading to conclusions. Formal languages

(such as First-Order Logic (CRUZ, 2021)), can be manually performed for small theories, however, to describe multi-connected theories (such as ontologies), while not contradicting itself, is rather unpleasant, therefore, we may implement automated reasoning. Automated reasoning is based on a descriptive logic knowledge base, and the concepts of deduction, abduction, and induction, hence, may be applied in a wide range of scenarios, for specific domains (DAVIS; MARCUS, 2016). Automated reasoning is an enormous field of study in computer science, for this document, we are going to attain the essentials of it. Furthermore, we might encapsulate the concept of automated reasoning as:

- The choice of the class of problems the software program has to solve (e.g., checking the consistency of a theory).
- The formal language in which to represent the problems (e.g., cardinality constraints, closing axioms).
- The way how the program has to compute the solution (using natural deduction or resolution).
- How to do this efficiently (either constraining the language into one of low complexity or optimizing the algorithms)

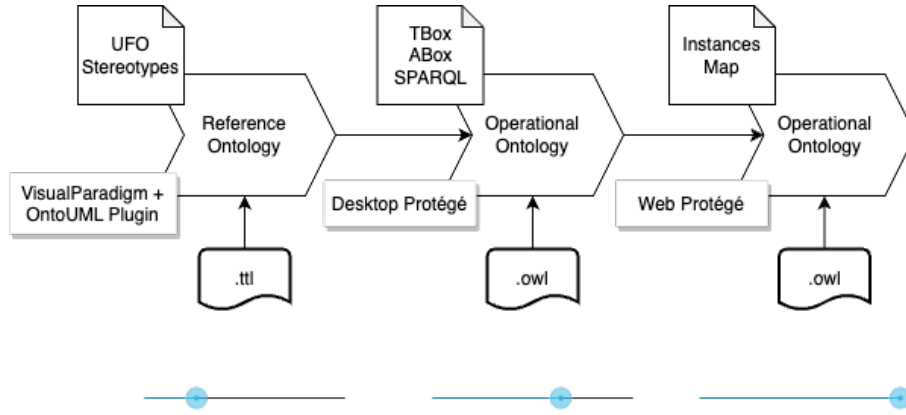
As a fragment of FOL, Description Logic (DL) divides a *knowledge base* in the *TBox*, which comprises the **T**erminology (class-level/types) and *ABox*, which defines the **A**ssertions (individuals/instances). Some literature may point to an *RBox*, which defines the automated **R**easoning, comprising the Tbox and Abox, nevertheless, for the development of *LegION-IDS*, we only provide in-depth details of the *TBox* and *ABox*, and shortly present the *RBox*.

5.3.1 Environment

For the development of *LegION-IDS* (operational), we used the automated reasoner Pellet (SIRIN et al., 2007), which supports OWL-DL and extends the functionalities of qualified cardinality restrictions, complex sub-property axioms (between a property and a property chain), local reflexivity restrictions, reflexive, irreflexive, symmetric and anti-symmetric properties, disjoint properties. Furthermore, we used Protégé desktop⁹ for the development of the *ABox* and *TBox*. For the querying, we employed the SPARQL language in order to retrieve data from the RDF. Finally, to visualize the instance maps, we exported the OWL file to Web Protégé¹⁰. The development lifecycle is represented in Figure 21

⁹ <<https://protege.stanford.edu/>>

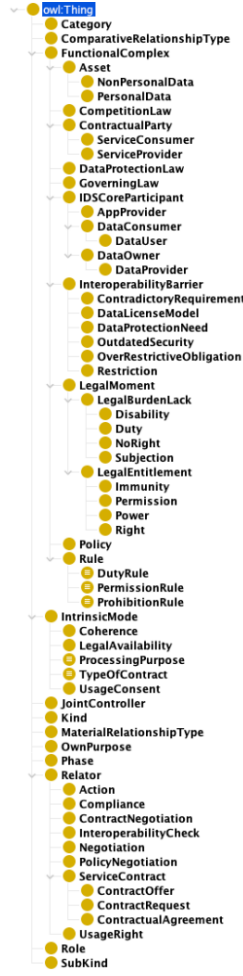
¹⁰ <<https://webprotege.stanford.edu/>>

Figure 21 – *LegION-IDS* development lifecycle and environment

Source: “Original Authorship (2025)”

5.3.2 Class Hierarchy

The Class Hierarchy comprises all the high-level concepts, which define the Taxonomy of our ontology, the *TBox* (KEET, 2018). As aforementioned, to develop *LegION-IDS*, we employed the *top-down* methodology, grounding the ontology into a foundational ontology (UFO) (GUIZZARDI, 2005), and their formalizations. The axioms hereby introduced describe the relationships between concepts. For the class hierarchy, the first layer of classes is composed of the UFO stereotypes, e.g., *category*, *kind*, *phase*. Within each stereotype, we approach the high-concepts of *LegION-IDS*, such as *asset*, which is the representation of interest of the *contractual parties*, in our case, data, and it may be *personal* or *non-personal*. Furthermore, we divide the *contractual parties* into *Service Customer*, and *service provider*, similarly to the reference ontology. the complete OWL file is available in the previously mentioned GitHub repository, and the classes hierarchy is available in Figure 22

Figure 22 – *LegION-IDS* Classes Hierarchy (Asserted)

Source: “Original Authorship (2025)”

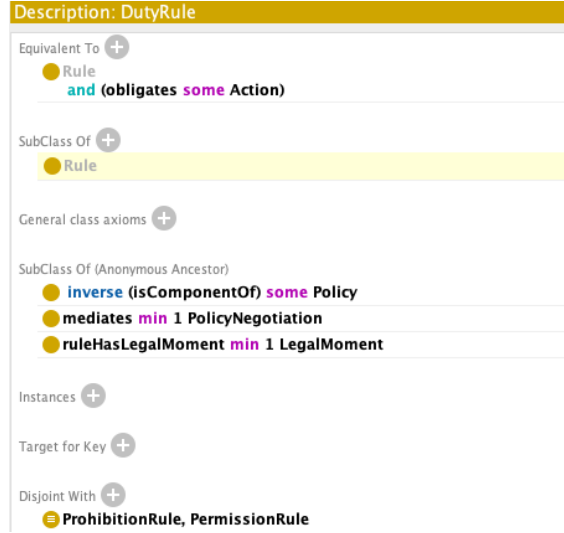
The *asserted* view of the class hierarchy presents the classes as we designed, hence, no reasoning whatsoever. For the design process of *LegION-IDS*, and aligned with the RQs, our ontology does not address reasoning on the *TBox*, but rather, only at the instance level, the *ABox*. (KEIL, 2018) proposes an approach to treat *static TBox*, which comprises a scenario on which, the *TBox* is considered invariant. Through *LegION-IDS*, we propose the representation of legal aspects regarding the Reference Architecture Model, which does not imply inferencing or reasoning among its classes. Rather, we propose an ontology that produces knowledge based on its instances. Furthermore, (KEIL, 2018) states the *instance-level update*, and *instance-level erasure*, the former, represents the addition, and the latter, the exclusion of instances. The set of models (axioms) that compose our Description Logic *knowledge base*, is solely dependent on the *ABox* (instance-level). The main differences between reference ontology and operational ontology, are presented through nuances such as *disjointness*, which is the definition of boundaries. RDF is composed of *sets*, hence, two siblings (same level) are *disjoint* classes, when an individual (instance) cannot

belong to the same classes at the same time. For instance, if we propose that the class *Rule* has its sub-classes disjoint, we cannot have an instance belonging to *DutyRule* and *PermissionRule* (or *ProhibitionRule*) at the same time. In FOL, we have:

$$\forall i \text{ Rule}(i) \rightarrow [\neg \text{ProhibitionRule}(i) \vee \neg \text{DutyRule}(i) \vee \neg \text{PermissionRule}(i)]$$

In Protégé, this axiom is represented as such in Figure 23.

Figure 23 – Example of *disjointness* in class *DutyRule*



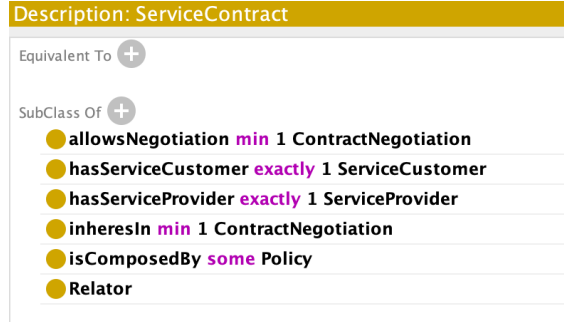
Source: “Original Authorship (2025)”

5.3.3 Axioms

As one of the main differences between a standard database and an ontology, the latter does not fully describe a *particular* state of the world, rather, it describes it by a set of statements, which are called *axioms*, and each of the proposed *axioms* must be true in its described situations. (GUIZZARDI, 2005) describe those axioms as *truth-makers*, serving as boundaries for the reasoning, and the unambiguous description of the pointed domain. Each *axiom* captures a specific statement (partial knowledge) about the situation described in the ontology, once there may be several different states of the world within the ontology (BORGIO et al., 2022). The *ABox* axioms describe assumptions about named individuals (instances), i.e., the high-level concept it belongs to, and the relationship with other individuals. Applying those premises to UFO, we may derive the concepts of *Role*, which as a high-level concept, may be incorporated by different individuals, and define further relations based on that. Terminological knowledge is expressed through the *TBox*, as a consequence of the relationships between concepts. For instance, in *LegION-IDS* one of the main concepts proposed, is the comparison between the high-level concept *Policy*, which was presented by (BADER et al., 2020), through the Information Model, and the

high-level concept of *Rule*, which was retrieved by the Open Digital Rights Language (Figure 18). The main concepts of *LegION-IDS* are endorsed by the classes *Service Contract*, which is axiomatically presented in Figure 24

Figure 24 – Axiomatic representation of the class *Service Contract*



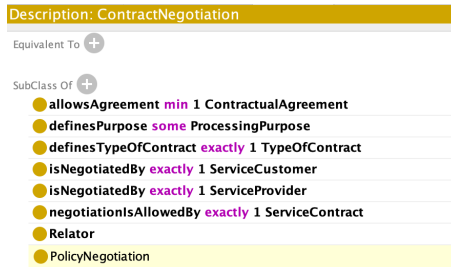
Source: “Original Authorship (2025)”

In this class, we may observe triples such as, *Service Contract* is *ComposedOf* some *Policy* (subject, predicate, object). All *triples* are *axioms*, but not the other way around. In FOL:

$$\forall t \text{ Triple}(t) \rightarrow \text{Axiom}(t) \wedge \neg(\exists a \text{ Axiom}(a) \rightarrow \text{Triple}(a))$$

Furthermore, we have *Service Contract* *hasServiceConsumer* some *Service Customer*. The reserved word *some*, expresses the cardinality of *min 1*. The same thing happens to a *ServiceProvider*, and a *Service Contract* allows the *Contract Negotiation*, which is a *deductive* thought, once a negotiation stands after an *offer* or *request*, being existentially dependent. Furthermore, we showcase the axiomatic vision of the class *Contract Negotiation* in Figure 25.

