U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

MESTRADO EM CIÊNCIA DA INFORMAÇÃO

INTEROPERABILIDADE SEMÂNTICA EM ESPAÇOS DE DADOS: DESENVOLVIMENTO DE UM PROCESSO DE SUPORTE À OPERACIONALIZAÇÃO

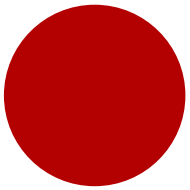
Rui Pedro Moreira Ribeiro

M

2026

UNIDADES ORGÂNICAS ENVOLVIDAS

FACULDADE DE ENGENHARIA
FACULDADE DE LETRAS



FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO
FACULDADE DE LETRAS DA UNIVERSIDADE DO PORTO

Interoperabilidade Semântica em Espaços de Dados: desenvolvimento de um processo de suporte à operacionalização

Rui Pedro Moreira Ribeiro



Mestrado em Ciência da Informação

Orientador: Prof. António Lucas Soares
Segundo Orientador: Henrique Diogo Silva

12 de Outubro de 2026

© Rui Pedro Moreira Ribeiro, 2026

Interoperabilidade Semântica em Espaços de Dados: desenvolvimento de um processo de suporte à operacionalização

Rui Pedro Moreira Ribeiro

Mestrado em Ciência da Informação

Aprovado em provas públicas pelo Júri:

Presidente: Prof. Gabriel de Sousa Torcato David
Arguente: Prof. João Nuno Lopes Barata
Vogal Orientador: Prof. António Manuel Lucas Soares

---

12 de Outubro de 2026

Resumo

A emergência dos espaços de dados enquanto infraestruturas federadas de partilha de informação representa um dos pilares estratégicos da União Europeia para a economia digital. Contudo, a materialização destes ecossistemas esbarra na heterogeneidade semântica e na frequente assunção errónea de que as organizações de um mesmo setor comunicam sob um vocabulário homogéneo. Embora arquiteturas de referência como a IDS-RAM 4.0 definam componentes técnicos (Connectors e Vocabulary Hubs) e o DSSC Blueprint 3.0 estabeleça diretrizes conceptuais de alto nível, persiste uma lacuna operacional: a ausência de um processo formal que oriente as entidades de governo na seleção, adaptação e validação de modelos de dados em fase de desenho (Design Time).

Adotando a metodologia Design Science Research (DSR), esta dissertação concebe um processo unificado de governação semântica para espaços de dados setoriais. O artefacto desdobra a entidade de governo em dois papéis funcionais estritos, o Fornecedor do Modelo de Dados e Autoridade de Governo, incorporando o papel de *Data Steward* em ambas segundo as boas práticas do Data Management Body of Knowledge (DAMA-DMBOK) e os padrões ISO/IEC 11179. A arquitetura é formalizada em UML 2.5.1 através de sete subprocessos que cobrem desde a seleção e definição do modelo de dados até à publicação no *Vocabulary Hub*, estando suportado em fichas operacionais de atividade ancoradas em padrões da Web Semântica (RDF, OWL, SHACL) do consórcio W3C e nas normas ISO/IEC.

A viabilidade do processo é demonstrada através de um cenário ilustrativo (walkthrough) centrado na rastreabilidade de lã reciclada através do Passaporte Digital do Produto (DPP) para a indústria têxtil, combinando o padrão eBIZ com a extensão TRICK e a especificação setorial de referência CEN CWA 18291:2025. A análise comparativa evidencia que o artefacto permite mitigar a dissonância ontológica e previne quebras nas cadeias de abastecimento de PMEs na utilização de espaços de dados, assegurando a conformidade direta com o Artigo 33.º do EU Data Act.

Abstract

The emergence of data spaces as federated infrastructures for information sharing represents one of the European Union's strategic pillars for the digital economy. However, the materialization of these ecosystems is hindered by semantic heterogeneity and the frequent erroneous assumption that organizations within the same sector communicate under a homogeneous vocabulary. Although reference architectures such as IDS-RAM 4.0 define technical components (Connectors and Vocabulary Hubs) and the DSSC Blueprint 3.0 establishes high-level conceptual guidelines, an operational gap persists: the absence of a formal process to guide governance entities in the selection, adaptation, and validation of data models during the Design Time phase.

Adopting the Design Science Research (DSR) methodology, this dissertation conceives a unified semantic governance process for sectoral data spaces. The artifact unfolds the governance entity into two strict functional roles, the Data Model Provider and Governance Authority, incorporating the Data Steward role into both in accordance with Data Management Body of Knowledge (DAMA-DMBOK) best practices and ISO/IEC 11179 standards. The architecture is formalized in UML 2.5.1 through seven subprocesses covering everything from the selection and definition of the data model to its publication in the Vocabulary Hub, supported by operational activity sheets anchored in Semantic Web standards (RDF, OWL, SHACL) from the W3C consortium and ISO/IEC norms.

The feasibility of the process is demonstrated through an illustrative walkthrough scenario centered on the traceability of recycled wool through the Digital Product Passport (DPP) for the textile industry, combining the eBIZ standard with the TRICK extension and the sectoral reference specification CEN CWA 18291:2025. The comparative analysis shows that the artifact helps to mitigate ontological dissonance and prevents breaks in SME supply chains when using data spaces, ensuring direct compliance with Article 33 of the EU Data Act.

UN Sustainable Development Goals

The United Nations Sustainable Development Goals (SDGs) provide a global framework for achieving a better and more sustainable future for all. It includes 17 goals[1] to address the world's most pressing challenges, including poverty, inequality, climate change, environmental degradation, peace, and justice.

First, it supports SDG 9 by developing a formal semantic governance artifact that empowers governance authorities of sectoral data spaces to systematically select, adapt, and manage the life cycle of interoperable data models. By establishing open application profiles and structured extension procedures, the process lowers technological integration barriers for Small and Medium-sized Enterprises (SMEs), preventing lock-in into proprietary silos and fostering collaborative innovation across industrial digital infrastructures.

