Legal Interoperability in International Data Spaces: A Systematic Literature Review

Victor Benoston Jaes de Oliveira[a] (victor.oliveira53295@alunos.ufersa.edu.br), Patrício de Alencar Silva[a] (patricio.alencar@ufersa.edu.br), João Luiz Rabelo Moreira[b] (j.luizrebelomoreira@utwente.nl)

[a] Federal Rural University of the Semiarid, Av. Francisco Mota, 572, Costa e Silva, Mossoró, 59625-900, Rio Grande do Norte, Brazil
[b] University of Twente, Enschede 7522, Netherlands

Corresponding Author:
Victor Benoiston Jales de Oliveira
Francisco Mota, 572 - Pres. Costa e Silva, Mossoró - RN, 59625-900. Brazil
Tel: +55 (84) 99902-8747
Email: victor.oliveira53295@alunos.ufersa.edu.br ORCID: 0009-0000-7026-6929

Legal Interoperability in International Data Spaces: A Systematic Literature Review

Victor Benoiston Jales de Oliveira[a,*], Patrício de Alencar Silva[a], João Luiz Rabelo Moreira[b]

[a] *Federal Rural University of the Semiarid, Av. Francisco Mota, 572, Costa e Silva, Mossoró, 59625-900, Rio Grande do Norte, Brazil*

[b] *University of Twente, Enschede 7522, Netherlands*

Abstract

The International Data Spaces (IDS) is built upon two major principles, trust within its participants, and data sovereignty. The latter is defined as the capability of data owners to define who, for how long, and to what means third parties may use them. The IDS initiative is grounded by the European Interoperability Framework, which provides six layers, the former, legal, organizational, semantic, and technical, and the recently added, interoperability governance, and integrated public service governance. Our SLR focuses on the former layers – Legal, organizational, semantic, and technical - which guide the Reference Architecture Model and set the standards for the implementation of IDS. Even though the Legal interoperability layer is fundamental, supporting the remaining layers, existing literature emphasizes organizational and semantical. This study proposes a Systematic Literature Review regarding the gap in Legal Interoperability within IDS. We retrieved 40 papers from ACM Digital Library, EI Compendex, IEEE, ScienceDirect, Scopus, and SpringerLink databases. Based on the review, we discuss the legal aspects retrieved from the literature, that somehow foster the availability of legal interoperability within IDS. The legal field of IDS still lacks research, hence, we broaden our population to similar data spaces, to retrieve important concepts, that may be applied to IDS. We propose a set of retrieved legal aspects and provide several open issues to be addressed in the legal field of IDS.

Keywords: International Data Space, Legal Interoperability, Systematic Literature Review, Legal Aspects

1. Introduction

Throughout the Industrial Revolution, organizations realized that in order to grow, remain competitive, and be up to date with commercial needs, they could not restrict their usage to internal and publicly available data sources (Jarke et al., 2019). To stand out in the Industrial race, they came upon the necessity to access other organizations' data, hence, creating the first concept of a thriving addition to the business ecosystems, a data space (Solmaz et al., 2022). Even though the

*Corresponding author.

Email addresses: `victor.oliveira53295@alunos.ufersa.edu.br` (Victor Benoiston Jales de Oliveira), `patricio.alencar@ufersa.edu.br` (Patrício de Alencar Silva), `j.luizrebelomoreira@utwente.nl` (João Luiz Rabelo Moreira)

first concept of Industrial Data Space dates far back as 2016 (Ahmadian et al., 2018), it is possible to retrieve similar approaches back in the 1990s (Ayres, 1996). Industrial Data Spaces were a safe and trustable environment, in which companies could rely on each other's data, by offering financial, services or data itself in exchange, it also developed the concept of Data Sovereignty (Jarke et al., 2019), which essentially, consists of providing the data owner the power to control the data access, for how long, by whom, and as restricted as it desires to be.

In 2016 a new initiative arose, inheriting the fundaments and needs of the Industrial Data Spaces, the International Data Spaces (IDS)[1] Initiative was developed to enforce a trustable and safe data ecosystem while upholding Data Sovereignty and expanding the approach across multiple countries. Diverging from the simplicity of an Industrial Data Space, IDSs have a complex lifecycle to ensure their principles, an IDS comprises 4 categories of participants, i.e.: Core Participants; Intermediary; Software/Service Provider; and Governance Body. A swift glimpse at the Core Participants consists of a data owner, data provider, data consumer data user, and app provider. Those participants are involved and required every time data is exchanged in the IDS, to have a deep understanding of IDS architecture, refer to IDS Reference Architecture Model (RAM)Otto et al. (2019).

Along with RAM, the International Data Spaces Association (IDSA), which is a non-profitable-Organization empowered to manage, validate, and set standards for the IDS initiative, delivers a so-called Dataspace Protocol[2], which defines dataspace Interoperability. Key roles in dataspace interoperability are the standards, one of which is referenced in the EU Data Act, the ISO 19941 Cloud Computing Interoperability and Portability[3], and the main one is declared as the European Interoperability Framework (EIF)[4], aiming to create a digital single market in Europe. The EIF defines interoperability in 4 different, yet hierarchical layers, i.e.: Legal Interoperability, Organizational Interoperability, Semantical Interoperability, and Technical Interoperability. Although in 2024 the EIF has been updated, with the inclusion of two more layers of interoperability (Interoperability governance and integrated public service governance), it does not reflect on an inquiry to change our protocol and research questions, once the new layers focus on the public integration amidst the European companies, and the IDS initiative relies on the industrial nuances of EIF. The present SLR has been performed from May 2023 to December 2023.

2. Related Work

Although the International Data Spaces initiative is a relatively new domain, data ecosystems have been out there since the 2000s, along with Industry 4.0. Hence, since the first glimpse of exchanging data amidst different companies in order to create value, challenges such as interoperability also arose. In its various forms, such as data interoperability, enterprise interoperability, and others, we may summarize all facets into a few definitions of interoperability, such as (1) 'the ability of organizations to interact towards mutually beneficial goals, involving the sharing of information and knowledge between these organizations, through the business processes they support, by means of exchanging of data between their ICT systems', the European Interoperability Framework delivers this definition, such framework guided the Reference Architecture Model, which defines the practices and constraints regarding the IDS infrastructure, this topic will be thoroughly addressed

[1] https://internationaldataspaces.org/

[2] https://docs.internationaldataspaces.org/ids-knowledgebase/v/dataspace-protocol/overview/readme

[3] https://www.iso.org/standard/66639.html

[4] https://ec.europa.eu/isa2/eif_en/

in further sections. Whereas da Silva Serapião Leal et al. (2019) defines interoperability as (2) the capability of systems to exchange data/information and use the exchanged information/data, adding the definition of data usage to the concept of interoperability. Finally, Rezaei et al. (2014) defines interoperability as (3) the ability of ICT systems and the business processes they support to exchange data and enable the sharing of information and knowledge. The latter, proposes a first glimpse at an important concept of interoperability, as the capability of sharing knowledge, which will be further addressed in web semantics, knowledge graphs, and ontologies. Active interoperability among companies, systems, or even sectors of the same companies, is reflected in numerous benefits, such as shown in Rezaei et al. (2014), improved efficiency, transparency, accountability, etc. At the same time, a lack of interoperability may lead to high costs, due to performance, data heterogeneity, and other related problems.

Rezaei et al. (2014) propose four different layers of interoperability, i.e., technical (this layer of interoperability focuses on the machine communication, hence, protocols and infrastructure), syntactic (summarized as the capability of exchanging data, it relates to data formats), semantic (definition of the content itself, and deals with human resources, instead of machines. this interoperability layer focuses on the essence of knowledge, and common understanding regarding the definition of the content being exchanged), and organizational (fostered by its name, this layer aims at the capability of companies to properly exchange data with minimal loss) interoperability. Furthermore, da Silva Serapião Leal et al. (2019) base their work on the EIF, adopting the four layers of interoperability (the research was conducted in 2018 and 2019, hence, the current EIF architecture comprised only four layers of interoperability), i.e., legal (englobes legislations and policies, usually expressed in legal elements and business rules), organizational (as the previous definition of organizational, it focuses on the organizational barriers regarding the incompatibility among companies), semantical (defined by the information syntactic and information semantics), and technical (applications and infrastructures linking ICT services) layers. After a swift analysis, it is possible to observe that the related works avoid approaching the legal layer of interoperability. Nevertheless, da Silva Serapião Leal et al. (2019) proposes the adoption of the EIF, however, it joints the legal and organizational layers, based on the assumption that legislation incompatibilities are included in the organizational barriers. On the other hand, the EIF defines legal interoperability as the capability of organizations operating under different legal frameworks to work together by aligning policies and strategies, requiring that current legislation does not block the proposed policies, by generating clear agreements on how to deal with those differences across borders, and even allowing the inclusion of new legislation. Hence, comprising legal interoperability into organizational interoperability may not be viable, and, even more problematic, not considered.

Even though the new version of EIF incorporated two more layers of interoperability, i.e., interoperability governance (decisions on interoperability frameworks, institutional agreements, and other aspects of monitoring interoperability at a national level), and Integrated public service governance (integration of public service for coordination and governance of multiple organizations), the four previous layers will lead the development of the present systematic literature review (SLR), due to the enough alignment with IDS architecture. With this SLR, we propose a centered approach on the legal interoperability layer, analyzing all retrieved works based on legal aspects of data ecosystems, and pointing to literature gaps, future works, challenges, and opportunities.

3. Systematic Literature Review

The afore-proposed Systematic Literature Review (SLR) follows the guidelines defined by Kitchenham (2004), where the author 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.

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 1. 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 1 by the bidirectional arrow).