Figure 25 – Axiomatic representation of the class *Contract Negotiation*

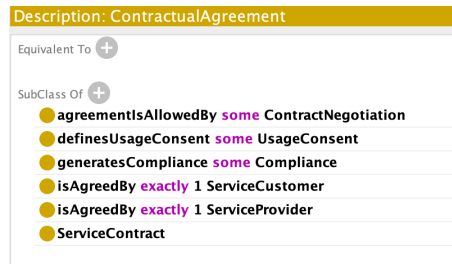


Source: “Original Authorship (2025)”

This set of *triples* defines the *Contract Negotiation* proposing its instauration by a *Service Customer* and *Service Provider*, which is another *inductive* thought, once it should be performed by its *Contractual Parties*. Further, it is only allowed once there is a *Service Contract*, as previously stated, an existential dependency. Finally,

the *Contract Negotiation* allows exactly 1 *Contractual Agreement*. The cardinality *exactly*, defines the exact number of *objects*, in this case, one agreement must take place. Finally, the triples *Contract Negotiation definesTypeOfContract* some *Type of Contract*, and *Contract Negotiation definesPurpose* some *Processing Purpose* regards the type of contract (which may be data purchase or data as a service), and the processing purpose, which varies as own purpose, or joint controller. Moreover, Figure 26 describes the *Contractual Agreement*.

Figure 26 – Axiomatic representation of the class *Contractual Agreement*



Source: “Original Authorship (2025)”

As outlined, the *Service Agreement* is existentially dependent on a *Service Contract* and also involves a *Service Provider* and a *Service Customer*. The agreement relies on the *UsageConsent*, provided by the commitment of the *Service Customer* and *Service Provider*. Finally, as outlined in Figure 26, all *Contractual Agreement* historically depends on a *Policy Negotiation*. If we take a look at the *Service Contract View*, in our reference ontology (Figure 17), we may observe that all *Service Contracts* are composed of a non-empty set of policies. Therefore, the negotiation of the so-called *Service Contract*, relies on the negotiation of the specific *Policies*.

Additionally, Protégé also allows the creation of *closing axioms*, which consists of a universal restriction (*constraint*) that acts along the property that only specified individuals may fulfill it (BORGIO et al., 2022). For instance, *LegION-IDS* introduces the concept of a *Duty Rule*. A *Duty Rule* is classified by a *Rule* that delegates an *obligation*, e.g., *entity A must pay taxes*. OWL is grounded by the Open World Assumption, which, unless clearly stated, there are possibilities for breach (BORGIO et al., 2022). To mitigate this assumption, we might add the *closure axioms*, which are *necessary and sufficient* conditions. In the axiom ‘duty rules delegates an obligation’, the subjectivity may lead the reasoner to interpret another rule, that has an obligation and a prohibition, to be characterized as such. To avoid this misunderstanding, the reasoner does not classify it, leaving it as a *necessary but not sufficient* condition, i.e., to be a *Duty Rule*, it must *obligates* an action. To provide the reasoner with objectivity, we may present this axiom as ‘a duty rule only obligates an action, or in

FOL:

$$\forall d \text{DutyRule}(d) \rightarrow \forall a (\text{obligates}(d, a) \rightarrow \text{Action}(a))$$

In Protégé, if we change to the *PermissionRule*, the axiom is represented in Figure 27

Figure 27 – Closure Axiom of class *PermissionRule*

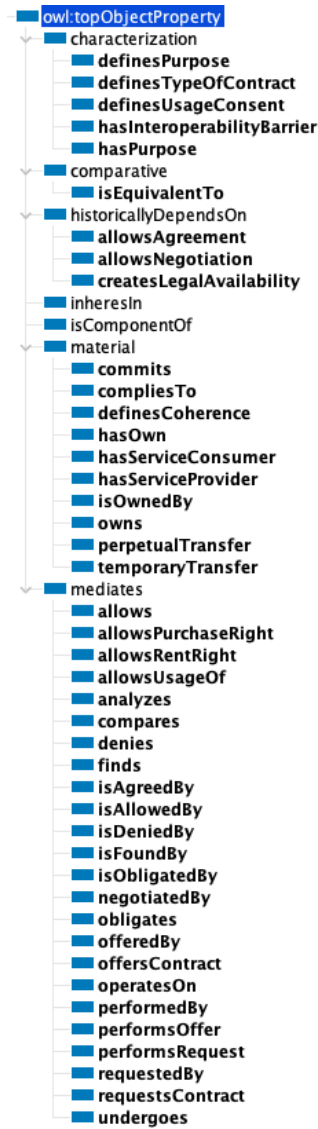


Source: “Original Authorship (2025)”

Finally, we present the *Enumerated Classes*. As in most computational languages, an enumerated class should only be addressed by pre-defined values, in our case, instances (KEET, 2018). An example of *Enumerated Class* is the class *Type of Contract*, which only accepts the instances of *Data as a service* or *Data purchase*.

5.3.4 Object Properties

In Description Logics, the so-called *Roles* are translated to the OWL’s *Object Properties*. An object property is a *predicate*, which links *classes* to other *classes*, hence, is a high-level connector (KEET, 2018). It characterizes, defines, or updates the *object*, by adding properties to it. In *LegION-IDS*, we may exemplify the object properties *hasLegalEntitlement*, and *hasLegalBurden*, which characterizes a *Legal Moment*. Object properties have a domain, which is the *subject*, and a range, which is the *object*. It links a subject A to an object C, by a property B. An object property may specialize into functional, inverse functional, transitive, symmetric, etc. For instance, a *funcitonal* object property, may only have a single range, e.g., a *Service Contract* may only have one *Type of Contract*. The *LegION-IDS*’ object property hierarchy is showcased in Figure 28

Figure 28 – *LegION-IDS'* object properties hierarchy

Source: “Original Authorship (2025)”

5.3.5 Data Properties

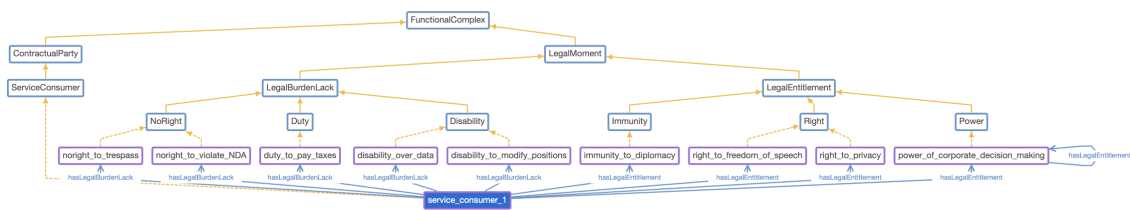
On the other hand, a *data property* connects a subject to a value. A data property denotes a binary relation between objects and values (KEET, 2018). For instance, all *Service Contracts* entails a *creation date*, hence, instead of creating a class, we might encapsulate the value into a data type, and assume certain values. The domain is going to be the **Service Contract** (what we need to characterize), to a data type, e.g., xsd:string. *LegION-IDS'* data properties are presented in Figure 29

Figure 29 – *LegION-IDS*’ data properties hierarchy

Source: “Original Authorship (2025)”

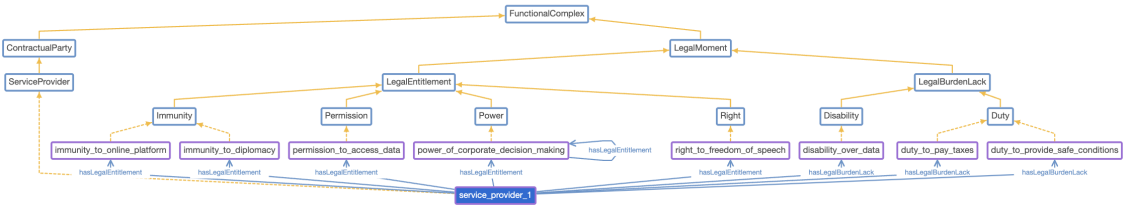
5.3.6 Instances Map

Whereas *types* are abstract things, rather, tools to classify the world, i.e., a set of characteristics we can expect to encounter in other particular things, and those particular things, are the *individuals* (GUIZZARDI, 2005). The individuals, or instances, are the set of specific characteristics related to a *type*. Inductive thinking allows one to extract the essence of the type, and classify instances of that kind, even though each individual is composed of a set of unique (or non-unique) characteristics (or features). Using the web version of Protégé, we might create *instances* map, convenient of the instantiated individuals using Desktop Protégé. Please note that the instances henceforth are *generic*, either retrieved by our SLR or open-source service contracts and governing laws. A road map towards this issue is proposed in Section 7. Figure 30 and Figure 31 provide the instances map of the main classes in *LegION-IDS*. Several snapshots of the instances map and the complete list of instances are stored in the aforementioned GitHub repository. Those two views allow the exhibition of the *Legal Moments* regarding the *Service Provider* and *Service Customer*.

Figure 30 – Instances Map (Service Customer)

Source: “Original Authorship (2025)”

Figure 31 – Instances Map (Service Provider)



Source: “Original Authorship (2025)”

6 Ontology Verification and Validation

This chapter provides an explanation of the proposed results based on our chosen methodology. It details the process of choosing the use cases from the literature provided by IDSA, showcasing how we partially instantiated the ontology and answering the proposed CQs, along with generating inferred results based on the reasoning, both verifying its completeness and validating the ontology.

6.1 Ontology Validation through Manufacturing Use Cases

To properly complete the Design Science Research lifecycle, validating our ontology is essential. For this purpose, we employ a methodical, though intentionally partial, instantiation approach using three representative use cases from IDSA reference literature. These selected use cases provide rich contextual frameworks that enable targeted testing of our ontology’s expressive capabilities and conceptual coverage through the formalized Competency Questions (CQs).

It is important to emphasize that this constitutes a partial validation strategy, deliberately focused on a subset of the ontology’s classes and relationships. While this approach is sufficient to fulfill the formal requirements of the DSR methodology’s evaluation phase, it does not represent a comprehensive validation of all ontological constructs. The scope was necessarily constrained to address the most critical aspects of legal interoperability within feasible research parameters.

The validation scenarios were extracted from the IDSA position paper “Data Sovereignty – Requirements Analysis of Manufacturing Use Cases,” (LARRINAGA, 2022), which represents a consensus view on data sovereignty requirements across the manufacturing industry. This document was developed through collaborative efforts of the IDS-Industrial Community, incorporating insights from multiple European research projects (OPEN DEI, QU4LITY, AI REGIO) and contributions from prominent research institutions, including Fraunhofer, TNO, and leading industrial partners.

Each selected use case—Collaborative Condition Monitoring (CCM), Smart Factory Web (SFW), and Sustainability Development Goals (SDG) impact assessment—presents distinct data sovereignty challenges that collectively exercise significant portions of our ontology’s capabilities. These scenarios contain sufficient granularity to instantiate key elements of the ontology’s core conceptual framework, encompassing:

- **IDS Core Participants** – data providers, consumers, app providers, and service brokers
- **Legal Entities** – contractual parties, service providers, and service customers
- **Normative Structures** – permissions, prohibitions, obligations, and policy rules
- **Data Sovereignty Mechanisms** – access control, usage policies, and provenance tracking
- **Legal Acts** – contract negotiations, agreement establishments, and consent declarations
- **Asset Management** – data ownership, rights assignments, and processing purposes

While our instantiation does not cover all classes defined in the ontology, it strategically addresses those most central to legal interoperability requirements in data spaces. The validation process involved creating individual instances for these selected classes, establishing appropriate property relationships between them, and verifying the ontology’s ability to represent key legal interoperability scenarios through SPARQL queries.