Second, it aligns with SDG 16 by institutionalizing formal mechanisms of transparency, accountability, and regulatory integrity within data administration. The process enforces the strict segregation of duties between entities proposing and approving models, mandates reasoned audit reports for validation decisions, and guarantees full metadata lineage traceability via the W3C PROV-O ontology, directly operationalizing the non-discriminatory transparency mandates established in Article 33 of the EU Data Act.

Third, the research aligns with SDG 17, as data spaces inherently serve as sociotechnical infrastructures designed to enable multi-stakeholder data-sharing partnerships. By delivering shared formal semantics, explicit co-creation rules, and transparent cataloguing within federated Vocabulary Hubs, the artifact operationalizes sovereign cooperation and knowledge exchange among autonomous organizations.

Finally, at the level of SDG 8, the research fosters economic resilience and industrial productivity by automating semantic validation and eliminating the substantial costs and operational inefficiencies stemming from manual data mapping. Moreover, establishing formal schema versioning, deprecation timelines, and backward-compatible coexistence safeguards automated supply chain data pipelines from abrupt disruptions, preventing the technical exclusion of smaller enterprises from federated markets.

The specific Sustainable Development Goals mentioned have the following names:

SDG 8 Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all

SDG 9 Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation

SDG 16 Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels

[1] https://sdgs.un.org/goals

SDG 17 Strengthen the means of implementation and revitalize the Global Partnership for Sustainable Development

SDG	Target	Contribution	Performance Indicators and Metrics
9	9.1	Conceives an open, standardized semantic governance process enabling interoperable and sovereign sectoral data spaces, providing industrial ecosystems with federated infrastructures that prevent vendor lock-in.	Rate of data models successfully published and catalogued in open, federated Vocabulary Hubs.
	9.b	Provides structured application profiles, data dictionaries, and explicit extension rules that lower technical barriers for SMEs, fostering technological modernization across industrial supply chains.	Reduction in engineering time and costs required for SME ERP integration in federated data spaces.
16	16.6	Enforces organizational accountability and transparent decision-making through the segregation of duties between Data Model Providers and Governance Authorities, supported by formal audit records.	Number of documented audit reports and reasoned rejections issued prior to cataloguing.
	16.10	Guarantees public, non-discriminatory access to data structures and metadata lineage using W3C PROV-O, fulfilling the legal transparency requirements of Article 33 of the EU Data Act.	Compliance rate with mandatory metadata publication requirements in sectoral repositories.
17	17.16	Establishes the governance mechanisms and common semantic foundation necessary to operationalize multi-stakeholder data-sharing partnerships across autonomous organizations within sectoral ecosystems.	Number of autonomous organizations and domain stakeholders adopting shared semantic standards for cross-enterprise collaboration.
8	8.2	Imposes formal governance mechanisms including change management, backward compatibility, and version coexistence (minor vs. major), preventing abrupt operational breaks in automated data pipelines across value chains.	Zero supply chain disruptions and transactional failures caused by unannounced or incompatible data schema modifications.

Acknowledgements

Poderia eu dizer que isto seria um prólogo, um epílogo ou outra qualquer forma de expressão literária desenvolvida em versos longos, em linhas inacabadas e na sede inesgotável de um conhecimento que quer explorar tudo, mas que precisa de se limitar para instanciar algo novo. Fazer algo nesta natureza, exige acedermos aos confins da consciência, de desligarmos o botão dos limites da desistência, o de olharmos a corrida de frente e vermos o quão longe ainda estamos da meta.

Escrever esta dissertação foi essa corrida, um verdadeiro circuito de curva e contracurva, que exigiu destreza, uma força cinética voraz, direção e a capacidade de ir buscar ferramentas onde tantas vezes julguei que não as poderia encontrar. Foi o ter de derrubar baluartes cada vez maiores quanto mais avançava, foi o ter que resistir ao cansaço, ao esgotamento, o ter de superar mudanças na estrada, de direções imprevisíveis e tempos adversos.

Foi uma missão quase impossível chegar até aqui, a este momento onde consegui construir algo, e de conseguir materializar o que durante meses foi um fervilhar de ideias e que julguei que não conseguiria materializar, o de entrar por caminhos sinuosos e filosóficos e de me trazerem de volta à realidade para chegar a algum resultado palpável. Como diz um famoso músico português, enquanto houver estrada para andar, a gente vai continuar e aqui não é diferente.

Quando digo a "gente", refiro-me às pessoas espetaculares que me deram força, ferramentas, ideias e...excelentes momentos de reflexão. Não escrevo o que gostaria de escrever, mas escrevo o que sou capaz de escrever.

Agradeço em particular:

Ao meu professor e orientador, António Lucas Soares, por me dar a oportunidade de regressar a uma casa que tão bem me acolheu na minha primeira passagem, e que desta vez, no Mestrado, não foi diferente. Agradeço todo o conhecimento que me foi transmitindo ao longo destes quase cinco anos entre licenciatura e mestrado, todas as reflexões que me colocou, de me proporcionar a oportunidade de estar num ambiente de excelência e de trocar ideias com profissionais. Aprendi muito com o professor desde aquela longínqua aula de ISCI em que perguntou "Quem tinha desenvolvido a Web?" e ninguém soube responder. Hoje, cinco anos depois, julgo já conseguir responder a essa pergunta.

Ao meu co-orientador, Henrique Diogo Silva, pela forma exímia e exemplar com que orientou ao longo destes meses, na minha segunda passagem, quer pelo INESC, quer pela sua orientação. Agradeço por todo o conhecimento transmitido, por me colocar no caminho certo quando tantas vezes estive a dispersar, pelas críticas construtivas, pela forma com que me levou a refletir, pelas constantes adaptações ao plano para que eu conseguisse fazer o melhor trabalho, com rigor e qualidade. A nível pessoal, por ser uma grande inspiração como profissional e como pessoa, exemplo com dedicação, empenho e compromisso, podemos chegar a onde quisermos, e por todo o apoio demonstrado ao longo deste percurso.

À minha supervisora, Ana Carolina Chaves, pela forma com que me acompanhou ao longo desta jornada, desde a comunicação, apoio, disponibilidade e conselhos que me foi transmitindo.