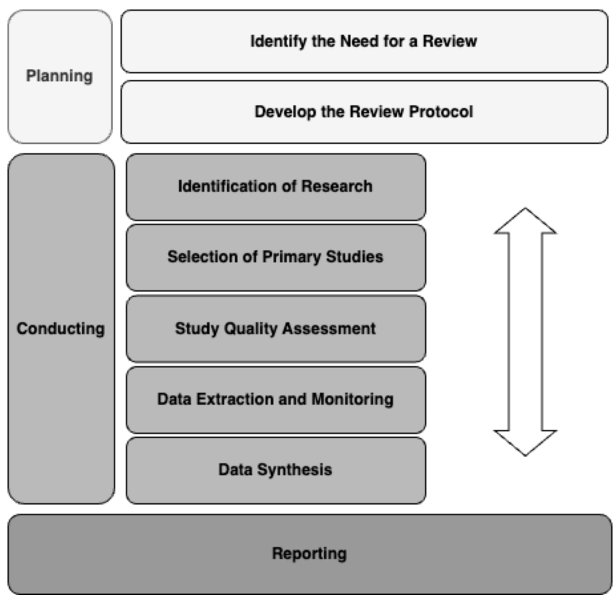


Figure 1: Systematic Literature Review Main Phases

3.1. Review Protocol

Motivated by the intrinsic need to enforce legal interoperability and legal digital service compliance within IDS, In this particular study, we propose to identify all relevant research regarding the Legal Aspects of data exchange/negotiation performed within an IDS architecture, to identify gaps in the current framework and define the state-of-the-art legal interoperability in IDS. The protocol will specify the methodology used to perform such a review and foster the reduction of research bias. The proposed SLR shall encompass nuances available in current literature, raising the awareness of IDS Participants, data governance specialists, protection officers, companies that would like to join IDS, and researchers/students in the field. The protocol is thoroughly stated in the next sections.

3.1.1. 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 Data Spaces?*
- *RQ2. What is the Conceptual relationship between Data Sovereignty and Legal Interoperability regarding IDS?*
- *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[5], it is possible to infer that it is still a new technology, and there is ongoing research (once it is on its first version);
- 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[6], 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

Finally, we set the keywords "Data Sovereignty" as a synonym of "Control Over Data" and "Power", addressed in the Outcome. For "International Data Spaces" we have the alternative of "Industrial Data Spaces", which is related to Intervention, and finally "Ontology", with a few synonyms/alternatives such as "Conceptual Model", "Knowledge Graph" and "Taxonomy", also addressing Intervention. Henceforth, we encompass a majority part of related words regarding our domain.

[5]https://internationaldataspaces.org/dataspace-protocol-ensuring-data-space-interoperability/
[6]https://internationaldataspaces.org/lets-talk-about-idsas-task-force-legal/

3.1.2. Paper Selection

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 first logic grid (Aromataris & Riitano, 2014), guided by the proposed PICOC, is available in Table 1 and will be further extended by related terms retrieved from the search.

Population	**Intervention**	**Comparison**	**Outcome**	**Context**
IDS Participants	Ontology	Reference Models	Legal Ontology Representation	IDS

Table 1: First Logic Grid, Aligned With PICOC Elements Available in the RQs

Due to the complexity and uniqueness of the proposed domain of research, we explored the main available databases (da Silva Serapião Leal et al., 2019) such as ScienceDirect, Taylor & Francis Online, and SpringerLink[7], and even expanded the search in order to retrieve the most related works. Other sources were ACM Digital Library[8], EI Compendex[9], IEEE Digital Library[10], and Scopus[11]. After several trial searches with different combinations of search terms (combinations of the PICOC elements), we enhanced the logic grid as shown in Table 2. Further, we randomly selected 10 papers from the database outputs and performed a data analysis, extracting the text, and populating a word cloud, available in Figure 2, allowing a comparison and assessment of the retrieved papers.

Population	**Intervention**	**Comparison**	**Outcome**	**Context**
IDS Participants	Ontology	Reference Models	Legal Ontology	IDS
Data Governance Specialists	Knowledge Graph	Reference Model Architecture	Legal Interoperability	Industrial Data spaces
Business Representants	Knowledge Representation	IDS Information Model	Data Sovereignty	Business Ecosystem
N/A	N/A	N/A	Data Exchange Policies	N/A

Table 2: Second Logic Grid, Aligned With PICOC Elements Available in the RQs

Finally, we reduced the main string twofold, a specific string for the IEEE Digital Library had

[7]https://link.springer.com/
[8]https://dl.acm.org/
[9]https://www.engineeringvillage.com/
[10]https://www.ieee.org/
[11]https://www.scopus.com/

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 3 summarizes the search strings 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*”))

Table 3: Search String by Database

Figure 2: Word Cloud From the Text of 10 papers Retrieved From the Databases

Upholding prior knowledge in the field, and initial discussion with specialists, we may establish the first round of inclusion and exclusion criteria, hereafter, referred to as step 1, as follows in

Table 4. The first step consists of reading the metadata available, i.e., title, abstract, and keywords (adaptive reading approach (Ali & Petersen, 2014)). All authors have performed the first step in the retrieved 127 papers to properly address such a technique.

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

Table 4: Step 1 Inclusion and Exclusion Criteria

Criteria such as "studies about interoperability" might sound ambiguous. Still, 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 et al., 2018), has been active for a longer period. Hence, we may expand the period from 5 to 10 years to better encompass parallel studies. Finally, it is important to embrace 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 5. Finally, the iterations lifecycle is synthesized in Figure 3.

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

Table 5: Step 2 Inclusion and Exclusion Criteria

To verify and validate the quality of the steps in the review protocol, a *pilot beding* has been performed, on which 10 random papers have been selected and read by at least two reviewers. The Data has been extracted, and the papers that didn't satisfy the inclusion criteria have been excluded. Furthermore, for better time management, reliability, and flexibility, we opted for the use of the SLR specialized tool Parsifal[12], which allows us to plan, manage, and conduct the SLR in a shared platform.

3.2. Paper Analysis and Discussions

Performing the developed systematic protocol, the initial search in the digital databases, specifically: ACM Digital Library (33), EI Compendex (3), IEEE Digital Library (23), Science@Direct (40), Scoups (3), and SpringerLink (23). The first set of papers for analysis is composed of 127

[12] `https://parsif.al/`

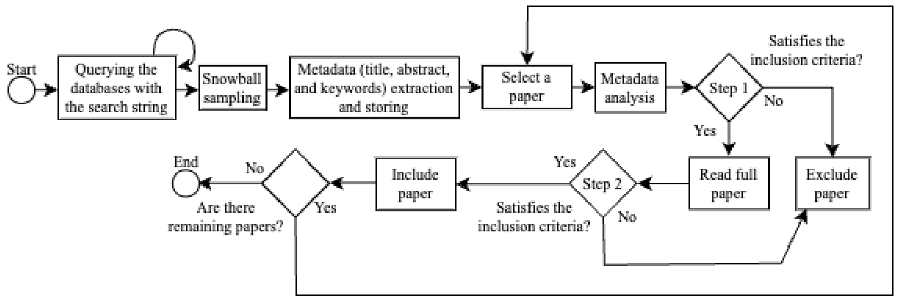


Figure 3: Paper Selection Iteration Lifecycle

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 the first set of selected papers in full, 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 4.

3.3. Quality Assessment and Data Extraction Fields

Strictly following the approached guideline (Kitchenham, 2004), and supported by parallel methodologies, such as Okoli (2015), 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 et al., 2020)), and the subjectivity of the quality assessment must be thoroughly concise and documented. The proposed quality assessment checklist and its respective founding assumption are presented in Table 6. Its sole goal is to provide a score for each particular work.

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.

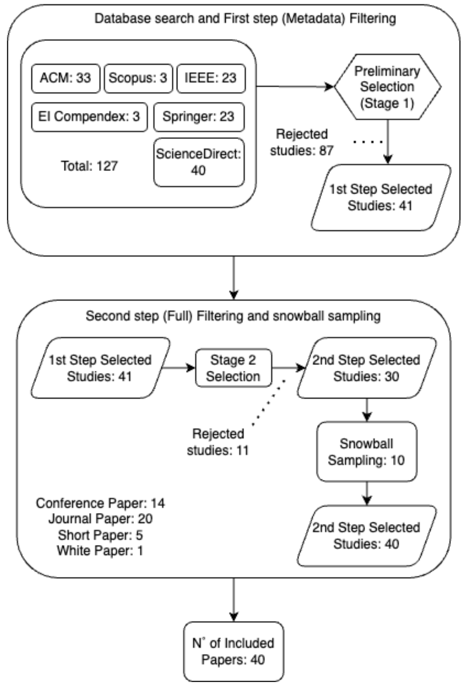


Figure 4: SLR Paper Selection Summary

Table 6: Quality Assessment Index, Field, and Founding Assumption

ID	Quality Assessment Field	Founding Assumption
(QAF1)	Does the model refer to any data regulation act? (E.g., LGPD, GDPR)	Encompassing all the papers that refer to some kind of data regulation act, hence, leveraging its credibility.
(QAF2)	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.
(QAF3)	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.
(QAF4)	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.
(QAF5)	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.

Continued on next page

Table 6 - *Continued from previous page*

ID	Quality Assessment Field	Founding Assumption
(QAF6)	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.
(QAF7)	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.
(QAF8)	Does the paper cover any aspect of legal interoperability?	One of the outcomes of our RSL aims towards the legal interoperability aspect within IDS.
(QAF9)	Does the first author have at least one publication about data sovereignty?	Engagement with the proposed research field enhances reliability.
(QAF10)	Are potential biases or limitations addressed and discussed?	Limitations and difficulties are primordial in a valid work, intrinsically, its credibility.

Moreover, each paper must undergo a data extraction and quality assessment. 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 18 unique 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.

Along with the paper metadata, we proposed an approach to assess the paper's relevance, instead of collecting just the citations, we propose a functional-like equation, based on weights, citations, and year of publication. Let C be the number of citations, Y the publication year, and CS the citations score, we have Equation (1).

$$CS = \frac{Citations}{CurrentYear - Y + 1} \tag{1}$$

By doing so, each citation will be a fairer comparison between papers written in the same year of the SLR (2023), and papers 9 years older (2014), with more time to be cited. Each citation in the same year carries the weight of 1 and each consecutive year, it lows $\frac{1}{2}$. 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[13]. Finally, Figure 5 represents the number of publications over the past 10 years (time constraint of our SLR).