This partial validation approach serves as a pragmatic compromise that satisfies the DSR lifecycle requirements while acknowledging the practical limitations of research scope and identifies areas where more comprehensive validation would be beneficial in future work research phases.

By grounding our validation in concrete manufacturing use cases recognized by industry stakeholders, we establish a meaningful empirical foundation for assessing the ontology’s core utility in addressing real-world legal interoperability challenges in data spaces, even as we recognize the need for more extensive validation in subsequent research.

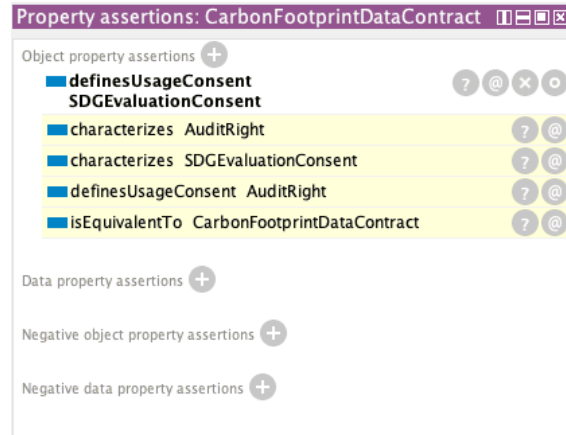
6.2 Ontology Completeness

As previously established in 3.4, we are able to partially showcase the ontology completeness by using the reasoner and generating inferred classes with not syntax issues. This is the first step to properly validate that our ontology axioms are well established.

The results of this partial instantiation have shown to be positively surprising. Displaying not only the capability to answer the proposed CQs, but to also display

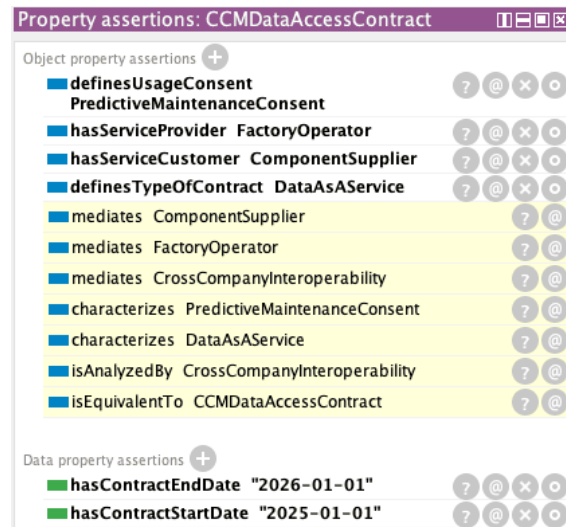
its reasoning capabilities. The core concepts from our ontology focus on the ability to describe the data usage contracts and their agreement. As the results have shown, we are able to infer the characterization generability of a contract, its equivalence, and its usage consent definition as seen in Figure 32 and Figure 33.

Figure 32 – CFD Contract



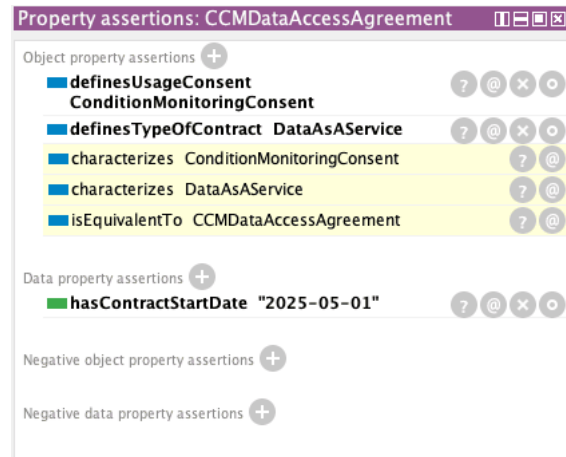
Source: “Original Authorship (2025)”

Figure 33 – CCM Data Access Contract

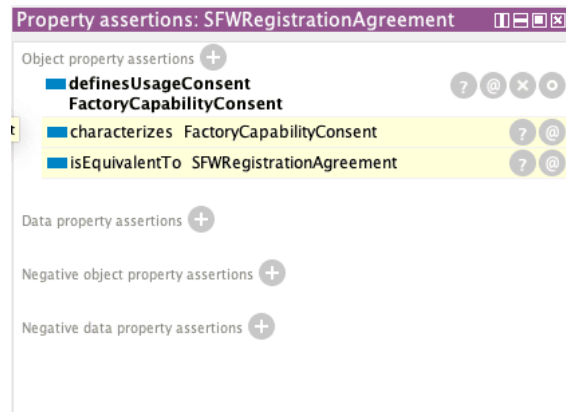


Source: “Original Authorship (2025)”

Furthermore, the ontology was able to infer the mediation capabilities, characterization, and equivalence of it, as seen in Figure 34 and Figure 35

Figure 34 – CCM Data Access Agreement

Source: “Original Authorship (2025)”

Figure 35 – SFWR Registration Agreement

Source: “Original Authorship (2025)”

6.3 Competency Questions

As proposed by our chosen methodology, a robust validation process is essential to ensure the ontology effectively models the domain’s complexity while maintaining logical consistency. To achieve this, we implement a comprehensive twofold validation approach. As we propose a framework to uphold the domain of legal interoperability, the competency questions went through a thorough process of refinement to properly express a junction from the literature enabling the core concepts of UFO-L (GRIFFO; GUIZZARDI; ALMEIDA, 2018) and the Information Model (OTTO; HOMPEL; WROBEL, 2019).

Firstly, in the previous section, we partially verified the ontology’s completeness through a systematic competency question evaluation using a preliminary set of

mock instances. This initial use demonstration confirmed the ontology’s structural integrity and its ability to represent basic legal interoperability concepts.

We now significantly extend this validation by applying the ontology to real-world manufacturing scenarios drawn from authoritative industry sources. Specifically, we selected three reference use cases from the International Data Spaces Association’s (IDSA) position paper "Data Sovereignty – Requirements Analysis of Manufacturing Use Cases" (LARRINAGA, 2022), which represents consensus from leading industrial partners across Europe, Japan, and other regions. These use cases—Collaborative Condition Monitoring (CCM), Smart Factory Web (SFW), and Sustainability Development Goals (SDG) impact assessment—were explicitly developed by the IDSA to capture the most pressing data sovereignty requirements in modern manufacturing environments.

This second phase of validation serves multiple critical purposes: (1) it tests the ontology against industry-recognized scenarios rather than theoretical constructs, (2) it verifies the ontology’s ability to represent the specific data sovereignty requirements identified as high-priority in industrial settings, and (3) it demonstrates the ontology’s practical utility for addressing real-world legal interoperability challenges in manufacturing data spaces.

For each use case, we instantiated the relevant ontology classes and properties, then formulated SPARQL queries addressing fundamental conceptual questions about legal interoperability. The successful execution of these queries, as demonstrated below, provides empirical evidence that our ontology effectively captures the legal relationships required for data sovereignty in complex manufacturing ecosystems.

The verification of an ontology can be assessed by its completeness, which should be ensured by answering the proposed competency questions presented on the ORSD. The Functional Requirements were previously translated into Competency Questions, as seen in Section 5.1.5. As an ongoing work, we partially answered the Competency Questions as follows in Table 7. The informal CQs are proposed in natural language, whereas the Formal CQs are presented in SPARQL queries. Concurrently to the SPARQL queries, we provide the outcome for each question, fostering one of the steps of validation, assessing the consistency and correctness of the ontology (MUNOZ-ARCENALES et al., 2019a).

Table 7 – Formal and Informal Questions and Equivalent Answers

Competency Questions
CQ1 (Informal): Who has primary control over specific <u>data assets</u> ?

CQ1 (Formal):

PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX legion: <http://www.legal_interoperability_ontology#>

SELECT ?owner ?asset
WHERE {
 ?owner legion:owns ?asset .
 OPTIONAL { ?owner legion:hasContractualPartyName ?ownerName }
}

Execute

?owner	?asset
legion:FactoryOperator	legion:MachineOperationalData
legion:FactoryOwner	legion:FactoryCapabilityDescription
legion:FraunhoferIOSB	legion:FactoryOperationalData
legion:SupplyChainParticipant	legion:EmissionData

Source: “Original Authorship (2025)”

Purpose: This first query translates the ownership of data, which is the foundation upon which all other rights and permissions are built in the legal interoperability aspect from EIF. This validation is particularly important for the IDS (International Data Spaces) architecture, which emphasizes data sovereignty as its core principle. This query ensures that we support and follow IDS’s core requirement that data owners maintain control over their data throughout its lifecycle.

Ontology Elements Validated:

- owns property: Validates the core ownership relationship
- Asset and DataOwner instances: Confirms existence of these fundamental instances
- hasContractualPartyName property: Tests identification of ownership entities

CQ2 (Informal): What is one legally entitled to do with this data?

CQ2 (Formal):

PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX legion: <http://www.legal_interoperability_ontology#>

SELECT ?entity ?entitlement ?relatedRight
WHERE {
 ?entity legion:hasLegalEntitlement ?entitlement .
 OPTIONAL { ?entitlement legion:allowsUsageOf ?relatedRight }
}

Execute

?entity	?entitlement	?relatedRight
legion:ThirdPartyCertificationBody	legion:AuditRight	legion:ResourceCirculationData
legion:ThirdPartyCertificationBody	legion:AuditRight	legion:EmissionData
legion:ThirdPartyCertificationBody	legion:SDGEvaluationConsent	legion:ResourceCirculationData
legion:ThirdPartyCertificationBody	legion:SDGEvaluationConsent	legion:EmissionData

Source: “Original Authorship (2025)”

Purpose: This query examines all legal entitlements assigned to entities and their optional connections to specific usage rights by connecting entities to their legal entitlements and, subsequently, to the rights these entitlements confer. This three-level relationship is essential for modeling complex legal scenarios in data spaces. Validating that our proposed ontology correctly implements the concept of rights derivation — showing how legal entitlements generate specific rights.

Ontology Elements Validated:

- hasLegalEntitlement property: Validates entitlement assignments
- LegalEntitlement instances: Confirms existence of proper entitlement instances
- allowsUsageOf property: Tests the connection between entitlements and usage rights

CQ3 (Informal): What specific conditions limit how data can be used?

CQ3 (Formal):

PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX legion: <http://www.legal_interoperability_ontology#>

SELECT ?rule ?constraint ?creationDate
WHERE {
 ?rule legion:hasRuleConstraint ?constraint .
 OPTIONAL { ?rule legion:hasRuleCreationDate ?creationDate }
}

Execute

?rule	?constraint	?creationDate
legion:DataIntegrityRule	Critical data must not be modified by untrusted nodes^^...	21-05-2025^^xsd:string
legion:DataSecrecyRule	Classified data must not be forwarded to unauthorized p...	21-05-2025^^xsd:string
legion:TimeToLiveRule	Data must be deleted after prescribed period^^xsd:string	21-05-2025^^xsd:string
legion:UsageScopeRule	Data may only serve as input within the connector but nev...	