Por me ensinar a importância de uma organização bem definida, pelos comentários construtivos, pela troca de ideias e esclarecimentos quando não compreendia algum conceito ou algum termo, e pela forma com que garantia que eu os ficava a compreender plenamente antes de avançar para outra fase. Muito aprendi contigo ao longo destes meses.

Ao INESC-TEC, em particular à equipa do SYSTEM, pela forma como me acolheram, por todas as sessões de CSI carregadas de conhecimento, debate, discussão e de boas memórias, pelas horas de almoço onde aprendi tanto, pelas experiências, pela oportunidade de ajudar a realizar uma conferência, pelo grupo de corrida e jogos, por me permitirem ser hoje uma pessoa muito mais rica de conhecimento do que há uns anos atrás.

Ao professores do MCI, LCI e outros tantos, por tudo o que me ensinaram ao longo destes anos, por cada aula, por cada questão, por cada momento de reflexão, por cada contribuição para me tornar uma melhor pessoa e melhor profissional.

Ás amigas de curso que me acompanharam no SYSTEM, a Sara e a Maria Fernanda, pelas conversas, pelo apoio, pelo companheirismo, pelas aventuras ao longo deste processo pelo qual passamos em conjunto, pela resiliência para chegarmos aos nossos objetivos.

Aos amigos de curso, pelo acompanhamento ao longo destes anos, por me permitirem conhecer diferentes realidades, diferentes trajetos, pelos diferentes projetos em que colaboramos, pelas discussões, pelas trocas de ideias e experiências, pelo apoio quando mais foi preciso, pela presença e contributos. Agradeço por tudo e que possamos contribuir para uma sociedade da informação mais rica e organizada.

Aos meus amigos. Seria injusto enumera-los todos porque provavelmente me iria esquecer de algum. Agradecer principalmente por acreditarem em mim, pelas diferentes conversas e momentos espetaculares que ajudaram a tornar este percurso mais leve, por me aconselharem, por me ouvirem quando só o stresse e o cansaço estavam presentes, por não me abandonaram nos momentos mais difíceis, por estarem para mim, por estarem presentes. Agradeço por muitas vezes me contrariarem, me ensinarem, me mostrarem um caminho.

Aos meus colegas e amigos, do Orfeão à Mentoria, das aulas de húngaro à saúde digital, nas humanidades digitais e tantos outros projetos, agradeço por ter vivido tantas experiências que me enriqueceram tanto como pessoa e culturalmente. Ao Orfeão, no qual pude voltar a uma das paixões antigas da música e integrar tanto o coro clássico/popular como o Cante Alentejano, onde tanto me diverti e tanto aprendi, como vencer o medo de cantar em público, o estar coordenado e a possibilidade de conhecer tanta gente espetacular dentro da Universidade. À Mentoria, o gosto de colaborar durante alguns anos na integração de novos estudantes na Universidade, algo que no meu primeiro ano foi difícil. Aos amigos das aulas de húngaro, por aceitarmos o desafio de aprendermos uma das línguas mais difíceis do mundo, que segundo o Chico Buarque, "é a única língua que o diabo respeita", e de me darem a oportunidade de conhecer diferentes culturas.

Ao meu núcleo familiar, agradeço por todo o amor, carinho e suporte ao longo desta grande jornada. À minha irmã, ao meu cunhado e ao meu sobrinho, que me incentivaram a ir mais longe, a jamais atirar a toalha ao chão mesmo que o desafio pareça não ter solução. À minha mãe, que me motiva todos os dias a levantar-me, a lutar pela vida, que está comigo desde sempre nas vitórias e nas derrotas, pela educação, por não deixar que me faltasse nada, pela resiliência.

Por último, um agradecimento especial à minha avó e ao meu "aliado", que infelizmente partiram antes de poder chegar a esta etapa.

A cada um de vós, o meu mais sincero agradecimento!

Rui Pedro Moreira Ribeiro

"É preciso imaginar Sísifo feliz"

Albert Camus

DECLARAÇÃO DE HONRA

Declaro que o presente trabalho de dissertação é de minha autoria e não foi previamente apresentado e avaliado noutro curso ou unidade curricular, desta ou de outra instituição.
As referências a outros autores (afirmações, ideias, pensamentos), assim como a utilização de trabalhos publicados respeitam escrupulosamente as regras da atribuição, e encontram-se devidamente indicadas no texto e nas referências bibliográficas, de acordo com as normas de referenciação. Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.
Declaro, ainda, que não utilizei ferramentas de Inteligência Artificial para a realização de parte(s) da presente dissertação, ou, quando tal utilização ocorreu, essa utilização e seus resultados estão devidamente assinalados e foram objeto de cuidada verificação para deteção de erros, imprecisões, atribuições infundadas ou outras formas de enviesamento de conteúdos e argumentos, bem como de determinação de autoria.
Mais declaro que incluí estudos realizados por vários co-autores, tendo especificado, com transparência e rigor, a colaboração de cada coautor, informado sobre a eventual existência de alguma outra dissertação em que o mesmo artigo tenha sido incluído, reproduzido o trabalho de forma integral com autorização da revista em que se encontra publicado e obtido o acordo expresso assinado por todos os co-autores para a inclusão do artigo.