[13] `https://github.com/VictorBenoiston/towards_legal_interoperability_SLR`

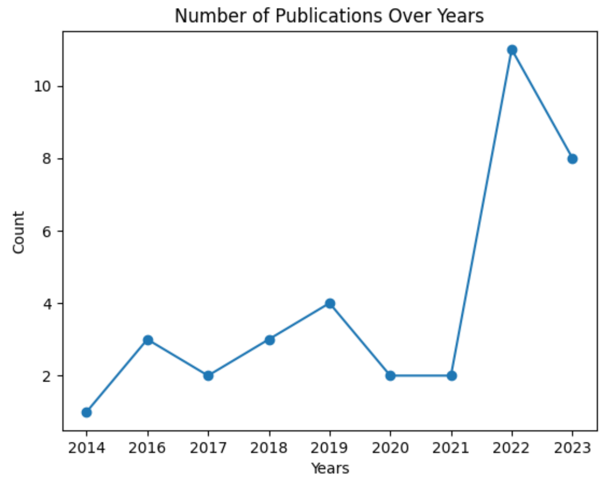


Figure 5: Number of Publications Over the Years

4. Results

As seen in the protocol, we have broadened the inclusion criteria in order to retrieve concepts used not only in IDS itself but also concepts that might be useful to foster legal interoperability within IDS, once previous data spaces are (in theory) more developed. Hence, other data ecosystems such as Industrial Data Spaces (Otto et al., 2021), and Open Data Spaces (Kirstein & Bohlen, 2022) were evaluated. The latter, surprisingly adopted the IDS connectors technology, and other concepts within the IDS RAM. Based on Kitchenham (2004), this section will describe the results obtained after the performed review, with qualitative and quantitative analysis, comparing common findings, and defining similar approaches/techniques. Furthermore, we define the hereby called Retrieved Legal Aspects (RLA) of IDS, and similar data spaces, literature gaps, challenges, opportunities, proposed future works, and interesting findings in the legal sphere of IDS.

4.1. Overview

As cited in section 3.3, 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 summarize their quality assessment scores in the histogram in Figure 6 (Please note, scores with floating numbers are encompassed in the previous integer, e.g., 9.5 appears in bin 9).

The only material graded at 9.5 is the Reference Architecture Model Otto et al. (2019) itself, it fosters all the policies and the architecture for the implementation of IDS. It is leveraging RQ5 (it shows that the interest in the field has grown over the past three years), and providing an overview of the publication years from the retrieved papers. Furthermore, as mentioned, the proposed SLR focuses on the Legal Aspects that foster legal interoperability among the participants, however, it also approaches other types of dataspaces, that use IDS technology or architecture or have similar approaches, due to the lack of research being performed specifically on IDS, and its implementation. An overview is available in Figure 7.

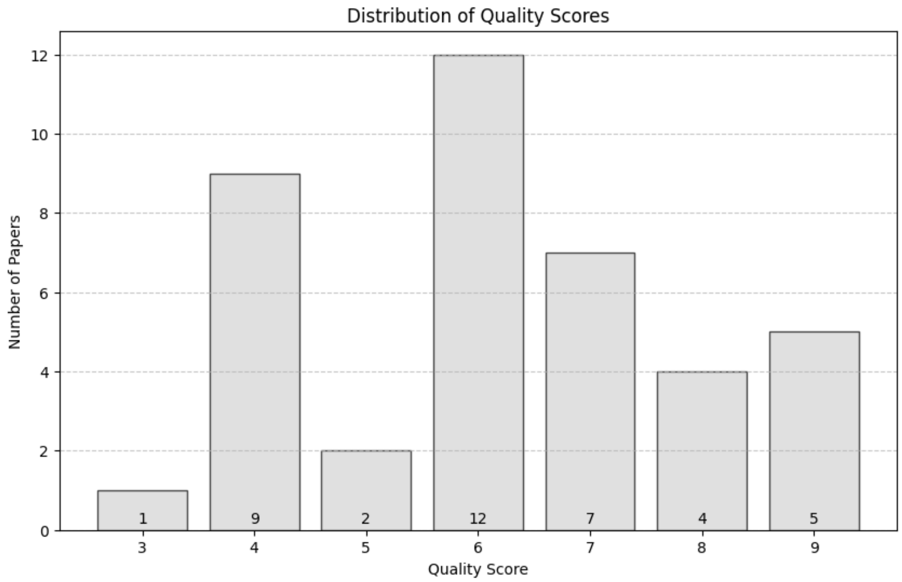


Figure 6: Quality Assessment Summary

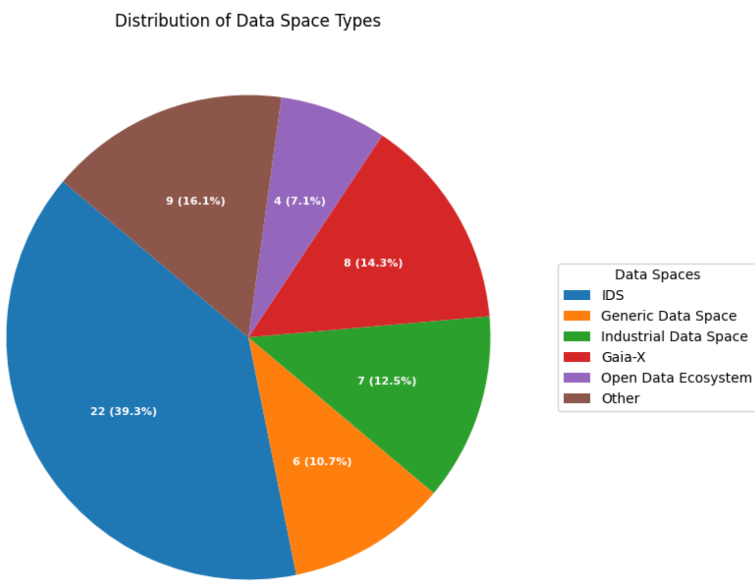


Figure 7: Distribution Of Data Spaces

4.2. Mapping Research Landscape: Venues and Geographical Distribution

One of the assumptions for developing a useful and valid SLR is the twofold meaningfulness, for practitioners and researchers. Leveraging this requirement, we ought to state the venues of publication from the 40 selected papers. Table 7 shows the distribution of Journals that fostered the retrieved papers. It is visible that they are not clusterized, rather, scattered in various venues, but it is possible to infer the relationship with data knowledge, systems modeling, information systems, and business and information fields.

Table 7: Journals and Number of Publications Used in our SLR (Through December 2023)

Journal	**No.**
Digital Communications and Networks	1
Procedia Computer Science	1
Data & Knowledge	2
Software Quality Journal	1
AI & Society	1
Electronic Markets	1
CEAS Space Journal	1
Software and Systems Modeling	1
Intereconomics	1
Information Systems	1
Cooperative Information Systems	1
Business and Information Systems Engineering - BISE	2
Business & Information Systems - Springer	1

Furthermore, we provide an overview of the conferences related to the field, which is available in Table 8.

Table 8: Conferences and Number of Publications Used in our SLR (Through December 2023)

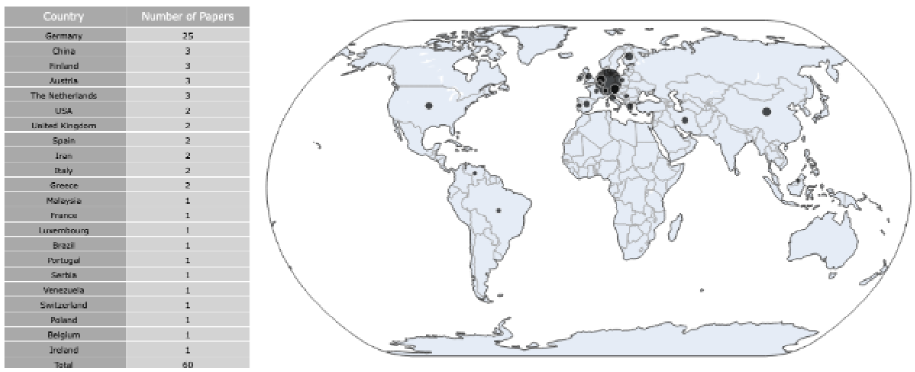
Conference	**No.**
CIRP Life Cycle Engineering Conference.	1
International Symposium on Emerging Information, Communication and Networks (EICN)	1
ACM Symposium on Applied Computing	1
ICIST: International Conference on Information Systems and Technologies	1
IEEE/ACM International Conference on AI Engineering – Software Engineering for AI (CAIN)	1
CoNEXT: The International Conference on emerging Networking EXperiments and Technologies	1
Digital government and solidarity	1
DSAI: International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion	1
WWW: The ACM Web Conference	2
WWW Companion: Companion Proceedings of the ACM Web Conference	1

Continued on next page

Table 8 - *Continued from previous page*

Conference	No.
IEEE International Conference on Emerging Technologies and Factory Automation	1
P2P, Parallel, Grid, Cloud, and Internet Computing (3PGCIC)	2
Conference: IEEE International Conference on Blockchain (Blockchain)	1
CEUR Workshop	1

Moreover, leveraging the answer of RQ3, we describe what countries were involved in the retrieved studies, we included the nationality of the first author's institution (university or company), and in papers addressing case studies (Munoz-Arcentales et al., 2019), (Ahmadian et al., 2018) Altendeitering et al. (2022), (Bigini et al., 2022), (Abramowicz et al., 2016), (Seidel et al., 2023), (Tchoffa et al., 2016), Figure 8 provides the geographical distribution.



Country	Number of Papers
Germany	25
China	3
Finland	3
Austria	3
The Netherlands	3
USA	2
United Kingdom	2
Spain	2
Iran	2
Italy	2
Greece	2
Malaysia	1
France	1
Luxembourg	1
Brazil	1
Portugal	1
Serbia	1
Venezuela	1
Switzerland	1
Poland	1
Belgium	1
Ireland	1
Total	60

Figure 8: Geographical Distribution of Research Facilities, and Case Studies per Country

Along with RAM, and the Information Model[14], IDSA also provides the Data Spaces Radar[15], 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 graph 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.

4.3. Retrieved Legal Aspects (RLA) of IDS and Similar Data Spaces

After thoroughly analyzing the 40 selected papers, and being grounded by Kitchenham (2004), it was possible to observe that some papers addressed the same aspects/approaches. Henceforth, we selected aspects related to our proposed research questions and legal aspects within data spaces. The proposed RLAs were analyzed by the authors, as domain experts. Table 9 summarizes the 10 Retrieved Legal Aspects (RLA) in the proposed SLR. Each topic will be further detailed.