Source: “Original Authorship (2025)”

Purpose: This query retrieves all rule instances along with their specified constraints and optional creation dates. Validating the rule framework of our proposed ontology, which forms the backbone of legal interoperability. By confirming that rules have specific constraints, you verify that the ontology can express concrete legal requirements rather than just abstract concepts. The presence of creation dates helps validate the temporal aspects of your legal model, ensuring the ontology can represent when rules came into effect — a critical aspect for legal validity.

Ontology Elements Validated:

- Rule class instances: Validates the existence of rule instances in the ontology
- hasRuleConstraint property: Confirms rules have properly defined constraints
- hasRuleCreationDate property: Verifies temporal metadata for rules

CQ4 (Informal): Why is this data being processed and is that processing authorized?

CQ4 (Formal):

Snap SPARQL Query:		
<pre> PREFIX owl: <http://www.w3.org/2002/07/owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> PREFIX legion: <http://www.legal_interoperability_ontology#> SELECT ?asset ?purpose ?relatedConsent WHERE { ?asset legion:hasProcessingPurpose ?purpose . OPTIONAL { ?relatedConsent legion:allowsUsageOf ?asset } }</pre>		
Execute		
?asset	?purpose	?relatedConsent
legion:CarbonFootprintData	legion:SDGImpactEvaluation	
legion:ConditionMonitoringConsent	legion:PredictiveMaintenance	
legion:EmissionData	legion:SDGImpactCalculation	legion:AuditRight
legion:EmissionData	legion:SDGImpactCalculation	legion:SDGEvaluationConsent
legion:MachineOperationalData	legion:ConditionMonitoring	legion:ConditionMonitoringConsent
legion:MachinePerformanceData	legion:ConditionMonitoring	
legion:ResourceCirculationData	legion:SDGImpactCalculation	legion:AuditRight
legion:ResourceCirculationData	legion:SDGImpactCalculation	legion:SDGEvaluationConsent

Source: “Original Authorship (2025)”

Purpose: This query identifies assets with defined processing purposes and any consents that allow usage of those assets. Validating one of the most critical aspects of modern data protection frameworks — purpose specification and consent. By confirming that assets have defined processing purposes and optional related consents, you validate that our proposed ontology aligns with legal requirements from frameworks like GDPR, which mandate purpose limitation.

Ontology Elements Validated:

- hasProcessingPurpose property: Validates purpose specification for assets
- ProcessingPurpose instances: Confirms existence of processing purposes
- allowsUsageOf property: Tests connection between consents and assets
- UsageConsent instances: Validates consent representation

CQ5 (Informal): How do parties come to legally binding arrangements about data sharing?

CQ5 (Formal):

```

PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX legion: <http://www.legal_interoperability_ontology#>

SELECT ?negotiation ?initiator ?responder ?resultingContract
WHERE {
  ?negotiation legion:performedBy ?initiator .
  OPTIONAL { ?responder legion:undergoes ?negotiation }
  OPTIONAL { ?resultingContract legion:createdBy ?negotiation }
}

```

Execute			
?negotiation	?initiator	?responder	?resultingContract
legion:DataSharingPolicy	legion:FraunhoferIOSB		
legion:SFWRRegistration	legion:FactoryOwner		

Source: “Original Authorship (2025)”

Purpose: This query traces contract negotiation processes, identifying initiators, responders, and resulting contracts. Validating the dynamic aspects of our proposed ontology — particularly how it models the process of establishing legal agreements rather than just their static existence. By confirming that negotiations have properly identified initiators, responders, and resulting contracts, you validate that your ontology can represent the full lifecycle of legal relationships.

Ontology Elements Validated:

- ContractNegotiation instances: Validates negotiation representation
- performedBy property: Confirms correct attribution of negotiation actions
- undergoes property: Tests passive participation in negotiations
- createdBy property: Validates outcome relationships

7 Conclusion

Thus far, grounded by DSR, and the ontology development methodology SABiO, we presented gathered knowledge, recovered by the performed SLR, grounding the prior steps towards the ontology development. Furthermore, we presented the reference ontology, which was based on Unified Foundational Ontology, and applied concepts from distinct ontologies (also provided by the SLR), e.g., SCO, IM, etc. Moreover, we translated the turtle version to an owl file, through desktop Protégé, which also allows the addition of detailed axioms, such as disjointness, and closing axioms. Finally, we populated the ontology, and performed the SPARQL queries, in order to partially validate the ontology (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; JUZGADO, 1997)- its concision and correctness. Finally, we generated the map of the instance using the web Protégé. As proposed by DSR, this work provides a novelty, while upholding scientific rigor. This section provides a glimpse of the already satisfactory contributions to the field and traces the next steps towards its validation, acceptance, and implementation.

7.1 Research Questions

As decomposed earlier, the RQs should ground a DSR project, setting boundaries, and goals. As proposed earlier, the General Research Question was **How to Achieve Legal Interoperability in IDS**. Even though we cannot yet properly answer this question, we have a foundation and a road map towards it. For now, we may answer it by providing a similarity between the Reference Architecture Model, Information Model, and ODRL. Currently, is possible to observe a relation among the *interoperability barriers*, the so-called *Legal Moments* (GRIFFO et al., 2021b), and the Policies and Rules. As for the decomposed RQs, we earlier introduced *What is legal interoperability in IDS*. The SLR has shown that we may answer it as *the capability of companies under different jurisdictions to cooperate*. The SLR has properly answered most of the conceptual questions, for the granular division of the GCP, we present the:

- *CQ1 - What is the current representation of service contracts in IDS?* Answer: The current description relies on the ODRL and a set of subkinds related to service contracts. Even though the IDSA proposes the *IDS usage control language*, there is no implementation.

- *CQ2 - How do legal aspects interfere with Data Sovereignty?* Answer: Data sovereignty is the capability of the data owner to retain ownership, by setting a data usage policy, defining who, for how long, and under what circumstances third parties may use their assets. Hence, the policy negotiation process, that leads to a contractual agreement directly interferes with data sovereignty (OLIVEIRA; SILVA; MOREIRA, 2024).
- *CQ3 - How are the legal norms and usage policies currently represented in IDS?* Answer: As observed, we discovered the relation between the legal norms, and overall policies by the ODRL, with no connection among themselves (OLIVEIRA, 2024).
- *CQ4 - What are the legal moments/positions of IDS participants?* Answer: (GRIFFO et al., 2021b) proposes the division presented in Section 5.2.4
- *CQ5 - Is personal-data management approached in IDS architecture?* Answer: The SLR has shown that, although the prior usage goal of IDS encompasses mainly non-personal data, other data spaces are using IDS technology, such as data connectors, hence, the architecture must comply with personal data management, which currently does not.

As for the GTQ, we earlier stated **How to effectively enforce legal interoperability in IDS?**. As a subjective and resource-demanding question, and as an ongoing work, we might summarize the answer to, an unambiguous representation language that allows the consent, and proper negotiation of *Data Provider* and *Data Consumer*. For an in-depth overview, and based on the performed SLR and artifacts, we have:

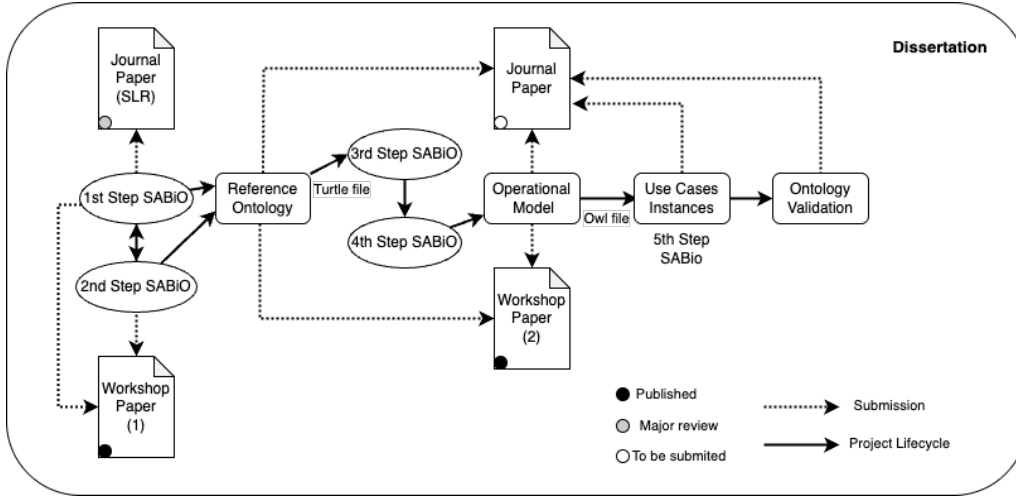
- *TQ1 - Which machine-readable specification language could address policy representation?* Answer: The SLR has revealed the lack of machine-readable representation, and seldom proposes the usage of ontologies to do so. Through our published papers and this dissertation, we also contributed to this lack by providing a domain ontology that answers a few of those open questions.
- *TQ2 - How to enforce a common understanding of contractual bindings (policies)?* Answer: As a subjective field, service contracts are composed of a non-empty set of policies. In order to represent those policies in a common understanding way, it is important to understand how each policy relates to each other, hence, the usage of ontologies is strongly advised, and suitable.

- *TQ3 - What constitutes the workflow of data exchange within IDS?* Answer: To properly answer this question, we may rely upon the reference version of *LegION-IDS*, and Figure 2.
- *TQ4 - What is the current state-of-the-art regarding contract automation?* Answer: As exposed by the SLR, there is no current (to the best of our knowledge) contract automation apparatus available. However, as previously mentioned, companies in Europe are exploring the so-called *legal testbed*. Nevertheless, there is no open discussion about it.
- *TQ5 - How could an application help resolve legal interoperability issues as a prelude to a contractual agreement?* Answer: As observed in *LegION-IDS*, a *contractual agreement* is formulated by a set of dependencies, such as *conset*, and the *offer* or *request* of a service contract. Hence, the automation of such processes would allow the fostering of legal agreements, and as a consequence, the legal availability research of each scenario, allowing the legal interoperability. Even though we do not propose a proper application, we define a set of tools to be further used in such an application, working as a foundation for this development.

Finally, the GPC **How does Proper Legal Interoperability affect IDS-based ecosystems?**. Despite the already substantial contribution, to properly answer the practical questions, we were not able to provide a full cycle of validation for DSR. We proposed a partial validation with cases retrieved from IDSA literature to verify the ontology, map, and validate the core concepts of the domain. This can be drawn as one of the greatest challenges while developing this dissertation.

7.2 Research Summary and Future Works

The research overview, along with our submission, is synthesized in Figure 36, providing a research summary.

Figure 36 – DSR treatment lifecycle and next steps

Source: “Original Authorship (2025)”

The culmination of our two-year research initiative through this Master’s thesis has yielded promising, although preliminary, results in the complex domain of legal interoperability for data spaces. Working within the inherent temporal constraints of academic research and navigating the sensitive legal landscape of legal data sovereignty presented significant methodological challenges, which constitute the primary limitation of our work.