Rui Pedro Moreira Ribeiro

Assinado por: **RUI PEDRO MOREIRA RIBEIRO**
Num. de Identificação: 30152266
Data: 2026.10.28 12:40:26+00'00'

CARTÃO DE CIDADÃO

Contents

1 Introdução **1**
- 1.1 Contexto e Motivação . . . 1
- 1.2 Descrição do Problema . . . 2
- 1.3 Objetivos e Resultados Esperados . . . 4
- 1.4 Estrutura da Dissertação . . . 5

2 Enquadramento Concetual (Background) **7**
- 2.1 Suporte aos Espaços de Dados: Data Space Support Centre e Blueprint 3.0 . . . 7
- 2.2 Gestão e Governação de Dados: O Papel do DAMA-DMBOK . . . 8

3 Revisão de Literatura **10**
- 3.1 A Economia dos Dados e a Necessidade de Soberania . . . 10
  - 3.1.1 Estratégia Europeia para os Dados: Do Regulamento (Data Act) à Realidade 10
  - 3.1.2 Ecossistemas de Dados: Definições, Atores e Dinâmicas de Valor . . . 11
  - 3.1.3 O Desafio da Fragmentação: Silos de Dados e Heterogeneidade . . . 12
- 3.2 Espaços de Dados: Arquitetura, Componentes e Suporte . . . 13
  - 3.2.1 Conceito e Arquiteturas de Referência (IDS e Gaia-X) . . . 14
  - 3.2.2 Componentes Técnicos Essenciais: O Papel do Connector, do Vocabulary Hub e do Metadata Broker . . . 17
  - 3.2.3 Atores e Papéis Funcionais-Data Provider, Data Consumer, Data Intermediary e Vocabulary Provider . . . 19
- 3.3 Interoperabilidade Semântica: O Motor da Comunicação . . . 22
  - 3.3.1 Níveis de Interoperabilidade: Da conexão técnica ao entendimento semântico . . . 22
  - 3.3.2 Artefactos Semânticos e Tecnologias de Suporte: Dos Vocabulários, Catálogos e Ontologias aos Knowledge Graphs e Linked Data . . . 25
  - 3.3.3 O Problema da "Dissonância Ontológica" em Ambientes Federados . . . 28
- 3.4 Instanciação e Governação de Espaços de Dados . . . 30
  - 3.4.1 Governação de Espaços de dados: Modelos Centralizados, Descentralizados e Federados . . . 30
  - 3.4.2 Processos de Instanciação e Onboarding: Como integrar novos participantes? . . . 31
  - 3.4.3 Ciclo de Vida dos Vocabulários: Gestão, Manutenção e Versão em Ecossistemas . . . 33
  - 3.4.4 Lacunas na Literatura: A falta de um processo unificado de governação semântica. . . . 34

4 Metodologia **40**
4.1 Introdução . . . 40
4.2 Desenvolvimento da Investigação-Design Science Research . . . 41
4.2.1 Ciclos de Investigação . . . 41
4.3 Concretização da Investigação . . . 44
4.3.1 Questões de Investigação . . . 44
4.3.2 Análise da Literatura: Fases de triagem . . . 45
4.4 Construção do Artefacto . . . 47

5 Construção do Artefacto **49**
5.1 Posicionamento do Artefacto no Ciclo de Vida de Espaços de Dados . . . 49
5.1.1 Enquadramento no Blueprint 3.0 e no Modelo de Co-Criação . . . 49
5.1.2 Âmbito do Processo: Início, Fim e Tarefas Essenciais . . . 50
5.2 Fundamentação da Conceção do Processo . . . 50
5.2.1 Base Normativa e Lacuna Identificada . . . 51
5.2.2 Perguntas de Co-Criação como Estrutura de Derivação do Processo . . . 52
5.2.3 Papéis e Responsabilidades no Processo de Governo . . . 54
5.3 Arquitetura do Processo . . . 55
5.3.1 Visão Geral da Arquitetura . . . 55
5.3.2 Fundamentação da Notação e Processo de Desenvolvimento . . . 57
5.3.3 Lógica de Decisão, Validação e Ciclos de Iteração . . . 57
5.3.4 Importância do limite de iterações nas decisões do processo . . . 60
5.3.5 Mapeamento dos Subprocessos Decompostos . . . 61
5.4 Alinhamento Normativo e Governação . . . 64
5.4.1 Normas Internacionais (International Standard Organization) . . . 64
5.4.2 Padrões W3C . . . 65
5.4.3 Cumprimento do Artigo 33º do Data Act . . . 67
5.5 Síntese de Resultados do Artefacto . . . 68
5.5.1 Síntese do Artefacto Desenvolvido . . . 68

6 Validação e Discussão do Artefacto **69**
6.1 Caracterização e Cenário de Caso de Uso Hipotético . . . 69
6.1.1 Caso de uso exemplo . . . 69
6.1.2 Cenários . . . 70
6.1.3 Cenário 1 — Só com as diretrizes da Blueprint 3.0 . . . 70
6.1.4 Cenário 2 — Aplicação do Processo . . . 71
6.2 Limites de demonstração e validação do artefacto . . . 72
6.3 Discussão: Resposta às Questões de Investigação . . . 73
6.3.1 Síntese das Respostas às Questões Formuladas . . . 73
6.3.2 Valor para a Governação Semântica Setorial . . . 74
6.3.3 Contributo Científico para a Ciência da Informação . . . 75

7 Conclusões e Trabalho Futuro **76**
7.1 Síntese do Trabalho Desenvolvido . . . 76
7.2 Contributos Principais . . . 76
7.3 Limitações do Estudo . . . 77
7.4 Linhas de Investigação Futura . . . 77

References **78**

A Prompts Utilizadas **84**
A.1 Prompt de avaliação e melhoria do artefacto 84
A.2 Prompt Análise Normas 84

B Subprocessos **87**

C Processo Inicial Pré-Validação **94**

D Fichas de Atividade **96**
D.1 96
D.2 97
D.3 98
D.4 99
D.5 100
D.6 101
D.7 102
D.8 103
D.9 103
D.10 104
D.11 105

List of Figures

4.1 Design Science Research, Peffers et al. (2007) 40
4.2 Design Science Research Cycles, Hevner (2007) 42
4.3 Expressão de Pesquisa 45
4.4 Diagrama PRISMA para Revisão de Literatura 48

5.1 Data Models Technical Building Block (Data Spaces Support Centre, 2026) . . . 52
5.2 Selecionar Modelo Candidato 56
5.3 processo de atividades de alto nível do processo *Selecionar e Publicar Modelo de Dados do Espaço de Dados*, com as duas partições de responsabilidade. 58
5.4 Legenda notacional utilizada nos diagramas UML do artefacto. 59
5.5 Data Model Scorecard (Hoberman, 2015; Henderson et al., 2017) 60
5.6 Exemplo de Ficha de Atividade Simples 64