[14] https://github.com/International-Data-Spaces-Association/InformationModel
[15] https://www.dataspaces-radar.org/radar/

Table 9: Retrieved Legal Aspects and its Relevance Towards RQs

ID	RLA	References	RQ1	RQ2	RQ3	RQ4	RQ5
RLA1	Domains of Business	(Munoz-Arcentales et al., 2019) (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) (Scheider et al., 2023) (Otto et al., 2021) (Seidel et al., 2023) (Pettenpohl et al., 2022) (Scerri et al., 2022)	X	X	X		X
RLA3	Usage and Data Policies	(Alexopoulos et al., 2023) (Altendeitering et al., 2022) (Scheider et al., 2023) (Bigini et al., 2022) (Rezaei et al., 2014) (Griffo et al., 2021) (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., 2021) (Bader et al., 2020) (Tardieu, 2022)	X	X	X	X	X

Continued on next page

Table 9 - *Continued from previous page*

ID	RLA	References	RQ1	RQ2	RQ3	RQ4	RQ5
RLA5	Smart contracts / Contract automation	(Weichhart & Naudet, 2014) (Pullmann et al., 2017) (Bigini et al., 2022) (Griffo et al., 2021) (Bader et al., 2020) (Duisberg, 2022) (Tardieu, 2022)		X	X	X	X
RLA6	Semantic Appeal	(Rezaei et al., 2014) (Janev et al., 2021) (Theissen-Lipp et al., 2023) (Meckler et al., 2023) (Bader et al., 2020) (Pullmann et al., 2017) (Abramowicz et al., 2016) (Immonen et al., 2018) (Jesse, 2018) (Firdausy et al., 2022)	X	X	X		
RLA7	AI usage in IDS	(Farahani & Monsefi, 2023) (Altendeitering et al., 2022) (Janev et al., 2021) (Solmaz et al., 2022) (Theissen-Lipp et al., 2023) (Hecker et al., 2022) (Kirstein & Bohlen, 2022) (Nagel & Lycklama, 2022) (Tardieu, 2022)		X		X	X
RLA8	Cloud	(Farahani & Monsefi, 2023) (Rezaei et al., 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 et al., 2016) (Immonen et al., 2018) (Kirstein & Bohlen, 2022) (Tardieu, 2022)	X		X	X	X

Continued on next page

Table 9 - *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., 2019) (Silva et al., 2023) (Weichhart & Naudet, 2014) (Sang et al., 2021) (Theissen-Lipp et al., 2023) (Meckler et al., 2023) (Bigini et al., 2022) (Abramowicz et al., 2016) (Seidel et al., 2023) (Firdausy et al., 2022) (Griffo et al., 2021) (Duisberg, 2022) (Kirstein & Bohlen, 2022)			X	X	X

4.3.1. Domains of Business

The Data Spaces Radar provides the current case of the application of IDS worldwide, crossing with the retrieved information, we may compose a high-level classification of domains as Industry, Academic, Governmental, Foundational, and Legal. Figure 9 summarizes the distribution of the domains, and further references are available in the provided GitHub Repository.

As for the classification, Industry refers to robust papers (Journal papers), addressing some validation (e.g., study case), and white papers. The Academic domain refers to ongoing research, lacking validation or no use of stakeholders. Governmental papers address the governmental layer, providing solutions and/or theories to foster data governance worldwide, through legislation or policies. Finally, Foundational papers leverage high-level concepts, such as validated ontologies or theories, such as the Information Model (Bader et al., 2020). After analyzing the 40 papers, it is possible to conclude that fields such as health, in which, Bigini et al. (2022) point out the intrinsic need for individuals' capability to own control over data assets, and companies must comply with the legislation (connects with RQ1 by providing a short definition of legal interoperability). Also, papers such (Breidenbach et al., 2023) and (Theissen-Lipp et al., 2023) address issues such as data privacy, and compliance with data usage policies. Furthermore, Cloud services also tackle legal interoperability awareness, Farahani & Monsefi (2023) and Tardieu (2022) propose the compliance for cloud services providers with the agreed policies for infrastructure and applications, before exchanging any data. The Information Model itself (Bader et al., 2020) also leverages the priority of data sovereignty of its participants, and points out cloud computing and the specifications on the formal usage restrictions, leveraging a better understanding of policies, and addressing machine-readable contracts (fostering RQ3 by addressing the current representation of policies and pointing to the lack of machine-readable policies). Other fields such as retail (Pinto et al., 2023), smart cities (Sang et al., 2021), and energy (Altendeitering et al., 2022) were also noticed.

The papers classified as Governmental, such as Klug & Prinz (2023), propose fairer principles of data exchange, and also allude to the implementation of smart contracts, whereas Tardieu (2022) and Scerri et al. (2022) make an in-depth study of how the European data spaces achieve interoper-

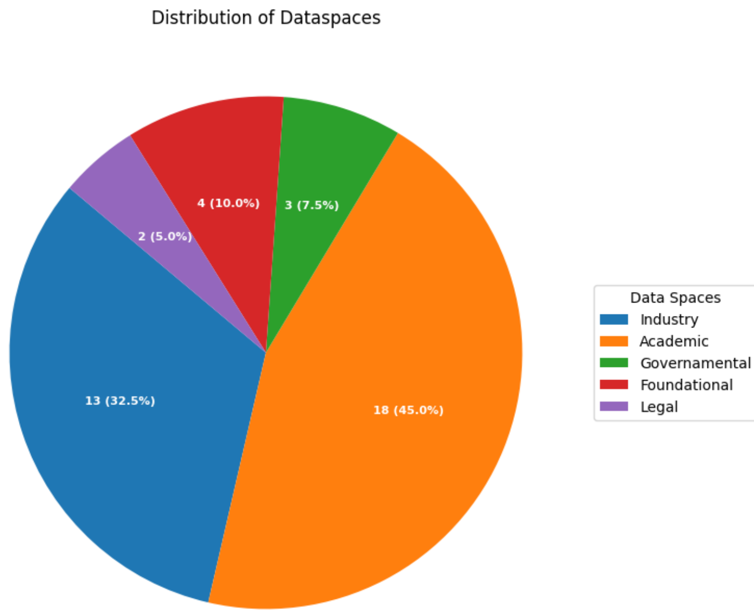


Figure 9: Distribution of Domains

ability, focusing in our domain, the papers propose a common set of policy and rules providing not only interoperability but portability as well. It is important to notice that Gaia-X uses IDS RAM to ensure usage controls and compliance assurance. Furthermore, a specific domain Finally, Legal papers such as (Griffo et al., 2021), propose a service contract ontology, that, although not clearly stated for IDS, provides important terms such as Data Provider and Data Consumer, which relate to the IDS architecture, hence, enabling its adoption. It thoroughly defines concepts such as legal entitlements, legal burdens, and legal agreements. Whereas Duisberg (2022) explicitly manages legal aspects of IDS itself, such as contract negotiation, legal compliance, and alignment with the legislation (addressing RQ1 by fostering the concept of legal interoperability, and RQ2 by linking data sovereignty to legal interoperability). Concepts retrieved based on this RLA were important to help answer RQ1, RQ2, and RQ3 by providing the current research conducted in the high-level approach and detailing a few observed fields of industry.

4.3.2. Personal/Non-Personal Data

The Reference Architecture Model refers almost solely to exchanging non-personal data, once it works with exchanging company and machine data. However, Meckler et al. (2023) propose a so-called Solid project, which resembles Gaia-X and IDS technology to exchange data in a trust and sovereign environment, focusing on personal data control, whereas IDS focuses on non-personal data for industry use cases. Hence, IDS may be used in a wider range, exploiting the industry field. As seen in (Tardieu, 2022), (Duisberg, 2022), and even in the Information Model, there is indeed a lack of privacy warranty regarding personal data in IDS architecture, such as the erasing of personal data after use, or data usage expiration date, as defined the Art. 5 of the General Data Protection Regulation (GDPR)[16], and pointed out by Tardieu (2022). The data must be used based on the data usage policy, as authorized in the usage policy. Although the IM proposes a model-based privacy, enabling the verification of the system that processes personal data supporting the service customers, it is still a glimpse of future work. Furthermore, Ahmadian et al. (2018) propose a privacy analysis for the previous Industrial Data Space. The results have shown that failures to protect personal data affect the data providers negatively, harming their reputations, and causing financial damages. IDS inherits its precursor's lack of consideration for private data and compliance with GDPR.

According to GDPR, personal data means 'any information relating to an identified or identifiable natural an identified or identifiable natural person; an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification, number, location data, on an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. Furthermore, Duisberg (2022) points out that the rise of Big Data and AI has shown that at a given point, an anonymous asset or other data with no connection or relevance to natural persons (as stated in GDPR), may turn out to be an element of personal data, if combined with other identifiers. Of the 40 analyzed papers, 21 propose personal data usage/exchange, hence, even though the SLR aggregates other kinds of Data Spaces, we might associate the personal data usage within IDS, fostering the RQ2, by exposing the lack of policies for personal data in IDS, employing gaps, opportunities and challenges within RQ4 and RQ5.

4.3.3. Usage and Data Policies

The IDS Reference Architecture Model proposes a data usage policy and states that the data consumer must comply with the data owner's usage policy to use the data Otto et al. (2019). The data usage policy should be available through the service contract (detailed in further sections). Each policy describes the permissions and obligations of an IDS resource. Furthermore, Munoz-Arcentales et al. (2019) relate the lack of attention within the usage policies, and its further enforcement and control, and propose an architecture adding XACML components such as Policy Execution Point (PXP), Policy Decision Point (PDP), and Policy Administration Point (PAP). The Information Model ontology and IDS architecture set the standard for machine-readable policies by addressing the Open Digital Rights Language (ODRL)[17].

The ODRL defines a rule as an abstract concept linking some action by an individual over an asset, and those rules may be rules of permission, duty, or prohibition. Tardieu (2022) identified the complexity of regulations employed on the data policies and usage policies, however, a deeper understanding of how data regulations interplay within data platforms (legal interoperability), allowing the dynamization of usage and data policies, the observed challenges will be tackled in the next topic (4.3.4). A few approaches are proposed to facilitate the understanding and application of usage and data policies, such as in Bader et al. (2020), with the concept of the minimum viable set (MVS) of policy rules (not yet adopted by IDS).