To adaptively complete the Design Science Research lifecycle, we strategically validated a representative subset of the ontology using authoritative data extracted from the International Data Spaces Association’s position paper. This carefully selected partial instantiation proved sufficient to address several critical questions articulated in our informal functional requirements and subsequently formalized as competency questions. The successful execution of our SPARQL queries demonstrates that *LegION-IDS* not merely accommodates but extends beyond the use cases documented in the literature, providing empirical evidence of its conceptual completeness, expressive adequacy, and proper documentation.

Looking forward, we identify significant opportunities for advancing this research through comprehensive instantiation with authentic IDS contractual frameworks, incorporating both core participant data and the granular specifications contained within data transaction agreements. Such instantiation would enable more sophisticated querying and reasoning capabilities across the foundational concepts that constitute our proposed framework for legal interoperability. This would move the ontology from theoretical validation to practical application in industrial data sovereignty scenarios.

Moreover, as evidenced by our Systematic Literature Review, legal interoper-

ability in data spaces remains an emergent research domain characterized by limited accumulated knowledge and standardized practices. This thesis therefore, serves as a significant contribution to the nascent body of knowledge, establishing conceptual foundations and methodological approaches for future research. By formalizing the semantic relationships between legal concepts in industrial data spaces, our work charts a clear trajectory toward a more comprehensive understanding and implementation of legal interoperability within the International Data Spaces framework.

As future work, we identify several promising research directions. First, implementing SHACL validation would provide more robust constraint checking and verification of ontology instances against defined shapes, ensuring data quality and consistency. Second, comprehensive instantiation with real industry use cases beyond the IDSA examples would further validate the ontology’s expressiveness and utility in diverse contexts.

Most significantly, we have begun exploring the use of our ontology as a foundation for machine learning applications that can automate aspects of legal interoperability assessment. Specifically, we have initiated the development of a text classification model capable of identifying key legal concepts within contractual documents. Our preliminary model has been trained on a dataset of 505 entries, each consisting of text snippets extracted from our Systematic Literature Review paired with corresponding ontological class labels. The current implementation can classify ten distinct `legion` classes: Right, Power, Immunity, Permission, No Right, Disability, Duty, Subjection, Data Consumer, and Data Provider.

Through systematic hyperparameter optimization, we determined that the most effective neural network architecture incorporates a learning rate of **0.004**, **50** units in the hidden layer, a dropout regularization of **0.2** (20%), and training across **100** epochs. While this configuration has produced promising results, we are implementing additional safeguards against overfitting, such as early stopping mechanisms based on validation loss plateaus. The current model achieves 66% accuracy on validation data—a promising baseline for this complex classification task.

Our ongoing efforts focus on expanding the training dataset, refining the neural network architecture, and exploring advanced natural language processing techniques to improve generalization capabilities. Additionally, we envision developing a continuous integration system that would automatically instantiate the ontology from newly encountered contractual documents, creating a self-improving framework for legal interoperability assessment. These computational approaches, combined with the semantic foundation provided by our ontology, represent a significant step toward the practical automation of legal interoperability assessment in industrial

data spaces.

Bibliography

ABRAMOWICZ, W.; AUER, S.; HEATH, T. Linked Data in Business. *Business & Information Systems Engineering*, v. 58, n. 5, p. 323–326, out. 2016. ISSN 1867-0202. Disponível em: <<https://doi.org/10.1007/s12599-016-0446-0>>. Citado 3 vezes nas páginas 106, 107, and 108.

AHMADIAN, A. S.; JÜRJENS, J.; STRÜBER, D. Extending model-based privacy analysis for the industrial data space by exploiting privacy level agreements. In: *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. New York, NY, USA: Association for Computing Machinery, 2018. (SAC '18), p. 1142–1149. ISBN 978-1-4503-5191-1. Disponível em: <<https://dl.acm.org/doi/10.1145/3167132.3167256>>. Citado na página 43.

ALEXOPOULOS, K.; WEBER, M.; TRAUTNER, T.; MANNS, M.; NIKOLAKIS, N.; WEIGOLD, M.; ENGEL, B. An industrial data-spaces framework for resilient manufacturing value chains. *Procedia CIRP*, v. 116, p. 299–304, jan. 2023. Citado na página 105.

ALEXY, R. *A Theory of Constitutional Rights*. [S.l.]: Oxford University Press UK, 2002. Citado 2 vezes nas páginas 30 and 63.

ALI, N. B.; PETERSEN, K. Evaluating strategies for study selection in systematic literature studies. In: *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. New York, NY, USA: Association for Computing Machinery, 2014. (ESEM '14). ISBN 9781450327749. Disponível em: <<https://doi.org/10.1145/2652524.2652557>>. Citado na página 43.

ALTENDEITERING, M.; PAMPUS, J.; LARRINAGA, F.; LEGARISTI, J.; HOWAR, F. Data sovereignty for AI pipelines: lessons learned from an industrial project at Mondragon corporation. In: *Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI*. New York, NY, USA: Association for Computing Machinery, 2022. (CAIN '22), p. 193–204. ISBN 978-1-4503-9275-4. Disponível em: <<https://dl.acm.org/doi/10.1145/3522664.3528593>>. Citado 3 vezes nas páginas 104, 105, and 106.

BADER, S.; PULLMANN, J.; MADER, C.; TRAMP, S.; QUIX, C.; MÜLLER, A.; AKYÜREK, H.; BÖCKMANN, M.; IMBUSCH, B.; THEISSEN-LIPP, J.; GEISLER, S.; LANGE, C. The International Data Spaces Information Model – An Ontology for Sovereign Exchange of Digital Content. In: . [S.l.: s.n.], 2020. p. 176–192. ISBN 978-3-030-62465-1. Citado 6 vezes nas páginas 20, 22, 30, 69, 105, and 106.

BADER, S. R.; MALESHKOVA, M. Towards Enforceable Usage Policies for Industry 4.0. In: . [s.n.], 2019. Disponível em: <<https://www.semanticscholar.org/paper/Towards-Enforceable-Usage-Policies-for-Industry-4.0-Bader-Maleshkova/ed7e2658946e35c5a43bb83253ad3a09f28e3315>>. Citado na página 105.

BIGINI, G.; ZICHICHI, M.; LATTANZI, E.; FERRETTI, S.; D'ANGELO, G. Decentralized Health Data Distribution: A DLT-based Architecture for Data

Protection. In: *2022 IEEE International Conference on Blockchain (Blockchain)*. [s.n.], 2022. p. 97–104. Disponível em: <<https://ieeexplore.ieee.org/document/9881818/>>. Citado 3 vezes nas páginas 104, 105, and 108.

BORGO, S.; FERRARIO, R.; GANGEMI, A.; GUARINO, N.; MASOLO, C.; PORELLO, D.; SANFILIPPO, E. M.; VIEU, L. Dolce: A descriptive ontology for linguistic and cognitive engineering. *Applied ontology*, SAGE Publications Sage UK: London, England, v. 17, n. 1, p. 45–69, 2022. Citado 5 vezes nas páginas 25, 27, 51, 69, and 71.

BREIDENBACH, M.; HAMITI, F.; GULUZADE, A.; HEIBA, N.; MOHAMAD, Y.; VELASCO, C.; BELNAP, B. H.; LÜHMANN, D. Development of a flexible and interoperable architecture to customize clinical solutions targeting the care of multimorbid patients. In: *Proceedings of the 10th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*. New York, NY, USA: Association for Computing Machinery, 2023. (DSAI '22), p. 12–17. ISBN 978-1-4503-9807-7. Disponível em: <<https://dl.acm.org/doi/10.1145/3563137.3563157>>. Citado 3 vezes nas páginas 104, 105, and 107.

BROCKE, J. vom; HEVNER, A.; MAEDCHE, A. Introduction to design science research. *Progress in IS*, p. 1–13, Sep 2020. Citado na página 16.

CHITUC, C.-M.; AZEVEDO, A.; TOSCANO, C. A framework proposal for seamless interoperability in a collaborative networked environment. *Computers in Industry*, v. 60, n. 5, p. 317–338, jun. 2009. ISSN 0166-3615. Citado na página 15.

CRUZ, L. 8.1 first-order theories. *Theoremus*, Springer, p. 121, 2021. Citado na página 66.

DAVIS, E.; MARCUS, G. The scope and limits of simulation in automated reasoning. *Artificial Intelligence*, Elsevier, v. 233, p. 60–72, 2016. Citado na página 66.

DUISBERG, A. Legal aspects of ids: Data sovereignty—what does it imply? In: _____. *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage*. Cham: Springer International Publishing, 2022. p. 61–90. ISBN 978-3-030-93975-5. Disponível em: <https://doi.org/10.1007/978-3-030-93975-5_5>. Citado na página 16.

DUISBERG, A. Legal Aspects of IDS: Data Sovereignty—What Does It Imply? In: OTTO, B.; HOMPEL, M. ten; WROBEL, S. (Ed.). *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage*. Cham: Springer International Publishing, 2022. p. 61–90. ISBN 978-3-030-93975-5. Disponível em: <https://doi.org/10.1007/978-3-030-93975-5_5>. Citado 5 vezes nas páginas 60, 61, 62, 105, and 108.

FALBO, R. de A. Sabio: Systematic approach for building ontologies. *Onto. Com/odise@ Fois*, v. 1301, 2014. Citado 7 vezes nas páginas 20, 34, 36, 51, 58, 59, and 65.

FARAHANI, B.; MONSEFI, A. K. Smart and collaborative industrial iot: A federated learning and data space approach. *Digital Communications and Networks*,

v. 9, n. 2, p. 436–447, 2023. ISSN 2352-8648. Citado 3 vezes nas páginas 105, 106, and 107.

FEAGIN, J. R.; ORUM, A. M.; SJOBERG, G. *A case for the case study*. [S.l.]: The University of North Carolina Press, 2016. Citado na página 19.

FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. Methontology: from ontological art towards ontological engineering. American Association for Artificial Intelligence, 1997. Citado na página 51.

FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A.; JUZGADO, N. J. Methontology: from ontological art towards ontological engineering. American Association for Artificial Intelligence, 1997. Citado 2 vezes nas páginas 33 and 86.

FIRDAUSY, D.; SILVA, P. D. A.; SINDEREN, M. J. v.; IACOB, M. E. Semantic discovery and selection of data connectors in international data spaces. In: . [s.n.], 2022. Disponível em: <<https://research.utwente.nl/en/publications/semantic-discovery-and-selection-of-data-connectors-in-internatio>>. Citado 2 vezes nas páginas 106 and 108.

FIRDAUSY, D. R.; SILVA, P. D. A.; SINDEREN, M. V.; IACOB, M.-E. Towards a reference enterprise architecture to enforce digital sovereignty in international data spaces. *2022 IEEE 24th Conference on Business Informatics (CBI)*, 2022. Citado na página 14.

FIRDAUSY, D. R.; SILVA, P. de A.; SINDEREN, M. van; IACOB, M.-E. A data connector store for international data spaces. *Cooperative Information Systems*, p. 242–258, 2022. Citado 2 vezes nas páginas 14 and 23.