B.1 Subprocesso 2: Analisar Adequação do Modelo. 88
B.2 Subprocesso 3: Especificar Modelo de Raiz. 89
B.3 Subprocesso 4: Estruturar Extensão. 90
B.4 Subprocesso 5: Refinar Extensão. 91
B.5 Subprocesso 6: Avaliar Qualidade do Modelo. 92
B.6 Subprocesso 7: Publicar e Governar Modelo. 93

C.1 Protótipo de Processo Pré-Validação 95

List of Tables

3.1 Comparação entre IDS-RAM e Gaia-X 15

5.1 Mapeamento dos subprocessos decompostos do artefacto. 62

List of Acronyms

API Application Programming Interface

CWA CEN Workshop Agreement

DAMA-DMBOK Data Management Body of Knowledge

DCAT Data Catalog Vocabulary

DPP Digital Product Passport

DSSC Data Space Support Centre

GRS Global Recycled Standard

IDS-RAM International Data Spaces Reference Architecture Model

IoT Internet of Things

ISO International Organization for Standardization

LLM Large Language Model

LOV Linked Open Vocabularies

ODRL Open Digital Rights Language

OWL Web Ontology Language

PME Pequenas e Médias Empresas

RDF Resource Description Framework

SKOS Simple Knowledge Organization System

TSLO Technical, Semantical, Legal, Organizational

UML Unified Modeling Language

URI Uniform Resource Identifier

W3C World Wide Web Consortium

XML Extensible Markup Language

Chapter 1

Introdução

1.1 Contexto e Motivação

“A Biblioteca é ilimitada e periódica. Se um viajante eterno a atravessasse em qualquer direção, comprovaria ao cabo de séculos que os mesmos volumes se repetem na mesma desordem (que, repetida, seria uma ordem: a Ordem)” (Borges et al., 2003).

A visão de Jorge Luís Borges, ao identificar a Biblioteca de Babel, encontrava-se muito à frente do seu tempo. Esta metáfora pode ser vista à luz atual como o meio digital onde nos encontramos. Babel ganha novo significado quando pensamos na metáfora da Torre de Babel: povos diferentes a falarem línguas distintas sem se entenderem, numa realidade que desafia as tecnologias atuais. Num tempo em que os dados são o motor da economia global, as organizações enfrentam o desafio de gerir dados de forma segura, eficaz e plenos de significado, garantindo a interoperabilidade e o valor ao longo do seu ciclo de vida. Neste cenário, os "espaços de dados" surgem como uma infraestrutura crítica para assegurar a soberania e a partilha de dados em ecossistemas de benefício mútuo (Curry et al., 2022).

Contudo, a abundância de dados gerados diariamente, impulsionada pela adoção generalizada de tecnologias digitais, resulta numa vasta fragmentação de formatos e de vocabulários (Ganzha et al., 2021; Möller et al., 2024). Tal como os construtores da Torre de Babel, as várias organizações e os seus sistemas de informação comunicam com "idiomas" diferentes, culminando na criação de silos de informação isolados (Ganzha et al., 2021; Dominik Lis et al., 2023). Para desbloquear o verdadeiro potencial dos dados na economia digital, é imperativa a transição de uma gestão de dados a nível interno para a partilha inter-organizacional, através dos ecossistemas de dados (Lis et al., 2021; Möller et al., 2024). Nestes ecossistemas o valor é gerado através da colaboração dinâmica entre fornecedores, consumidores e intermediários na exploração, procura, publicação e consumo de dados, atuando como um conjunto de redes socio-técnicas no suporte a novos modelos de negócio (Oliveira et al., 2019; Möller et al., 2024).

Os espaços de dados fornecem a infraestrutura que permite gerar estes ecossistemas de dados, assentando numa arquitetura descentralizada ou federada, que assegura a soberania, permitindo a

quem participa no espaço de dados possa partilhar informação com confiança, mantendo o controlo total sobre quem acede aos dados, em que condições e para que finalidade (Möller et al., 2024).

Não obstante a arquitetura dos espaços de dados, o estabelecimento de confiança e a criação de ligações técnicas, não são suficientes se a informação partilhada não for integralmente compreendida, dependendo do aspeto crítico da interoperabilidade semântica (Cimmino et al., 2024; Steinbuss; Almeida, et al., 2025). A falta de clareza ou a ambiguidade no significado dos dados partilhados resulta em falhas de comunicação, erros de interpretação e tomadas de decisão incorretas, levando as organizações a despenderem custos avultados no mapeamento da informação para formatos em conformidade com o sistema.

Sem uma semântica partilhada e formalizada, a troca de dados pode ser tecnicamente bem sucedida, mas perde-se o contexto, impossibilitando que as máquinas e as pessoas da organização que recebe os dados processem e interpretem a informação com confiança.

Neste sentido, a motivação central desta investigação, reside na necessidade de mitigar as barreiras semânticas que limitam a escalabilidade dos espaços de dados, desenvolvendo abordagens centradas na interoperabilidade semântica para garantir partilha de dados sem perdas de significado, com geração de valor, transformando ao nível dos espaços de dados, o fator caótico da metáfora da Torre de Babel num ecossistema estruturado.

1.2 Descrição do Problema

Na atual economia digital, onde os dados são um ativo fundamental na geração de valor, a sua exploração entre organizações esbarra frequentemente no isolamento de sistemas e na relutância das empresas em partilhar informações sensíveis, que coloquem em causa a sua competitividade (Brechtel, 2025). Mesmo quando existe vontade de partilhar, as organizações deparam-se com a fragmentação da informação como obstáculo técnico persistente, que consistem em repositórios isolados onde os dados estão estruturados em formatos incompatíveis, impedindo a sua reutilização tanto por barreiras técnicas como pela ausência de significado partilhado (Auer, 2022; Casolari et al., 2023; Harmelink et al., 2024).

Para ultrapassar estas barreiras, os espaços de dados emergiram como infraestruturas descentralizadas de partilha de dados, assentes em mecanismos técnicos e de governo que garantem a soberania, a confiança e o controlo dos participantes sobre os dados que disponibilizam (Möller et al., 2024; Kernstock et al., 2025).