Additionally, Bader & Maleshkova (2019) cite a current deficiency of formal descriptions for permissions and obligations in data usage, used to build the policies. Furthermore, the author says 'The IDS IM further details these constructs and defines their implications, focusing on their

[16] https://gdpr-info.eu/
[17] https://www.w3.org/TR/odrl-model/

publication, negotiation, acknowledgment, and enforcement.', however, the constructs available in the IM, proposed by Bader et al. (2020), are top-level classes, with no relations and no nested components. Hence, it becomes a shallow taxonomy, with low classification and generalization power, e.g., 'Usage policy class', has no connections or nested classes, and the model does not address the policy negotiation step, only states the top-level class. The present RLA directly answers the RQ3, and points out future works and gaps, fostering RQ5.

4.3.4. Legal Inteorperability Constraints

The European Interoperability Framework defines legal interoperability as the capability of organizations under different legal frameworks, policies, and strategies to work together. It requires a free flow of data, not being blocked by legislation, in a European scenario, it also encompasses the integration among public services and member states, and defining clear arrangements about how to deal with different legislations across borders, even allowing the creation of new legislation (complied by all involved actors). Scerri et al. (2022) propose the concept of legal compliance challenges, and the risk of navigating around legal constraints due to potential data policy breaches, leading to not compliance with GDPR (for European countries), and exposure of sensitive and/or personal data.

The first legal compliance challenge is data protection, based on the absence of understanding of how data platforms connect within data regulations, the average citizen ought to realize their rights over data, and its implications, therefore, a dynamic use awareness, and better guidance from regulators and data platform developers. Another challenge is the free-flowing data, data ownership persists as a problem, surrounded by access, and portability directly aggravated by the AI context (Bader & Maleshkova, 2019). Data ownership is not a trivial concept within the data market setting. The third legal challenge is preserving privacy, although IDS architecture promises a trustful environment for industrial purposes, its technology may be used in other applications, such as Open Data Spaces (Immonen et al., 2018).

Moreover, Bader & Maleshkova (2019) point out the lack of a comprehensive end-to-end usage control framework (which, would foster a quicker integration of legal interoperability), however, research is being conducted in the machine-readable sphere, and later, the author proposes the Information Model, shortly addressing a few of the approached topics, in a shallow manner. The authors also provide the need for the unambiguity of policies and contracts, guiding to the next legal constraint, the lack of formulation and implementing tools, with no space for no particular definitions, but a clear set of policies. The lack of knowledge of opposite parties (usually, the data provider defines the usage police offer), leads to the need for a semantic representation, fostering the RQ1 and RQ4.

Further, Rezaei et al. (2014) conceptualize the definition of Rules Interoperability, which stands for 'the ability of entities to align and match their business and legal rules for conducting legitimate automated transactions that are also compatible with the internal business operation rules of each other. Consequently, it is reasonable to compare it to legal interoperability, once the authors propose a further division of concepts into rules modeling, alluding to the documentation of the development of expandable and re-usable models to represent rules - a clear link to ODRL. Business and legal rules homogenization/alignment eagers to derivate an alignment of different legislations (of different territories), enabling legitimate conduct and transaction - we might refer to the concept of interoperability checks and interoperability barriers provided by the EIF. Finally, rules execution addresses the dynamic processing and handling of rules, complying with both legislation and business rules, and embracing business, processes and transactions. The aforementioned concepts lead to the answer of RQ1 and foster RQ2 and RQ3.

4.3.5. Smart contracts / Contract Automation

IDSA provides the concept of Task Force Legal[18] as a working group that supports legal situations within IDS. Hence, it takes care of the contractual analysis, agreements, and enforcement, and it also defines the legal aspects of the Reference Architecture Model. Therefore, it has a high implicit cost, hence, Bader et al. (2020) propose as a future work on the Information Model ontology, the further usability of semantics (mainly ontologies), to use machine-readable policies to foster possible automation of contracts and negotiation. It is observed in the literature, a request by the researchers and actors of IDS to a possible automation of service contract negotiation, agreement, and enforcement. Sang et al. (2021) links the possibility of smart contracts to blockchain, due to its traceability, and the authors emphasize the condition checks for different business rules.

Furthermore, Scerri et al. (2022) mention the domain of data as evidence for policy compliance, by exploiting the increasing phenomenon of a data-driven society, on which, standard compliance mechanisms are challenged due to the increasing velocity, complexity, and volume of data, while regulation cyber-physical systems (such as IDS). With the current technological advances, especially IoT (Farahani & Monsefi, 2023), data can serve as evidence for validating whether specific policy-related conditions are met. Data Markets are leading the development of machine-readable policies to leverage the implementation of smart contracts. Duisberg (2022) provides a wide view of the legal aspects, and it tackles the need for further development of automation in contract execution, not only in the conclusion (compliance), but in its performance, enforcement, and negotiation. The authors also indicate that the current scenario (through December of 2023) still sets the scene for such implementation. However, the group Plattform Industrie 4.0 is performing research on the scenario of contract automation, by simulating legal contract executions, and the same team has created the so-called 'terms of use for an Industrie 4.0 platform', these terms of Use were developed aiming toward the definition of a reasonable balance between the interest of the platform operator, and the users (data consumer and data provider), facilitating data transactions. Nevertheless, the legal testbed is not yet available to the community.

Currently, the IDSA has two model contracts[19], which consist of two basic templates that cover data purchases and data as a service type of licensing transactions, designed within the German rules domain, terms, and conditions. The intention of providing those templates is not about prescription, but rather to endorse the overarching principle of freedom of contract, and an attempt to reduce the legal costs addressed to contract formulation and negotiation. Finally, supporting the RQ1, RQ2, and RQ4, it is possible to identify the concepts of data lending, data as a service, data trade, and data purchase.

4.3.6. Semantic Appeal

Such RLA relates to noticing the formal description of knowledge, which has become a major barrier for all kinds of interoperability, but mostly affects the legal interoperability layer, once it is a subjective layer and one of the identified barriers in previous sections, is the lack of unified representation of domains, breaching new interpretations, and dificulting processes such as the automation, and reasoning. Pullmann et al. (2017) address a semantic limitation regarding architectures of generic data spaces, such as the limitation of the architecture to the functional, implementation, and business, however, they overlook targeting high-level aspects such as legal, and governmental.

[18] `https://internationaldataspaces.org/lets-talk-about-idsas-task-force-legal/`

[19] `https://github.com/International-Data-Spaces-Association/IDS-G/blob/main/UsageControl/Contract/README.md`

The same applies to IDS, with shallow classes in the IM ontology, addressing topics such as legal agreement and contracts in a taxonomical behavior, with no connections or relationships.

A semantical approach based on Resource Description Framework (RDF) `https://www.w3.org/RDF/` or Ontology Web Language (OWL)[20] allows the expression of high-level concepts in a machine-readable manner. Bader & Maleshkova (2019) emphasize the machine-readable language ODRL, later, building the IM following its principles, even though, in a superficial sphere. Solmaz et al. (2022) exposes an additional approach to the semantical field, which states an old challenge within the IDS field, the concern regarding losing its control over data, due to its competitive advantage. Hence, the authors propose a data integration at a semantic level, based on an ontology schema, sharing sensitive data indirectly (ontologies describe domains with metadata, fostering the RQ2), and then fully integrating it. The IDS architecture already integrates a resembling architecture, on which the metadata is exchanged by the connectors, on each edge (data provider and data consumer), and analyzed by a broker, aligned with Firdausy et al. (2022). The data broker is responsible for managing the data exchange metadata, hence, not providing the final raw data before the legal agreement.

Moreover, Sardis et al. (2013) provide challenges, such as data heterogeneities, different security specifications and requirements, and the lack of semantics regarding the applications and services. All problems might mislead common definitions and policies, making it difficult for the software engineers and developers to automate such processes, and for systems to perform inferences and predictions - as previously stated, researchers and developers must comply with definitions, and policies must be unambiguous. The authors also specify that ontologies are the cornerstone of semantic web technology, allowing the creation of structured vocabularies for a formal specification of domains.Theissen-Lipp et al. (2023) highlight the implementation of ontologies on data spaces, using them as the foundation for knowledge graphs, sharing a common understanding of data. Then, the authors define important roles in the data spaces initiative, such as science, by providing trustworthy research evidence (especially on semantics and reusability), implementation, by providing data spaces for a broader population, and finally, the role of political bodies including the forming of mutual view on common aspects. Finally, the concept of further development of semantic cohesion in data spaces over time, by providing new semantics mapping will lead to a better adoption worldwide (as the foundation of IDS), relying on specific sets of standards. Therefore, essentially the majority of the authors propose a semantic approach in order to foster a unique understanding of policies and contracts. Hence, FAIR ontologies for describing the domain of IDS as seen in Bader et al. (2020), and aggregate value to the Information Model are pointed as good solutions. The retrieved concepts address not only RQ1, RQ3, RQ4, and RQ5, by defining a unambiguous definition, and further proposed works.

4.3.7. AI Usage in IDS

Another retrieved aspect is using IDS technology to foster AI applications, or conversely. Janev et al. (2021) propose that the usage of knowledge graphs for building data connectors would facilitate the integration, and improve the explainability and generalization capability of machine learning models. Solmaz et al. (2022) also propose that machine learning models would benefit from an easier setup, better prediction, and alignment with legislation, to improve the generalization and interoperability of models. This interoperability would also lead to portable and replicable solu-

[20] `https://www.w3.org/OWL/`

tions (fostering the concepts of ontology usage for unambiguous definitions). Furthermore, Hecker et al. (2022) emphasize the data heterogeneity among companies, leveraging the federated learning approach. The authors also point out the importance of using the IDS architecture to foster a swift implementation of federated models, on which the companies do not necessarily exchange data, rather, they exchange the metadata to train the models, and then, the final steps are performed with their own data. Hence, IDS technology, such as data brokers and data connectors leverages the collaboration of different actors and different roles with a common goal, of creating value over data (Theissen-Lipp et al., 2023).