GANCK, A. D. *The New European Interoperability Framework*. 2017. Disponível em: <https://ec.europa.eu/isa2/eif_en/>. Citado na página 14.

GANZHA, M.; PAPRZYCKI, M.; PAWŁOWSKI, W.; SZMEJA, P.; WASIELEWSKA, K.; MOREIRA, J. Towards semantic interoperability between internet of things platforms. *Internet of Things*, p. 103–127, 2017. Citado na página 14.

GRAS, N. *Let's talk about IDSA's Task Force Legal*. [S.l.]: Dr. Söntje Julia Hilberg, 2023. <<https://internationaldataspaces.org/lets-talk-about-idsas-task-force-legal>>. Citado na página 15.

GRIFFO, C.; ALMEIDA, J. P. A.; GUIZZARDI, G.; NARDI, J. C. Service contract modeling in enterprise architecture: An ontology-based approach. *Information Systems*, v. 101, p. 101454, nov. 2021. ISSN 0306-4379. Citado na página 20.

GRIFFO, C.; ALMEIDA, J. P. A.; GUIZZARDI, G.; NARDI, J. C. Service contract modeling in Enterprise Architecture: An ontology-based approach. *Information Systems*, v. 101, p. 101454, nov. 2021. ISSN 0306-4379. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S030643791930506X>>. Citado 8 vezes nas páginas 29, 51, 60, 63, 86, 87, 105, and 108.

GRIFFO, C.; GUIZZARDI, G.; ALMEIDA, J. *Conceptual Modeling of Legal Relations*. [S.l.: s.n.], 2018. Citado 4 vezes nas páginas 29, 53, 54, and 79.

GUARINO, N. *Formal Ontologies and Information Systems*. [S.l.: s.n.], 1998. Citado na página 27.

GUARINO, N. *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*. [S.l.]: IOS press, 1998. v. 46. Citado 2 vezes nas páginas 25 and 27.

GUIZZARDI, G. Ontological foundations for structural conceptual models. 2005. Citado 6 vezes nas páginas 26, 28, 29, 67, 69, and 74.

GUIZZARDI, G. Ontology, ontologies and the “i” of fair. *Data Intelligence*, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , v. 2, n. 1-2, p. 181–191, 2020. Citado na página 19.

GUIZZARDI, G.; BENEVIDES, A. B.; FONSECA, C. M.; PORELLO, D.; ALMEIDA, J. P. A.; SALES, T. P. Ufo: Unified foundational ontology. *Applied ontology*, IOS Press, v. 17, n. 1, p. 167–210, 2022. Citado na página 20.

GUIZZARDI, G.; BERNASCONI, A.; PASTOR, O.; STOREY, V. C. Ontological unpacking as explanation: the case of the viral conceptual model. In: SPRINGER. *International Conference on Conceptual Modeling*. [S.l.], 2021. p. 356–366. Citado na página 29.

HECKER, D.; VOSS, A.; WROBEL, S. Data Ecosystems: A New Dimension of Value Creation Using AI and Machine Learning. In: OTTO, B.; HOMPEL, M. ten; WROBEL, S. (Ed.). *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage*. Cham: Springer International Publishing, 2022. p. 211–224. ISBN 978-3-030-93975-5. Disponível em: <https://doi.org/10.1007/978-3-030-93975-5_13>. Citado na página 106.

IMMONEN, A.; OVASKA, E.; KALAOJA, J.; PAKKALA, D. A service requirements engineering method for a digital service ecosystem. *Service Oriented Computing and Applications*, Springer, v. 10, n. 2, p. 151–172, 2016. Citado 2 vezes nas páginas 56 and 58.

IMMONEN, A.; OVASKA, E.; PAASO, T. Towards certified open data in digital service ecosystems. *Software Quality Journal*, v. 26, n. 4, p. 1257–1297, dez. 2018. ISSN 1573-1367. Disponível em: <<https://doi.org/10.1007/s11219-017-9378-2>>. Citado 2 vezes nas páginas 106 and 107.

JANEV, V.; VIDAL, M. E.; ENDRIS, K.; PUJIC, D. Managing Knowledge in Energy Data Spaces. In: *Companion Proceedings of the Web Conference 2021*. New York, NY, USA: Association for Computing Machinery, 2021. (WWW '21), p. 7–15. ISBN 978-1-4503-8313-4. Disponível em: <<https://dl.acm.org/doi/10.1145/3442442.3453541>>. Citado 2 vezes nas páginas 105 and 106.

JESSE, N. Internet of Things and Big Data: the disruption of the value chain and the rise of new software ecosystems. *AI & SOCIETY*, v. 33, n. 2, p. 229–239, maio 2018. ISSN 1435-5655. Disponível em: <<https://doi.org/10.1007/s00146-018-0807-y>>. Citado na página 106.

KEET, C. M. *An introduction to ontology engineering*. [S.l.]: University of Cape Town, 2018. Citado 7 vezes nas páginas 19, 25, 32, 65, 67, 72, and 73.

KEIL, J. M. Ontology abox comparison. In: SPRINGER. *European Semantic Web Conference*. [S.l.], 2018. p. 240–250. Citado na página 68.

KIRSTEIN, F.; BOHLEN, V. IDS as a Foundation for Open Data Ecosystems. In: OTTO, B.; HOMPEL, M. ten; WROBEL, S. (Ed.). *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage*. Cham: Springer International Publishing, 2022. p. 225–240. ISBN 978-3-030-93975-5. Disponível em: <https://doi.org/10.1007/978-3-030-93975-5_14>. Citado 4 vezes nas páginas 48, 106, 107, and 108.

KITCHENHAM, B. Procedures for performing systematic reviews. *Keele, UK, Keele Univ.*, v. 33, ago. 2004. Citado 6 vezes nas páginas 20, 38, 40, 45, 47, and 48.

LARRINAGA, F. Data sovereignty-requirements analysis of manufacturing use cases. International Data Spaces Association, 2022. Citado 3 vezes nas páginas 20, 76, and 80.

LEAL, G. da S. S.; GUÉDRIA, W.; PANETTO, H. Interoperability assessment: A systematic literature review. *Computers in Industry*, v. 106, p. 111–132, abr. 2019. ISSN 0166-3615. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0166361518303476>>. Citado na página 41.

LI, F.-L.; HORKOFF, J.; MYLOPOULOS, J.; GUIZZARDI, R. S.; GUIZZARDI, G.; BORGIDA, A.; LIU, L. Non-functional requirements as qualities, with a spice of ontology. In: IEEE. *2014 IEEE 22nd international requirements engineering conference (RE)*. [S.l.], 2014. p. 293–302. Citado 4 vezes nas páginas 52, 53, 56, and 57.

LIU, M.; TU, Z.; XU, X.; WANG, Z.; WANG, Y. A data-driven approach for constructing multilayer network-based service ecosystem models. *Software and Systems Modeling*, v. 22, n. 3, p. 919–939, jun. 2023. ISSN 1619-1374. Disponível em: <<https://doi.org/10.1007/s10270-022-01029-6>>. Citado na página 107.

LUQI, L.; STEIGERWALD, R. Rapid software prototyping. In: *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*. [S.l.: s.n.], 1992. ii, p. 470–479 vol.2. Citado na página 19.

MECKLER, S.; DORSCH, R.; HENSELMANN, D.; HARTH, A. The Web and Linked Data as a Solid Foundation for Dataspaces. In: *Companion Proceedings of the ACM Web Conference 2023*. New York, NY, USA: Association for Computing Machinery, 2023. (WWW '23 Companion), p. 1440–1446. ISBN 978-1-4503-9419-2. Disponível em: <<https://dl.acm.org/doi/10.1145/3543873.3587616>>. Citado 3 vezes nas páginas 104, 106, and 108.

MENGIST, W.; SOROMESSA, T.; LEGESE, G. Method for conducting systematic literature review and meta-analysis for environmental science research. *MethodsX*, v. 7, p. 100777, jan. 2020. ISSN 2215-0161. Citado na página 45.

MUNOZ-ARCENTALES, A.; LÓPEZ-PERNAS, S.; POZO, A.; ALONSO, ; SALVACHÚA, J.; HUECAS, G. An architecture for providing data usage and access control in data sharing ecosystems. *Procedia Computer Science*, v. 160, p. 590–597, jan. 2019. ISSN 1877-0509. Citado 2 vezes nas páginas 16 and 80.

MUNOZ-ARCENTALES, A.; LÓPEZ-PERNAS, S.; POZO, A.; ALONSO, ; SALVACHÚA, J.; HUECAS, G. An Architecture for Providing Data Usage and Access Control in Data Sharing Ecosystems. *Procedia Computer Science*, v. 160, p. 590–597, jan. 2019. ISSN 1877-0509. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1877050919317429>>. Citado 2 vezes nas páginas 104 and 108.

NAGEL, L.; LYCKLAMA, D. How to Build, Run, and Govern Data Spaces. In: OTTO, B.; HOMPEL, M. ten; WROBEL, S. (Ed.). *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage*. Cham: Springer International Publishing, 2022. p. 17–28. ISBN 978-3-030-93975-5. Disponível em: <https://doi.org/10.1007/978-3-030-93975-5_2>. Citado na página 106.

NOY, N. F.; MCGUINNESS, D. L. et al. *Ontology development 101: A guide to creating your first ontology*. [S.l.]: Stanford knowledge systems laboratory technical report KSL-01-05 and . . . , 2001. Citado na página 33.

OLIVEIRA, V. An ontology-based framework to provide legal interoperability within international data spaces. 2024. Citado na página 87.

OLIVEIRA, V.; SILVA, P.; MOREIRA, J. Legion-ids: Legal interoperability ontology for international data spaces. 2024. Citado na página 87.

OTTO, B. Interview with reinhold achatz on “data sovereignty and data ecosystems”. *Business amp; amp; Information Systems Engineering*, v. 61, n. 5, p. 635–636, 2019. Citado na página 14.

OTTO, B.; HOMPEL, M. t.; WROBEL, S. International data spaces. In: _____. *Digital Transformation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2019. p. 109–128. ISBN 978-3-662-58134-6. Disponível em: <https://doi.org/10.1007/978-3-662-58134-6_8>. Citado 3 vezes nas páginas 14, 15, and 79.

OTTO, B.; WROBEL, S.; HOMPEL, M. *Designing data spaces*. [S.l.]: Springer International Publishing, 2022. Citado na página 15.

PAN, J. Z.; STAAB, S.; ASSMANN, U.; EBERT, J.; ZHAO, Y. *Ontology-driven software development*. [S.l.]: Springer Science & Business Media, 2012. Citado na página 19.

PETTENPOHL, H.; SPIEKERMANN, M.; BOTH, J. R. International Data Spaces in a Nutshell. In: OTTO, B.; HOMPEL, M. ten; WROBEL, S. (Ed.). *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage*. Cham: Springer International Publishing, 2022. p. 29–40. ISBN 978-3-030-93975-5. Disponível em: <https://doi.org/10.1007/978-3-030-93975-5_3>. Citado 2 vezes nas páginas 23 and 104.

PORELLO, D.; GUIZZARDI, G.; SALES, T. P.; AMARAL, G. A core ontology for economic exchanges. In: SPRINGER. *International Conference on Conceptual Modeling*. [S.l.], 2020. p. 364–374. Citado na página 29.