Para que estas infraestruturas funcionem plenamente, as organizações necessitam de ultrapassar múltiplas camadas de interoperabilidade. A camada técnica garante que os sistemas conseguem comunicar e transferir dados num formato legível. A camada semântica garante que o significado desses dados é preservado e interpretado de forma consistente pelos sistemas recetores (Cimmino et al., 2024; Steinbuss; Almeida, et al., 2025).É neste cenário que arquiteturas de referência como a *International Data Spaces Reference Model 4.0* (IDS-RAM 4.0) (International Data Spaces Association, 2025) se destacam.

A arquitetura IDS-RAM 4.0 (International Data Spaces Association, 2025) fornece as bases para interoperabilidade sintática, garantindo que os dados circulam entre sistemas em formatos reconhecíveis e estabelece os componentes para suportar a interoperabilidade semântica. Contudo, como identificam Auer (2022) e Yan et al. (2024), a garantia de que o significado dos dados é preservado ao longo das trocas de dados permanece sem resposta sistemática, com destaque para o nível setorial.

A interoperabilidade semântica é responsável por assegurar que os dados ao serem trocados entre atores e organizações, não perdem o seu significado original, independentemente dos sistemas de informação utilizados. A medição da temperatura reflete um exemplo prático, se considerarmos organizações que utilizam medidas como *Celsius*,*Fahrenheit*, ou o *Kelvin* quando armazenam dados. O objetivo é o mesmo com diferentes representações, daí que durante o processo de troca, se os atores não identificarem claramente as unidades utilizadas, pode acontecer um conflito semântico.

A troca de dados do ponto de vista técnico pode ocorrer sem problemas, mas o significado fica comprometido. A arquitetura IDS-RAM foi concebida para permitir a instanciação de estruturas como os espaços de dados para resolução das questões da interoperabilidade e entendimento comum. No entanto, grande parte dos trabalhos desenvolvidos para os espaços de dados está focado numa visão de alto nível do sistema, quando observamos organizações de diferentes setores como a manufatura ou a indústria.

Esta lacuna agrava-se quando a análise desce do nível intersetorial para o nível setorial. Segundo Kovach et al. (2026), menos de 35% dos estudos identificados integram modelos semânticos de forma coesa, e a investigação concentra-se predominantemente na visão intersetorial, negligenciando as necessidades específicas de cada sector. Fica evidente a pré-conceção da ideia de que as organizações do mesmo setor são semanticamente homogéneas, tanto nos sistemas de informação como nos artefactos semânticos a que recorrem para comunicar (Kovach et al., 2026). Segundo Brechtel (2025), a investigação tende a falhar por negligenciar a diversidade dos grupos de atores nas cadeias de abastecimento industriais, assumindo que todos partilham os mesmos objetivos, motivos e necessidades para a partilha de dados. O Catena-X, um espaço de dados na indústria automóvel caracterizado por cadeias de abastecimento altamente complexas e globais, serve como um caso de estudo para desconstruir a visão homogénea dentro dos setores, evidenciando as diferenças em termos de dimensão, presença no mercado, poder e interdependência. Neste âmbito, é possível observar uma tensão entre política europeia, que assume uma visão intersetorial da interoperabilidade, e a realidade prática, que demonstra que a interoperabilidade setorial ainda não está resolvida (Steinbuss; Almeida, et al., 2025).

Assim, na ausência de um processo que governe os artefactos semânticos ao nível setorial, a infraestrutura dos espaços de dados fornece o canal de comunicação, mas não a linguagem comum (Nagel et al., 2022). Na ausência deste processo, a ambição europeia de ecossistemas de dados interoperáveis fica comprometida na sua fundação mais essencial, mais concretamente, o significado partilhado entre os participantes do espaço.

Assim sendo, o problema central desta dissertação reside na interoperabilidade semântica, com

a necessidade de garantir que os dados são utilizados de forma eficaz em diferentes aplicações e contextos, contribuindo para superar o fenómeno que a metáfora da Torre de Babel descreve.

1.3 Objetivos e Resultados Esperados

O objetivo primário desta dissertação consiste em desenvolver um processo unificado de suporte à interoperabilidade semântica em espaços de dados setoriais, centrado na gestão do ciclo de vida de artefactos semânticos, que permitam às organizações de um mesmo setor partilhar dados com significado consistente e sem perdas de contexto, bem como auxiliar as entidades de governo. O processo abrange as fases de seleção, avaliação de maturidade, publicação e governo destes artefactos no âmbito da arquitetura IDS-RAM 4.0 (International Data Spaces Association, 2025; Steinbuss; Almeida, et al., 2025).. Através de uma lente conceptual e prática, a investigação inclui como objetivos específicos:

- Caracterizar a realidade semântica dos setores industriais, caracterizando os artefactos semânticos existentes, o seu grau de maturidade e as lacunas de cobertura que comprometem a interoperabilidade setorial.
- Mapear os fluxos de informação e os mecanismos de governo de dados adotados em espaços de dados setoriais, identificando os requisitos funcionais e de governo que um processo de suporte semântico deve satisfazer.
- Validar o processo desenvolvido através de um caso de uso hipotético-dedutivo, de forma descritiva do contexto setorial, utilizando dados ilustrativos a partir de um caso real (Venable et al., 2012).

Os resultados esperados pretendem contribuir para a inovação na gestão da informação, oferecendo um quadro estruturado para que as organizações alcancem resultados superiores através da colaboração efetiva e da soberania de dados. Para operacionalizar estes objetivos, a investigação é orientada pelas seguintes questões de investigação:

Questão Principal *De que modo podemos auxiliar as entidades de governo em espaços de dados setoriais para garantir a interoperabilidade semântica?*

Questões Complementares

- Que elementos-chave de suporte à interoperabilidade semântica deverão estar contemplados neste processo?
- De que modo os espaços de dados resolvem a dissonância ontológica e a interoperabilidade semântica entre sistemas heterogéneos ao nível setorial?