Farahani & Monsefi (2023) mention the usage of IDS technology in Gaia-X. Gaia-X also provides a reference architecture, with a federated/cloud approach, hence, it resembles IDS, yet, uses IDS technology to exchange data in a trustful and sovereign way. Even though Gaia-X and IDS may lead to a balance between the creation of value for data, compliance, and privacy regulations, those technologies are not designed to tackle the current gap between privacy and AI. Further, they do address remote access to data (cloud), distributed data governance, and access control, however, they are not sufficient to erase barriers regarding privacy, legal obstacles, policies, and conflict of interest. It all relates to the central challenge of using those data spaces, i.e., the concern of losing a competitive edge by revealing industry secrets or sensitive data. Other works such as Munoz-Arcentales et al. (2019) and Pinto et al. (2023) also tackle the usage of AI, and decision-making, resembling legal aspects. Therefore, this RLA enables the answer to questions such as RQ2, RQ3, RQ4, and RQ5 by addressing the usage of IDS technology and stating its outcome.

4.3.8. Cloud

Along with Industry 4.0, the migration to cloud environments to reduce costs is a reality. It is impossible to tackle the cloud sphere without mentioning Gaia-X (Otto et al., 2021). Gaia-X is a project, started in 2019 by the German and French Governments of Economy, to ensure data sovereignty over data, when hosted by non-EU cloud service providers (Tardieu, 2022). Bader et al. (2020) emphasize the developments in cloud computing, and formal usage restrictions, leading to a new identification, and description of resources, promoting the requirement of machine-readable formal contracts, and embedding them into the IDS infrastructure. As stated in previous sections, the Gaia-X policy rules committee has developed a minimum viable set (MVS) of policy and rules, advancing the regularization of cloud environments for data exchange.Tardieu (2022) indicates that the main concerns preventing companies from sharing industrial data are often referred to as security and data usage control. The author proposes that one way of resolving such a challenge is for cloud service providers to demonstrably comply with the agreed set of policies for infrastructure and applications, leveraging the traceability and free flow of data. A common set of policy rules and specifications is also proposed, harnessing portability and interoperability.

As aforementioned, IDS architecture and technology are being used beyond its initial boundaries. Gaia-X imposes IDS technology, in order to maintain sovereignty, trust, and ensure data usage controls and compliance. Seidel et al. (2023) state the main services available in Gaia-X, such as transparent and verifiable descriptions, compared to data protection, data security, and compliance with technical and legal requirements. The eligibility for compliance with legislations such as GDPR, EU requirements, etc., accredits the initiative in terms of trustworthiness. Furthermore, other topics approached were identity and trust services, federated catalog, data sovereignty services, and compliance. Promoting Gaia-X cloud services and liabilities. Solmaz et al. (2022) highlight the focus of Gaia-X on comprehending the requirements for cloud services and data exchange in a particular matter, in order to standardize the vocabulary encompassing such elements in formal rules and policies. Hence, contributing to the control of participants, and creating a balanced holistic environment for cloud service providers, and consumers. However, Gaia-X focuses on the meta-level, describing the federated services, participants and data to be exchanged, whereas the data level itself is comprised of IDS architecture. Summarizing, both initiatives (Gaia-X and IDS) provide rules, frameworks, and architectures for enabling trustful and data sovereign exchanging, but with no proper definition of data models and detailed interactions needed to achieve complete interoperability.

Otto et al. (2021) point out that Gaia-X is not at the same level of maturity as IDS, still, it aligns with the vision of upholding data sovereignty, value creation, and a trustful data ecosystem environment. Research conducted in Theissen-Lipp et al. (2023) and Meckler et al. (2023) also proposes a higher level of maturity in IDS. Grounded by those common goals, Gaia-X and IDS complement each other to ensure cloud data sovereignty for end-to-end data value in federated ecosystems. Regarding the collaboration between Gaia-X and IDS, Theissen-Lipp et al. (2023) define IDS RAM as the main building block of Gaia-X. Furthermore, Meckler et al. (2023), propose that although they complement each other, IDS provides sophisticated concepts for dataspaces, regarding the proper raw data exchange, whereas Gaia-X architecture provides solutions for data storage, cloud components, and federated services. Nevertheless, one of the fundamental challenges of Gaia-X is also the legal interoperability capability. The proposed RLA allows us to tackle future opportunities for using IDS, and IDS technology, such as the Cloud environment (RQ5), and pointing out challenges of implementing legal interoperability in such domains, even though, Gaia-X proposes a better alignment with legislations (RQ4)

4.3.9. Usage of IDS in Open Spaces

The sections above pointed out the usage of IDS, and IDS technologies to promote different data spaces, and through this SLR, it was possible to observe, the rising usage of IDS for Open Spaces, as vastly proposed in (Abramowicz et al., 2016), (Immonen et al., 2018), (Kirstein & Bohlen, 2022), and (Tardieu, 2022). (Meckler et al., 2023) showed that IDS technology is being used in the open-data ecosystems, and it proposes that companies utilizing IDS to share industrial data could improve data analysis by consuming open data, and even publishing non-confidential data under an open license. Technically, exchanging open data is not different from exchanging confidential data, enabling the usage of IDS to distribute both, closed and open data. The IDS data broker and open data portals act as gateways with similar functionalities, providing information (meta-data) about the available data. In addition, the open data initiative constantly seeks practical and resource-efficient methods to publish their data, such technologies are incorporated by IDS architecture, allowing the publisher to choose which metadata broker to register to. Further, providing data sovereignty to the data provider, however, the control over data concludes once downloaded, given there are currently no usage controls applied to open data.

Notwithstanding the usage of open data for the creation of value, Immonen et al. (2018) demonstrated that open data upholds challenges such as an arduous task, once the data needed is not entirely available (due to industrial secrets), or is attached to a usage price (classifying a data rent, as viewed in section XX), and most importantly, there is uncertainty about the data quality. However, the authors also demonstrated the same vision about the possible advantages of collaboration among different data ecosystems, as collaborative innovation, and value co-creation, acting as joint controllers (Scerri et al., 2022). The authors share the same point of view in (Meckler et al., 2023) tracing the upsides of using IDS technology in open spaces, such as the efficiency in value co-creation and lowering the threshold for using and publishing open data. Finally, it may lower the threshold for companies' adoption of IDS, acting as a catalyst.

Furthermore, Scerri et al. (2022) propose several opportunities for business, such as the enhancement of the current contractual-binding agreements, with a complementary adoption of a large-scale marketplace whose participation is open to all kinds of data producers and consumers. This way, enabling the sharing of industrial data in and out of value networking, while guaranteeing data sovereignty. Other opportunities are the increased availability of vast and heterogeneous data ecosystems for AI, Innovative and data-driven business models enabled by new value ecosystems, and opportunities to tap into safe personal data. This RLA partially answers RQ2, by addressing the concept of data sovereignty linked to the concept of legal interoperability. Further, it also enhances the understanding of the legal challenges in IDS, and other dataspaces that use IDS architecture, alluding to RQ4.

4.3.10. Future Works Addressing Legal Aspects

Analyzing the future works established by the proposed 40 papers is a clear methodology to determine the current gaps in the literature. In order to summarize our findings, Figure 10 presents the distribution of future works proposed by the analyzed literature. An in-depth presentation is available in the previously mentioned GitHub repository.

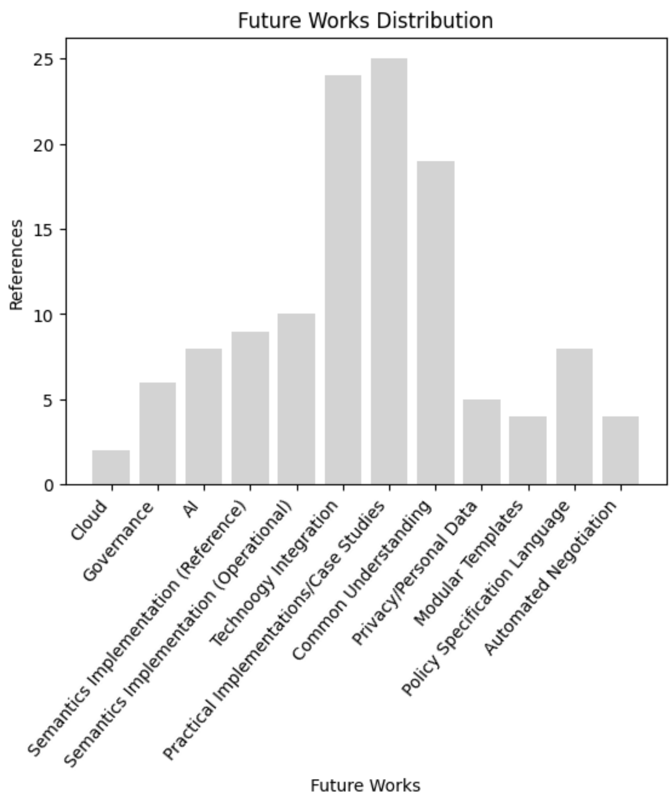


Figure 10: Future Works Distribution

The first kind of future work addressed by the literature is the implementation of semantics. As observed, semantics is usually addressed by an ontological approach, leading to knowledge graphs, or feeding AI models. We might split this kind of implementation into referential, on which the designer proposes a graphical approach, such as the development of ontologies in OntoUML[21], plain UML, or other technologies. Whereas, operational semantics, would be developed in RDF, and for ontologies, further translated to OWL. Bader et al. (2020) propose the so-called information model, which shallowly comprises the main architecture of IDS, and states as future work the further improvement of it, in a referential and operational scope. Further, Altendeitering et al. (2022) present the creation of an information hub, composed of knowledge graphs to facilitate the integration and provide explainability of machine learning services. In their work, Meckler et al. (2023) propose a balance between enough complexity to guarantee dataspace features (i.e., trust and data sovereignty), while leading with a low threshold for integration. The authors state the necessity of a clear domain description by fostering the linked data platform. Pullmann et al. (2017) suggest the modeling of mappings, defining context, and providing description through semantic models. We also recognized the dependency of referential and operational models. Usually, when research conducts referential models, such as architectures, reference ontologies, or conceptual models, future works are stated as the translation to RDF/OWL. Examples are (Solmaz et al., 2022), (Klug & Prinz, 2023), (Firdausy et al., 2022), and (Griffo et al., 2021).