PULLMANN, J.; PETERSEN, N.; MADER, C.; LOHMANN, S.; KEMENY, Z. Ontology-based information modelling in the industrial data space. In: *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. Limassol, Cyprus: IEEE Press, 2017. p. 1–8. Disponível em: <<https://doi.org/10.1109/ETFA.2017.8247688>>. Citado 2 vezes nas páginas 105 and 106.

REZAEI, R.; CHIEW, T.; LEE, S. A review on e-business interoperability frameworks. *Journal of Systems and Software*, v. 93, jul. 2014. Citado 3 vezes nas páginas 105, 106, and 107.

SANG, G. M.; XU, L.; VRIEZE, P. de; BAI, Y.; PAN, F. Predictive Maintenance in Industry 4.0. In: *Proceedings of the 10th International Conference on Information Systems and Technologies*. New York, NY, USA: Association for Computing Machinery, 2021. (ICIST '20), p. 1–11. ISBN 978-1-4503-7655-6. Disponível em: <<https://dl.acm.org/doi/10.1145/3447568.3448537>>. Citado 2 vezes nas páginas 104 and 108.

SARDIS, E.; GOGOUVITIS, S. V.; BOURAS, T.; GOUVAS, P.; VARVARIGOU, T. Secure Enterprise Interoperability Ontology for Semantic Integration of Business to Business Applications. In: *2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*. [s.n.], 2013. p. 68–75. Disponível em: <<https://ieeexplore.ieee.org/document/6681211/>>. Citado 2 vezes nas páginas 104 and 107.

SCERRI, S.; TUIKKA, T.; VALLEJO, I.; CURRY, E. Common European Data Spaces: Challenges and Opportunities. In: . [S.l.: s.n.], 2022. p. 337–357. ISBN 978-3-030-98635-3. Citado 2 vezes nas páginas 104 and 105.

SCHEIDER, S.; LAUF, F.; GELLER, S.; MÖLLER, F.; OTTO, B. Exploring design elements of personal data markets. *Electronic Markets*, v. 33, n. 1, p. 28, jun. 2023. ISSN 1422-8890. Disponível em: <<https://doi.org/10.1007/s12525-023-00646-3>>. Citado 2 vezes nas páginas 104 and 105.

SEIDEL, A.; WENZEL, K.; HÄNEL, A.; TEICHER, U.; WEIß, A.; SCHÄFER, U.; IHLENFELDT, S.; EISENMANN, H.; ERNST, H. Towards a seamless data cycle for space components: considerations from the growing European future digital ecosystem Gaia-X. *CEAS Space Journal*, maio 2023. ISSN 1868-2510. Disponível em: <<https://doi.org/10.1007/s12567-023-00500-4>>. Citado 2 vezes nas páginas 104 and 108.

SILVA, L. C. e.; SOBRINHO, A. d. C. C.; CORDEIRO, T. D.; MELO, R. F.; BITTENCOURT, I. I.; MARQUES, L. B.; MATOS, D. D. M. d. C.; SILVA, A. P. d.; ISOTANI, S. Applications of convolutional neural networks in education: A systematic literature review. *Expert Systems with Applications*, v. 231, p. 120621, 2023. ISSN 0957-4174. Citado 2 vezes nas páginas 39 and 108.

SIMPERL, E. Reusing ontologies on the semantic web: A feasibility study. *Data & Knowledge Engineering*, Elsevier, v. 68, n. 10, p. 905–925, 2009. Citado na página 32.

SIRIN, E.; PARSIA, B.; GRAU, B. C.; KALYANPUR, A.; KATZ, Y. Pellet: A practical owl-dl reasoner. *Journal of Web Semantics*, Elsevier, v. 5, n. 2, p. 51–53, 2007. Citado na página 66.

SMITH, P. A.; ALLEE, V. Value-creating networks: organizational issues and challenges. *The learning organization*, Emerald Group Publishing Limited, 2009. Citado na página 25.

SNYDER, H. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, v. 104, p. 333–339, 2019. Citado 2 vezes nas páginas 19 and 38.

SOLMAZ, G.; CIRILLO, F.; FÜRST, J.; JACOBS, T.; BAUER, M.; KOVACS, E.; SANTANA, J. R.; SÁNCHEZ, L. Enabling data spaces: existing developments and challenges. In: *Proceedings of the 1st International Workshop on Data Economy*. New York, NY, USA: Association for Computing Machinery, 2022. (DE '22), p. 42–48. ISBN 978-1-4503-9923-4. Disponível em: <<https://dl.acm.org/doi/10.1145/3565011.3569058>>. Citado 2 vezes nas páginas 104 and 106.

STELLMACH, C.; MUZOORA, M. R.; THUN, S. Digitalization of health data: Interoperability of the proposed european health data space. *Studies in Health Technology and Informatics*, 2022. Citado na página 15.

STUMME, G. Formal concept analysis on its way from mathematics to computer science. *Conceptual Structures: Integration and Interfaces*, p. 2–19, 2002. Citado na página 19.

SUÁREZ-FIGUEROA, M. C.; GÓMEZ-PÉREZ, A.; VILLAZÓN-TERRAZAS, B. How to write and use the ontology requirements specification document. In: SPRINGER. *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*. [S.l.], 2009. p. 966–982. Citado 3 vezes nas páginas 33, 51, and 58.

SUÁREZ-FIGUEROA, M.; GOMEZ-PEREZ, A.; TERRAZAS, B. V. *How to Write and Use the Ontology Requirements Specification Document*. [S.l.: s.n.], 2009. v. 5871. 982 p. ISBN 978-3-642-05150-0. Citado na página 19.

TAMBASSI, T. Completeness in information systems ontologies. *Axiomathes*, v. 32, n. S2, p. 215–224, 2021. Citado na página 19.

TARDIEU, H. Role of Gaia-X in the European Data Space Ecosystem. In: OTTO, B.; HOMPEL, M. ten; WROBEL, S. (Ed.). *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage*. Cham: Springer International Publishing, 2022. p. 41–59. ISBN 978-3-030-93975-5. Disponível em: <https://doi.org/10.1007/978-3-030-93975-5_4>. Citado 4 vezes nas páginas 104, 105, 106, and 107.

TCHOFFA, D.; FIGAY, N.; GHODOUS, P.; EXPOSITO, E.; KERMAD, L.; VOSGIEN, T.; MHAMED, A. E. Digital factory system for dynamic manufacturing network supporting networked collaborative product development. *Data & Knowledge Engineering*, v. 105, n. C, p. 130–154, set. 2016. ISSN 0169-023X.

Disponível em: <<https://doi.org/10.1016/j.datak.2016.02.004>>. Citado 2 vezes nas páginas 107 and 108.

THEISSEN-LIPP, J.; KOCHER, M.; LANGE, C.; DECKER, S.; PAULUS, A.; POMP, A.; CURRY, E. Semantics in Dataspaces: Origin and Future Directions. In: *Companion Proceedings of the ACM Web Conference 2023*. New York, NY, USA: Association for Computing Machinery, 2023. (WWW '23 Companion), p. 1504–1507. ISBN 978-1-4503-9419-2. Disponível em: <<https://dl.acm.org/doi/10.1145/3543873.3587689>>. Citado 4 vezes nas páginas 104, 106, 107, and 108.

WEICHHART, G.; GUÉDRIA, W.; NAUDET, Y. Supporting interoperability in complex adaptive enterprise systems. *Data Knowledge Engineering*, v. 105, n. C, p. 90–106, set. 2016. ISSN 0169-023X. Citado na página 16.

WEICHHART, G.; NAUDET, Y. Ontology of Enterprise Interoperability Extended for Complex Adaptive Systems. In: *Proceedings of the Confederated International Workshops on On the Move to Meaningful Internet Systems: OTM 2014 Workshops - Volume 8842*. Berlin, Heidelberg: Springer-Verlag, 2014. v. 8842, p. 219–228. ISBN 978-3-662-45549-4. Disponível em: <https://doi.org/10.1007/978-3-662-45550-0_23>. Citado 2 vezes nas páginas 105 and 108.

WIERINGA, R. Design science as nested problem solving. *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology - DESRIST '09*, May 2009. Citado 2 vezes nas páginas 16 and 18.

WIERINGA, R. J. *Design science methodology for information systems and software engineering*. [S.l.]: Springer, 2014. Citado na página 16.

WILKINSON, M. D.; DUMONTIER, M.; AALBERSBERG, I. J.; APPLETON, G.; AXTON, M.; BAAK, A.; BLOMBERG, N.; BOITEN, J.-W.; SANTOS, L. B. da S.; BOURNE, P. E.; BOUWMAN, J.; BROOKES, A. J.; CLARK, T.; CROSAS, M.; DILLO, I.; DUMON, O.; EDMUNDS, S.; EVELO, C. T.; FINKERS, R.; GONZALEZ-BELTRAN, A.; GRAY, A. J. G.; GROTH, P.; GOBLE, C.; GRETHE, J. S.; HERINGA, J.; HOEN, P. A. C. 't; HOOFT, R.; KUHN, T.; KOK, R.; KOK, J.; LUSHER, S. J.; MARTONE, M. E.; MONS, A.; PACKER, A. L.; PERSSON, B.; ROCCA-SERRA, P.; ROOS, M.; SCHAIK, R. van; SANSONE, S.-A.; SCHULTES, E.; SENGSTAG, T.; SLATER, T.; STRAWN, G.; SWERTZ, M. A.; THOMPSON, M.; LEI, J. van der; MULLIGEN, E. van; VELTEROP, J.; WAAGMEESTER, A.; WITTENBURG, P.; WOLSTENCROFT, K.; ZHAO, J.; MONS, B. The fair guiding principles for scientific data management and stewardship. *Scientific Data*, Nature Publishing Group, v. 3, n. 11, p. 160018, mar. 2016. ISSN 2052-4463. Citado na página 41.

WOHLIN, C.; KALINOWSKI, M.; FELIZARDO, K. R.; MENDES, E. Successful combination of database search and snowballing for identification of primary studies in systematic literature studies. *Information and Software Technology*, v. 147, p. 106908, jul. 2022. ISSN 0950-5849. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0950584922000659>>. Citado na página 44.