O principal resultado desta investigação é um processo estruturado de suporte à interoperabilidade semântica em espaços de dados setoriais, que define como selecionar, avaliar, publicar e governar artefactos semânticos, como ontologias e vocabulários de domínio com a arquitetura

IDS-RAM 4.0 como referência. Este processo permite que a autoridade de governo de um espaço de dados setorial assegure que os dados partilhados entre participantes do mesmo setor preservem o seu significado ao longo de todo o ciclo de troca, reduzindo a ambiguidade semântica e a necessidade de mapeamento manual entre sistemas (Steinbuss; Almeida, et al., 2025; Kovach et al., 2026). O contributo desta investigação situa-se na interseção da gestão de informação, do governo de dados e da arquitetura de espaços de dados, respondendo a uma lacuna identificada na literatura relativamente ao nível setorial, que carece de validação empírica.

A revisão do estado da arte da interoperabilidade semântica em espaços de dados e consequentemente o desenvolvimento do processo de suporte, permitirão responder às questões de investigação definidas. Com esse propósito, a secção seguinte estabelece a estrutura para guiar esta dissertação.

1.4 Estrutura da Dissertação

A dissertação está estruturada de acordo com os seguintes capítulos:

- **Capítulo 1** - Introdução

Permite estabelecer o contexto e a motivação para o desenvolvimento do projeto de dissertação, formular o problema e as questões de investigação, bem como os objetivos e os resultados esperados. Adicionalmente, introduz a relevância da interoperabilidade semântica e da governação na perspetiva atual dos ecossistemas de dados.

- **Capítulo 2** - Background

Contextualiza os instrumentos nucleares e normativos que suportam a construção de um artefacto, detalhando as diretrizes do Data Space Support Center (DSSC), o método de co-criação aplicado aos diferentes componentes do espaço de dados, além das boas práticas de gestão e governo de dados do Data Management Body of Knowledge (DAMA-DMBOK).

- **Capítulo 3** - Revisão de Literatura

Apresenta uma revisão do estado da arte sobre os domínios centrais do estudo: a economia e ecossistemas de dados, arquiteturas de referência para Espaços de Dados (como a IDSA e Gaia-X), interoperabilidade semântica e modelos de governação de dados. Delineia as principais perspetivas teóricas, os desafios da fragmentação de dados e as lacunas na literatura que fundamentam a necessidade do processo proposto.

- **Capítulo 4** - Metodologia

Detalha a abordagem metodológica baseada na *Design Science Research* (DSR). Explica a fundamentação para a adoção desta metodologia orientada à resolução de problemas e descreve a forma como os ciclos de Relevância, Rigor e Design são aplicados para a construção e avaliação iterativa do artefacto da investigação.

- **Capítulo 5** - Construção do Artefacto

Apresenta a arquitetura e formalização do processo de governação semântica em UML 2.5.1, suportado nas diretrizes do DSSC Blueprint 3.0, no DAMA-DMBOK, nas normas ISO/IEC 11179 e nos standards W3C articulado com a arquitetura IDS-RAM 4.0.

- **Capítulo 6** - Validação e Discussão do Artefacto

Demonstra a exequibilidade do artefacto através de um cenário *walkthrough* no Passaporte Digital do Produto (lã reciclada), comparando a abordagem ad-hoc com o processo governado e discutindo as respostas às questões de investigação.

- **Capítulo 7** - Conclusões e Trabalho Futuro

Sintetiza os principais contributos teóricos e práticos alcançados, discute as limitações assumidas em Design-Time e define as linhas de investigação futura.

Chapter 2

Enquadramento Concetual (Background)

A conceção de um processo formal de suporte à interoperabilidade semântica em espaços de dados exige a articulação prévia entre dois referenciais metodológicos e normativos: as diretrizes de desenho federado promovidas pela União Europeia através do *Data Space Support Centre* (DSSC) e o corpo de conhecimento internacionalmente consolidado em gestão e governação de dados do DAMA-DMBOK.

O presente capítulo estabelece a base conceptual e operacional destes dois instrumentos, delimitando o contexto tecnológico sobre o qual a revisão da literatura científica (Capítulo 3) e a construção da solução (Capítulo 5) se sustentam.

2.1 Suporte aos Espaços de Dados: Data Space Support Centre e Blueprint 3.0

Enquanto as arquiteturas de referência para espaços de dados (como a IDS-RAM 4.0, aprofundada na Secção 3.2) especificam os blocos técnicos de software para a partilha de informação (tais como *Connectors*, *Vocabulary Hubs* e *Metadata Brokers*), a sua operacionalização no terreno carece de uma abordagem orientadora para as entidades gestoras desde a fase conceptual até à produção. É neste âmbito que surge o *Data Space Support Centre* (DSSC) (Data Spaces Support Centre, 2026), iniciativa criada sob o programa *Digital Europe* da Comissão Europeia para facilitar e harmonizar a criação de espaços comuns de dados setoriais.

A principal entrega do Data Space Support Centre (DSSC) materializa-se na sua *Blueprint 3.0* (Data Spaces Support Centre, 2026), instrumento que organiza o desenvolvimento de ecossistemas através de um conjunto coerente de Blocos de Construção (*Building Blocks* técnicos, organizacionais e de negócio). Em termos de ciclo de vida de implementação, a *Blueprint* categoriza a evolução do ecossistema em três fases fundamentais:

- **Fase Exploratória:** Identificação de modelos de negócio, casos de uso e avaliação preliminar de viabilidade;
- **Fase Preparatória ou de Desenho (*Design-Time*):** Momento em que se estruturam a governação, os modelos de dados, os contratos e o caderno de encargos normativo (*Rulebook*) (Steinbuss, 2020; Data Spaces Support Centre, 2026);
- **Fase de Implementação e Execução (*Runtime*):** Instanciação física dos conectores, execução dinâmica de contratos e troca efetiva de dados entre participantes (Data Spaces Support Centre, 2026).