The concepts of Technology Integration and Practical Implementation usually are mentioned jointly. Technology integration means developing active tools, and interfaces, extending the infrastructure, or combining different technologies. At the same time, practical implementation explores the usage of stakeholders, or even final users, conduction of a trial experiment or empirical studies, and especially case studies. However, a set of studies propose technology integration with no practical implementation, such as (Theissen-Lipp et al., 2023), (Meckler et al., 2023), (Bader et al., 2020), and (Firdausy et al., 2022). Studies that propose practical implementation with no technology integration, refer to use cases for validation of already established technology, i.e., (Ahmadian et al., 2018), (Sardis et al., 2013), (Scheider et al., 2023), and (Tardieu, 2022). Another important topic for our SLR is based on governance, several studies propose further research regarding governance. Klug & Prinz (2023) point out the lack of details on the governance framework, whereas Nagel & Lycklama (2022) propose a governance structuration and definition of operational processes towards such governance.

As for AI usage, authors mostly propose the regulation of AI based on dataspaces, fostering the implementation of data exchange (in a sovereign and trustful environment), while upholding personal (as mentioned in section 4.3.7, machine learning models have the potential of using non-personal data to generate or trace personal data) and industrial data protection, and clear privacy policies. Altendeitering et al. (2022) also propose the improvement and explainability of machine learning services and analytical applications. Solmaz et al. (2022) point out the data value regarding data spaces by addressing the time-consuming process of data scientists manually extracting, analyzing, and selecting the features. Studies conducted in (Bader & Maleshkova, 2019), (Hecker et al., 2022), and (Scerri et al., 2022) also propose the usage of IDS or data spaces for exchanging data fostering further usage in machine learning models. Several papers tackled a lack of common understanding among companies, industries, users, and government. Theissen-Lipp et al. (2023) state that semantic models that use domain vocabularies may lead to a common understanding between data providers and consumers. Bader & Maleshkova (2019) propose that policies and

[21] `https://ontouml.org/`

contracts must be shared and unambiguously understood, pointing to future work in formulating particular vocabularies for the usage control domain.

Through this SLR, we establish that indeed, there is a current lack of research regarding privacy and personal data handling. Although the IDS initiative aims at the exchange of industrial data (i.e., non-personal), different data spaces are using IDS technologies to exchange personal data, as pointed out in section 4.3.9. As a European initiative, the standard architecture must comply with the GDPR. For instance, dataspaces that rely on exchanging health data, such as observed in (Theissen-Lipp et al., 2023) and (Bigini et al., 2022). Developing modular templates (i.e., with high generalization potential) was also proposed as future work, as seen in (Tchoffa et al., 2016), which manage the concept of modular blueprints, with multi-scale capabilities. Other works such as (Solmaz et al., 2022), and (Spiekermann, 2019) also propose flexible models. Another interesting point observed is cloud computing.

Regarding cloud computing, Gaia-X leads the path, fostering an architecture for federated cloud services, (Tardieu, 2022). IDS architecture does not comprise the cloud environment, but it proposes the decentralized approach, reflected in federated services, in which, no central database exists. Gaia-X proposes using a few IDS technologies such as IDS connectors, and data broker. Seidel et al. (2023) propose a use-case for the space sector with a plan-do-check-act roadmap for implementing Gaia-X architecture in hardware, which could be easily transferred to other domains. Alexopoulos et al. (2023) also foster the usage of cloud virtualization, as a way to combine fast components and blueprint templates, to optimize the place and size required to store data.

Numerous papers suggest a specification language, such as Munoz-Arcentales et al. (2019), which define the future work of their research as a definition of a policy specification language, based on ODRL. Weichhart & Naudet (2014) cite a domain-specific language to support core components and their decentralized nature. This characteristic is also reproduced in (Theissen-Lipp et al., 2023), (Griffo et al., 2021), and (Bader & Maleshkova, 2019). Finally, we reach the automated negotiation, which is addressed in (Weichhart & Naudet, 2014) as the capability of agents to be automatically selected, leading to a federated infrastructure. Bader et al. (2020) define as next step as the development of tools for automated extraction of the IM metadata from different data sources, which is conditional to the topic of a common understanding of knowledge. Duisberg (2022) states the development of the so-called legal test bed[22], which would be able to perform contract negotiation automation. However, it is still a future work. This RLA showcases a clear connection with RQ5 by proposing the ongoing research and future directions.

5. 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. As observed in section 4.1., the proposed papers had an average of 5.5 on the quality assessment. The overall results indicate that the legal aspects of data spaces, specifically IDS 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

[22] https://legaltestbed.org/en/start/

representing policies and rules among companies while respecting the so-called state sovereignty (for a detailed answer, refer to topics 4.3.1, 4.3.4,4.3.5, and 4.3.6). 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 (a fine-grained answer is proposed throughout the topics 4.3.1, 4.3.2, 4.3.4 - 4.3.7, and 4.3.9). 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 (please refer to topics 4.3.1, 4.3.3, 4.3.4, 4.3.6, and 4.3.7).

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 (sections 4.3.2, 4.3.4 - 4.3.9 present an in-depth approach). 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 (different future work, and specific gaps may be found in sections 4.3.2, 4.3.3, 4.3.6, 4.3.7, 4.3.8, and especially 4.3.10). 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.

6. Conclusion

It is possible to conclude, that several studies have shown that one of the proper ways to tackle legal interoperability, is through semantic representation. The semantic representation, usually comprised of ontologies, could unambiguously define high-level concepts while working with metadata (fostering data ownership). Additionally, not sharing industry secrets is related to the review as one of the main concerns of companies in sharing their data), and connecting concepts through relationships. One of the main works pointed out by our review, is the Information Model, proposed by Bader & Maleshkova (2019), however, such approaches, have failed to address the legal aspects regarding legal compliance, legal agreements, policies representation, and contractual arrangements. Finally, IDS is an ongoing project, hence, research is being conducted to reach its prime, and with this review, we may present that although legal aspects (leading to legal interoperability) are foundational, they still do not attain the attention of researchers.

Acknowledgements

The present work was endorsed by CNPq - National Council for Scientific and Technological Development - Brazil.

References

Abramowicz, W., Auer, S., & Heath, T. (2016). Linked Data in Business. *Business & Information Systems Engineering*, *58*, 323–326. URL: https://doi.org/10.1007/s12599-016-0446-0. doi:10.1007/s12599-016-0446-0.

Ahmadian, A. S., Jürjens, J., & Strüber, D. (2018). 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* SAC '18 (pp. 1142–1149). New York, NY, USA: Association for Computing Machinery. URL: https://dl.acm.org/doi/10.1145/3167132.3167256. doi:10.1145/3167132.3167256.

Alexopoulos, K., Weber, M., Trautner, T., Manns, M., Nikolakis, N., Weigold, M., & Engel, B. (2023). An industrial data-spaces framework for resilient manufacturing value chains. *Procedia CIRP*, *116*, 299–304. doi:10.1016/j.procir.2023.02.051.

Ali, N. B., & Petersen, K. (2014). Evaluating strategies for study selection in systematic literature studies. In *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement* ESEM '14. New York, NY, USA: Association for Computing Machinery. URL: https://doi.org/10.1145/2652524.2652557. doi:10.1145/2652524.2652557.

Altendeitering, M., Pampus, J., Larrinaga, F., Legaristi, J., & Howar, F. (2022). 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* CAIN '22 (pp. 193–204). New York, NY, USA: Association for Computing Machinery. URL: https://dl.acm.org/doi/10.1145/3522664.3528593. doi:10.1145/3522664.3528593.

Aromataris, E., & Riitano, D. (2014). Constructing a search strategy and searching for evidence. a guide to the literature search for a systematic review. *The American journal of nursing*, *114*, 49–56. doi:10.1097/01.NAJ.0000446779.99522.f6.

Ayres, R. U. (1996). Creating industrial ecosystems: a viable management strategy? *International Journal of Technology Management*, *12*, 608–624.

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. (2020). The International Data Spaces Information Model – An Ontology for Sovereign Exchange of Digital Content. (pp. 176–192). doi:10.1007/978-3-030-62466-8_12.

Bader, S. R., & Maleshkova, M. (2019). Towards Enforceable Usage Policies for Industry 4.0. URL: https://www.semanticscholar.org/paper/Towards-Enforceable-Usage-Policies-for-Industry-4.0-Bader-Maleshkova/ed7e2658946e35c5a43bb83253ad3a09f28e3315.

Bigini, G., Zichichi, M., Lattanzi, E., Ferretti, S., & D'Angelo, G. (2022). Decentralized Health Data Distribution: A DLT-based Architecture for Data Protection. In *2022 IEEE International Conference on Blockchain (Blockchain)* (pp. 97–104). URL: https://ieeexplore.ieee.org/document/9881818/. doi:10.1109/Blockchain55522.2022.00023.

Breidenbach, M., Hamiti, F., Guluzade, A., Heiba, N., Mohamad, Y., Velasco, C., Herbeck Belnap, B., & Lühmann, D. (2023). 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* DSAI '22 (pp. 12–17). New York, NY, USA: Association for Computing Machinery. URL: https://dl.acm.org/doi/10.1145/3563137.3563157. doi:10.1145/3563137.3563157.

Duisberg, A. (2022). Legal Aspects of IDS: Data Sovereignty—What Does It Imply? In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage* (pp. 61–90). Cham: Springer International Publishing. URL: https://doi.org/10.1007/978-3-030-93975-5_5. doi:10.1007/978-3-030-93975-5_5.

Farahani, B., & Monsefi, A. K. (2023). Smart and collaborative industrial iot: A federated learning and data space approach. *Digital Communications and Networks*, *9*, 436–447. doi:https://doi.org/10.1016/j.dcan.2023.01.022.

Firdausy, D., Silva, P. D. A., Sinderen, M. J. v., & Iacob, M. E. (2022). Semantic discovery and selection of data connectors in international data spaces. URL: https://research.utwente.nl/en/publications/semantic-discovery-and-selection-of-data-connectors-in-internatio.

Griffo, C., Almeida, J. P. A., Guizzardi, G., & Nardi, J. C. (2021). Service contract modeling in Enterprise Architecture: An ontology-based approach. *Information Systems*, *101*, 101454. URL: https://www.sciencedirect.com/science/article/pii/S030643791930506X. doi:10.1016/j.is.2019.101454.