APPENDIX A – Ontology Requirements

Specification Document - ORSD

Legal Interoperability Ontology for International Data Spaces - LegION.
1. Purpose
To provide a legal interoperability domain description in order to foster an unambiguous representation of policies, i.e., service contracts, within International Data Spaces architecture.
2. Scope
A knowledge-based approach ranging from a semi-informal to a semi-formal degree of formality. Based on a middle-out architecture. The level of granularity is defined by a continuous integration with the NFRs and FRs. The scope of the ontology is bounded by the description of legal concepts, legal agreements, legal lacks/burdens and entitlements, data sovereignty, contractual compliance, and grounded by legal norms. The proposed ontology is strictly a domain ontology, with foundational integration of top-level ontologies, such as UFO.
3. Implementation Language
The ontology is retrieved in two different layers, the reference ontology, and the operational ontology. For this ontology, we developed a reference ontology implemented in OntoUML, which is a graphical extension of the Unified Foundational Ontology (UFO) applied through UML-like notation, with further translation for the operational ontology in Web Ontology Language (OWL).
4. Intended End-Users
User 1. IDS participant User 2. Potential IDS participants User 3. Legal and Compliance Professionals User 4. Data Governance Specialists and Protection Officers User 5. Software engineers and researchers who wish to acknowledge or expand the domain of legal interoperability in IDS
5. Intended Uses
Use 1: Negotiation of contractual terms regarding legal aspects Use 2: Compliance cost reduction Use 3: Semi-automation capability of data exchange contract negotiation Use 4: Data sovereignty and legal norms enforcement Use 5: Verified model for worldwide policies overview Use 6: Legal Interoperability endeavor
5.1 Intended Uses Grounding Assumptions
Use 1: By allowing the observance of legal interoperability barriers, legal availability, and coherence between the governing laws Use 2: Fostering faster compliance by pre-validating the legal availability of possible data providers and data users Use 3: Delivering a machine-readable RDF syntax with the further possibility of machine learning models integration to leverage participants' negotiation of contractual lacks and entitlements Use 4: Providing a legal foundation for legal norms compliance, yet, upholding data owners' data sovereignty principle Use 5: Allowing the implementation, regardless of the governing law by legal professionals, data governance specialists, and potential IDS participants Use 6: Addressing the EIF, the ontology allows to achieve a legal interoperability overview of potential IDS participants, comparing their governing laws, and unpacking their potential coherence and legal availability.
6. Ontology Requirements
6.1. Non-Functional Requirements
NFR 1: Support for Natural Language (English) -> Design concern NFR 2: Addresses the European Interoperability Framework -> Adaptation Concern NFR 2.1: Explicitly Defines and Leverages Legal Interoperability NFR 3: Follow the Fair Principles: Follow the FAIR principles -> Design/Performance Concern NFR 3.1: Findable: It must have an open access; NFR 3.2: Accessible: It shall possess unique URIs; NFR 3.3: Interoperable: Use a formal and broadly applicable set of concepts and languages for representation; NFR 3.4: Reusability: Data meets domain-relevant accepted standards and holds a clear and accessible data usage license. For the Prior, we provide an ISO dictionary of terms, and for the latter, we employ the MIT license. NFR 4: Unified Foundational Ontology (UFO) as foundational ontology NFR 5: Cover the aspects retrieved from Service Contract Ontology (SCO) NFR 6: Specific NFRs, documentation, OWL file, and VPP file are available at the open-access repository: < http://www.github.com/VictorBenoiston/ >

Source: “Original Authorship (2025)”

6.2. Functional Requirements
6.2.1. Informal Competency Questions
Competency Questions Group 1: SCO and EIF Related
CQ1: What are the <u>legal entitlements</u> of the <u>service provider</u> ?[SCO] CQ2: What are the <u>legal burdens/lacks</u> of the <u>service provider</u> ?[SCO] CQ3: What are the <u>legal entitlements</u> of the <u>service consumer</u> ?[SCO] CQ4: What are the <u>legal burdens/lacks</u> of the <u>service consumer</u> ?[SCO] CQ5: What are the <u>Interoperability barriers in Service Contract 1</u> ?[EIF] CQ6: What <u>contracts</u> represent <u>joint controllership</u> ? [EIF]
Competency Questions Group 2: IM and RAM Related
CQ7: What are <u>data user's permissions and duties</u> ? [IM] CQ8: What <u>contracts</u> characterize <u>data rent</u> ? [RAM] CQ9: What <u>contracts</u> characterize <u>data purchase</u> ? [RAM]
6.2.1 Formal Competency Questions
CQ1: <pre> PREFIX owl: <http://www.w3.org/2002/07/owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> PREFIX legion: <http://www.legal_interoperability_ontology#> SELECT ?owner ?asset WHERE { ?owner legion:owns ?asset . OPTIONAL { ?owner legion:hasContractualPartyName ?ownerName } }</pre> CQ2: <pre> PREFIX owl: <http://www.w3.org/2002/07/owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> PREFIX legion: <http://www.legal_interoperability_ontology#> SELECT ?entity ?entitlement ?relatedRight WHERE { ?entity legion:hasLegalEntitlement ?entitlement . OPTIONAL { ?entitlement legion:allowsUsageOf ?relatedRight } }</pre> CQ3: <pre> PREFIX owl: <http://www.w3.org/2002/07/owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> PREFIX legion: <http://www.legal_interoperability_ontology#> SELECT ?rule ?constraint ?creationDate WHERE { ?rule legion:hasRuleConstraint ?constraint . OPTIONAL { ?rule legion:hasRuleCreationDate ?creationDate } }</pre>

Source: “Original Authorship (2025)”

6.2.1 Formal Competency Questions

CQ4:

```

PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX legion: <http://www.legal_interoperability_ontology#>

SELECT ?asset ?purpose ?relatedConsent
WHERE {
  ?asset legion:hasProcessingPurpose ?purpose .
  OPTIONAL { ?relatedConsent legion:allowsUsageOf ?asset }
}

```

CQ5:

```

PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX legion: <http://www.legal_interoperability_ontology#>

SELECT ?negotiation ?initiator ?responder ?resultingContract
WHERE {
  ?negotiation legion:performedBy ?initiator .
  OPTIONAL { ?responder legion:undergoes ?negotiation }
  OPTIONAL { ?resultingContract legion:createdBy ?negotiation }
}

```

7. Pre-Glossary of Terms

7.1. Data retrieved from RSL

After performing a SLR in 8 databases, and extracting the keywords from 41 papers, the most relevant keywords are represented as follows.

• Legal Agreement	• Governing Law	• Policy
• Legal Aspect	• Coherence	• Rule
• Compliance	• IDS Core Participant	• Action
• Data Usage Policy	• Service Consumer	• Asset
• Industrial ecosystems	• Service Provider	• Data

7.2. Data retrieved from ISO standards

In order to foster the Reuse (enhancing the ontology FAIRness), we propose the retrieval and verification of the acquired terms by the SLR grounding the ontology on ISO standards. We employed the usage of the general ISO dictionary of terms and the specific ISO 17442:2020 - Legal entity identifier (LEI). The complete dictionary is available in:

7.3. Terms from Competency Questions + Frequency

Contract	3	Legal Entitlement	2	Permission	1
Service Provider	2	Legal Lack/Burden	2	Duty	1
Legal Agreement	2	Data Rent	1		
Service Consumer	2	Data Purchase	1		

8. Objects

GDPR, Rights, Duties, Immunities, No Powers, Liabilities, Legal Moments, etc.

Source: “Original Authorship (2025)”

APPENDIX B – Systematic Literature Review Papers

Table 8 – Retrieved Legal Aspects and their Relevance Towards RQs

ID	RLA	References	RQ1	RQ2	RQ3	RQ4	RQ5
RLA1	Domains of Business	(MUNOZ-ARCENTALES et al., 2019b) (SANG et al., 2021) (SOLMAZ et al., 2022) (BREIDENBACH et al., 2023) (THEISSEN-LIPP et al., 2023) (MECKLER et al., 2023) (SARDIS et al., 2013) (BIGINI et al., 2022) (TARDIEU, 2022)			X		X
RLA2	Personal / Non-Personal Data	(ALTENDEITERING et al., 2022) (MECKLER et al., 2023) (SCHEIDER et al., 2023) (??) (SEIDEL et al., 2023) (PETTENPOHL; SPIEKERMANN; BOTH, 2022) (SCERRI et al., 2022)	X	X	X		X

Continued on next page

Table 8 – *Continued from previous page*

ID	RLA	References	RQ1	RQ2	RQ3	RQ4	RQ5
RLA3	Usage and Data Policies	(ALEXOPOULOS et al., 2023) (ALTENDEITERING et al., 2022) (SCHEIDER et al., 2023) (BIGINI et al., 2022) (REZAEI; CHIEW; LEE, 2014) (GRIFFO et al., 2021b) (BADER; MALESHKOVA, 2019) (SCERRI et al., 2022)	X	X	X	X	X
RLA4	Interoperability Constraints	(FARAHANI; MONSEFI, 2023) (WEICHHART; NAUDET, 2014) (ALTENDEITERING et al., 2022) (JANEV et al., 2021) (BREIDENBACH et al., 2023) (PULLMANN et al., 2017) (GRIFFO et al., 2021b) (BADER et al., 2020) (TARDIEU, 2022)	X	X	X	X	X
RLA5	Smart Contracts / Contract Automation	(WEICHHART; NAUDET, 2014) (PULLMANN et al., 2017) (BIGINI et al., 2022) (GRIFFO et al., 2021b) (BADER et al., 2020) (DUISBERG, 2022b) (TARDIEU, 2022)		X	X	X	X

Continued on next page

Table 8 – *Continued from previous page*

ID	RLA		References	RQ1	RQ2	RQ3	RQ4	RQ5
RLA6	Semantic peal	Ap-	(REZAEI; CHIEW; LEE, 2014) (JANEV et al., 2021) (THEISSEN-LIPP et al., 2023) (MECK- LER et al., 2023) (BADER et al., 2020) (PULL- MANN et al., 2017) (ABRAMOWICZ; AUER; HEATH, 2016) (IMMONEN; OVASKA; PAASO, 2018) (JESSE, 2018) (FIRDAUSY et al., 2022a)	X	X	X		
RLA7	AI Usage IDS	in	(FARAHANI; MON- SEFI, 2023) (AL- TENDEITERING et al., 2022) (JANEV et al., 2021) (SOL- MAZ et al., 2022) (THEISSEN-LIPP et al., 2023) (HECKER; VOSS; WROBEL, 2022) (KIRSTEIN; BOHLEN, 2022) (NAGEL; LYCK- LAMA, 2022) (TARDIEU, 2022)		X		X	X

Continued on next page

Table 8 – Continued from previous page

ID	RLA	References	RQ1	RQ2	RQ3	RQ4	RQ5
RLA8	Cloud	(FARAHANI; MONSEFI, 2023) (REZAEI; CHIEW; LEE, 2014) (TCHOFFA et al., 2016) (BREIDENBACH et al., 2023) (THEISSEN-LIPP et al., 2023) (SARDIS et al., 2013) (LIU et al., 2023) (TARDIEU, 2022)		X	X	X	X
RLA9	IDS Usage in Open Spaces	(ABRAMOWICZ; AUER; HEATH, 2016) (IMMONEN; OVASKA; PAASO, 2018) (KIRSTEIN; BOHLEN, 2022) (TARDIEU, 2022)	X		X	X	X

Continued on next page

Table 8 – *Continued from previous page*

ID	RLA	References	RQ1	RQ2	RQ3	RQ4	RQ5
RLA10	Future Works Addressing Legal Aspects	(TCHOFFA et al., 2016) (MUNOZ-ARCENTALES et al., 2019b) (SILVA et al., 2023) (WEICH-HART; NAUDET, 2014) (SANG et al., 2021) (THEISSEN-LIPP et al., 2023) (MECKLER et al., 2023) (BIG-INI et al., 2022) (ABRAMOWICZ; AUER; HEATH, 2016) (SEIDEL et al., 2023) (FIR-DAUSY et al., 2022a) (GRIFFO et al., 2021b) (DUISBERG, 2022b) (KIRSTEIN; BOHLEN, 2022)			X	X	X