Hecker, D., Voss, A., & Wrobel, S. (2022). Data Ecosystems: A New Dimension of Value Creation Using AI and Machine Learning. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage* (pp. 211–224). Cham: Springer International Publishing. URL: https://doi.org/10.1007/978-3-030-93975-5_13. doi:10.1007/978-3-030-93975-5_13.

Immonen, A., Ovaska, E., & Paaso, T. (2018). Towards certified open data in digital service ecosystems. *Software Quality Journal*, *26*, 1257–1297. URL: https://doi.org/10.1007/s11219-017-9378-2. doi:10.1007/s11219-017-9378-2.

Janev, V., Vidal, M. E., Endris, K., & Pujic, D. (2021). Managing Knowledge in Energy Data Spaces. In *Companion Proceedings of the Web Conference 2021* WWW '21 (pp. 7–15). New York, NY, USA: Association for Computing Machinery. URL: https://dl.acm.org/doi/10.1145/3442442.3453541. doi:10.1145/3442442.3453541.

Jarke, M., Otto, B., & Ram, S. (2019). Data Sovereignty and Data Space Ecosystems. *Business & Information Systems Engineering*, *61*, 549–550. URL: https://doi.org/10.1007/s12599-019-00614-2. doi:10.1007/s12599-019-00614-2.

Jesse, N. (2018). Internet of Things and Big Data: the disruption of the value chain and the rise of new software ecosystems. *AI & SOCIETY*, *33*, 229–239. URL: https://doi.org/10.1007/s00146-018-0807-y. doi:10.1007/s00146-018-0807-y.

Kirstein, F., & Bohlen, V. (2022). IDS as a Foundation for Open Data Ecosystems. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage* (pp. 225–240). Cham: Springer International Publishing. URL: https://doi.org/10.1007/978-3-030-93975-5_14. doi:10.1007/978-3-030-93975-5_14.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele Univ.*, *33*.

Klug, L., & Prinz, W. (2023). Fair prices for sustainability in agriculture and food. Requirements and design options for a data-based transparency system. In *Proceedings of the 24th Annual International Conference on Digital Government Research* DGO '23 (pp. 496–507). New York, NY, USA: Association for Computing Machinery. URL: https://dl.acm.org/doi/10.1145/3598469.3598525. doi:10.1145/3598469.3598525.

Liu, M., Tu, Z., Xu, X., Wang, Z., & Wang, Y. (2023). A data-driven approach for constructing multilayer network-based service ecosystem models. *Software and Systems Modeling*, *22*, 919–939. URL: https://doi.org/10.1007/s10270-022-01029-6. doi:10.1007/s10270-022-01029-6.

Meckler, S., Dorsch, R., Henselmann, D., & Harth, A. (2023). The Web and Linked Data as a Solid Foundation for Dataspaces. In *Companion Proceedings of the ACM Web Conference 2023* WWW '23 Companion (pp. 1440–1446). New York, NY, USA: Association for Computing Machinery. URL: https://dl.acm.org/doi/10.1145/3543873.3587616. doi:10.1145/3543873.3587616.

Mengist, W., Soromessa, T., & Legese, G. (2020). Method for conducting systematic literature review and meta-analysis for environmental science research. *MethodsX*, *7*, 100777. doi:10.1016/j.mex.2019.100777.

Munoz-Arcentales, A., López-Pernas, S., Pozo, A., Alonso, , Salvachúa, J., & Huecas, G. (2019). An Architecture for Providing Data Usage and Access Control in Data Sharing Ecosystems. *Procedia Computer Science*, *160*, 590–597. URL: https://www.sciencedirect.com/science/article/pii/S1877050919317429. doi:10.1016/j.procs.2019.11.042.

Nagel, L., & Lycklama, D. (2022). How to Build, Run, and Govern Data Spaces. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage* (pp. 17–28). Cham: Springer International Publishing. URL: https://doi.org/10.1007/978-3-030-93975-5_2. doi:10.1007/978-3-030-93975-5_2.

Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, *37*. URL: https://hal.science/hal-01574600.

Otto, B., Rubina, A., Eitel, A., Teuscher, A., Schleimer, A. M., Lange, C., Stingl, D., Loukipoudis, E., Brost, G., Böge, G., Pettenpohl, H., Langkau, J., Gelhaar, J., Mitani, K., Hupperz, M., Huber, M., Jahnke, N., Brandstädter, R., Wessel, S., & Bader, S. (2021). *GAIA-X and IDS*.

Otto, B., Steinbuß, S., Teuscher, A., & Lohmann, S. (2019). Reference architecture model—international data spaces (version 3.0).

Pettenpohl, H., Spiekermann, M., & Both, J. R. (2022). International Data Spaces in a Nutshell. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces : The Ecosystem Approach* to *Competitive Advantage* (pp. 29–40). Cham: Springer International Publishing. URL: https://doi.org/10.1007/978-3-030-93975-5_3. doi:10.1007/978-3-030-93975-5_3.

Pinto, P., Sousa, C., & Cardeiro, C. (2023). Data spaces based approach for b2b data exchange: A footwear industry case. *Procedia Computer Science*, *219*, 933–940. doi:10.1016/j.procs.2023.01.369.

Pullmann, J., Petersen, N., Mader, C., Lohmann, S., & Kemeny, Z. (2017). Ontology-based information modelling in the industrial data space. In *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)* (pp. 1–8). Limassol, Cyprus: IEEE Press. URL: https://doi.org/10.1109/ETFA.2017.8247688. doi:10.1109/ETFA.2017.8247688.

Rezaei, R., Chiew, T., & Lee, S. (2014). A review on e-business interoperability frameworks. *Journal of Systems and Software*, *93*. doi:10.1016/j.jss.2014.02.004.

Sang, G. M., Xu, L., de Vrieze, P., Bai, Y., & Pan, F. (2021). Predictive Maintenance in Industry 4.0. In *Proceedings of the 10th International Conference on Information Systems and Technologies* ICIST '20 (pp. 1–11). New York, NY, USA: Association for Computing Machinery. URL: https://dl.acm.org/doi/10.1145/3447568.3448537. doi:10.1145/3447568.3448537.

Sardis, E., Gogouvitis, S. V., Bouras, T., Gouvas, P., & Varvarigou, T. (2013). 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* (pp. 68–75). URL: https://ieeexplore.ieee.org/document/6681211/. doi:10.1109/3PGCIC.2013.17.

Scerri, S., Tuikka, T., Vallejo, I., & Curry, E. (2022). Common European Data Spaces: Challenges and Opportunities. (pp. 337–357). doi:10.1007/978-3-030-98636-0_16.

Scheider, S., Lauf, F., Geller, S., Möller, F., & Otto, B. (2023). Exploring design elements of personal data markets. *Electronic Markets*, *33*, 28. URL: https://doi.org/10.1007/s12525-023-00646-3. doi:10.1007/s12525-023-00646-3.

Seidel, A., Wenzel, K., Hänel, A., Teicher, U., Weiß, A., Schäfer, U., Ihlenfeldt, S., Eisenmann, H., & Ernst, H. (2023). Towards a seamless data cycle for space components: considerations from the growing European future digital ecosystem Gaia-X. *CEAS Space Journal*, . URL: https://doi.org/10.1007/s12567-023-00500-4. doi:10.1007/s12567-023-00500-4.

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. (2023). Applications of convolutional neural networks in education: A systematic literature review. *Expert Systems with Applications*, *231*, 120621. doi:https://doi.org/10.1016/j.eswa.2023.120621.

da Silva Serapião Leal, G., Guédria, W., & Panetto, H. (2019). Interoperability assessment: A systematic literature review. *Computers in Industry*, *106*, 111–132. URL: https://www.sciencedirect.com/science/article/pii/S0166361518303476. doi:10.1016/j.compind.2019.01.002.

Solmaz, G., Cirillo, F., Fürst, J., Jacobs, T., Bauer, M., Kovacs, E., Santana, J. R., & Sánchez, L. (2022). Enabling data spaces: existing developments and challenges. In *Proceedings of the 1st International Workshop on Data Economy* DE '22 (pp. 42–48). New York, NY, USA: Association for Computing Machinery. URL: https://dl.acm.org/doi/10.1145/3565011.3569058. doi:10.1145/3565011.3569058.

Spiekermann, M. (2019). Data Marketplaces: Trends and Monetisation of Data Goods. *Intereconomics*, *54*, 208–216. URL: https://doi.org/10.1007/s10272-019-0826-z. doi:10.1007/s10272-019-0826-z.

Tardieu, H. (2022). Role of Gaia-X in the European Data Space Ecosystem. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage* (pp. 41–59). Cham: Springer International Publishing. URL: https://doi.org/10.1007/978-3-030-93975-5_4. doi:10.1007/978-3-030-93975-5_4.

Tchoffa, D., Figay, N., Ghodous, P., Exposito, E., Kermad, L., Vosgien, T., & El Mhamedi, A. (2016). Digital factory system for dynamic manufacturing network supporting networked collaborative product development. *Data & Knowledge Engineering*, *105*, 130–154. URL: https://doi.org/10.1016/j.datak.2016.02.004. doi:10.1016/j.datak.2016.02.004.

Theissen-Lipp, J., Kocher, M., Lange, C., Decker, S., Paulus, A., Pomp, A., & Curry, E. (2023). Semantics in Dataspaces: Origin and Future Directions. In *Companion Proceedings of the ACM Web Conference 2023* WWW '23 Companion (pp. 1504–1507). New York, NY, USA: Association for Computing Machinery. URL: https://dl.acm.org/doi/10.1145/3543873.3587689. doi:10.1145/3543873.3587689.

Weichhart, G., & Naudet, Y. (2014). 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* (pp. 219–228). Berlin, Heidelberg: Springer-Verlag volume 8842. URL: https://doi.org/10.1007/978-3-662-45550-0_23. doi:10.1007/978-3-662-45550-0_23.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., 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., 't Hoen, P. A. C., 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., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., & Mons, B. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. doi:10.1038/sdata.2016.18.

Wohlin, C., Kalinowski, M., Romero Felizardo, K., & Mendes, E. (2022). Successful combination of database search and snowballing for identification of primary studies in systematic literature studies. *Information and Software Technology*, *147*, 106908. URL: https://www.sciencedirect.com/science/article/pii/S0950584922000659. doi:10.1016/j.infsof.2022.106908.