Para orientar a progressão entre estas etapas, o DSSC disponibiliza o Método de Co-criação (*Co-creation Method*). No bloco técnico de *Data Models*, este método sublinha que as autoridades devem assegurar a reutilização de esquemas consolidados, definir perfis de aplicação e publicar catálogos de metadados. Contudo, a documentação limita-se a um nível de abstração alto: descreve os resultados esperados, mas omite o fluxo sequencial, as regras de decisão e a segregação de funções requerida entre quem propõe e quem valida os modelos semânticos. Esta lacuna processual no ciclo de vida é o ponto de partida para o artefacto concebido nesta dissertação.

2.2 Gestão e Governação de Dados: O Papel do DAMA-DMBOK

Se a *Blueprint* da DSSC fornece a arquitetura macro para espaços de dados, a gestão da qualidade e a curadoria dos artefactos que neles circulam requerem a adoção de boas práticas consagradas na Ciência da Informação e na gestão organizacional de dados.

O *Data Management Body of Knowledge* (DAMA-DMBOK) (Henderson et al., 2017) constitui o referencial normativo global que organiza as funções e práticas de gestão de informação. No âmbito da interoperabilidade e governação semântica tratadas nos capítulos subsequentes, relevam-se três das suas áreas funcionais:

- **Governação de Dados (*Data Governance*):** Estabelece a autoridade formal, as políticas e o controlo sobre os ativos informacionais. Destaca-se o princípio da segregação de funções (*segregation of duties*), que postula uma separação nítida entre as entidades que produzem e gerem modelos e as autoridades que auditam a sua conformidade regulatória. É nesta disciplina que se ancora o papel nuclear do *Data Steward*, enquanto agente de garantia da integridade concetual dos dados (Henderson et al., 2017).
- **Modelação e Desenho de Dados (*Data Modeling and Design*):** Sistematiza a transição dos conceitos abstratos de negócio para esquemas lógicos e físicos implementáveis. Destaca-se aqui o instrumento de auditoria *Data Model Scorecard* (Hoberman, 2015; Henderson et al., 2017), uma grelha de avaliação métrica estruturada em dez dimensões que permite quantificar a consistência, normalização e qualidade de um modelo antes da sua homologação em produção.

- **Gestão de Metadados (*Metadata Management*):** Define as diretrizes para catalogação, preservação e controlo de versões de esquemas, vocabulários e ontologias. O DAMA-DMBOK fundamenta esta área na família de normas internacionais ISO/IEC 11179 (*Metadata Registries*), sustentando que a interoperabilidade semântica entre sistemas distribuídos depende da existência de repositórios federados regidos por definições unívocas e regras padronizadas de nomenclatura (Henderson et al., 2017).

Ao articular o referencial setorial europeu (DSSC) com a gestão de dados (DAMA-DMBOK), este enquadramento antecipa os conceitos-chave mobilizados na revisão do estado da arte sobre espaços de dados e interoperabilidade semântica apresentada de seguida no Capítulo 3.

Chapter 3

Revisão de Literatura

A seguinte secção de revisão de literatura está definida em quatro pontos com o objetivo de explorar o estado da arte dos ecossistemas de dados, estabelecer claramente a distinção entre ecossistemas de dados e espaços de dados, identificar onde se insere a interoperabilidade semântica na partilha de dados e consolidar as lacunas que ainda existem na investigação para este aspeto.

3.1 A Economia dos Dados e a Necessidade de Soberania

Esta primeira secção enquadra a investigação, partindo da análise do conjunto de políticas europeias para chegar aos fundamentos técnicos e organizacionais dos ecossistemas de dados. Num primeiro ponto, é examinada a Estratégia Europeia para os Dados e de que forma o Data Governance Act e o Data Act estabelecem as condições legais para a partilha de dados num mercado digital único, identificando o conjunto de aspetos em que ambos os diplomas se complementam e as tensões que levantam ao nível da interoperabilidade (Casolari et al., 2023). Na segunda secção, é aprofundado o conceito de ecossistema de dados, caracterizando os seus atores, papéis e dinâmicas de criação de valor, com base numa perspectiva sociotécnica (Susha et al., 2023). Por fim, é analisado o desafio da fragmentação da informação, com exploração dos diferentes níveis de heterogeneidade que dificultam a partilha de dados, cujo o enquadramento legal não consegue resolver por si só. A heterogeneidade semântica emerge como uma barreira profunda e complexa, que abre caminho para o problema central que esta dissertação procura resolver (Yan et al., 2024).

3.1.1 Estratégia Europeia para os Dados: Do Regulamento (Data Act) à Realidade

Nos objetivos da União Europeia, uma economia assente em dados é definida como algo crucial, harmonizando uma visão estratégica entre todos os Estados Membros num mercado digital único. Esta ambição atenta no objetivo de fortalecer, de forma soberana, a competitividade no bloco europeu face aos grandes atores globais, posicionando o desenvolvimento de ecossistemas de dados no centro da inovação a nível industrial. De acordo com Casolari et al. (2023), esta autonomia estratégica não depende apenas da capacidade técnica de armazenamento de dados, mas da criação de um ambiente em que estes possam circular entre vários setores com total segurança estando

alinhados com regras claras. Não obstante, neste objetivo é fundamental alavancar a economia europeia, com a materialização desta visão, exigindo um equilíbrio delicado entre a partilha exaustiva de informação e o respeito pelos direitos fundamentais de pessoas e organizações dentro da União Europeia.

A necessidade de estabelecer uma legislação mais robusta e específica, contribuiu, assim, para o desenvolvimento do Data Governance Act e do Data Act. Apesar de ambos partilharem uma base comum de fomentar a partilha de dados, a sua visão e aplicação divergem na parte estratégica. O Data Governance Act, conforme analisado por Möller et al. (2024), é focado na criação de uma infraestrutura de confiança, regulamentando o papel dos intermediários neutros, promovendo altruísmo em relação aos dados. Trata-se de um regulamento focado no "como" os dados deverão ser partilhados, garantindo uma atuação sem conflitos de interesse (Casolari et al., 2023; Möller et al., 2024). Em contrapartida, o Data Act assume uma dimensão mais substantiva, na definição dos direitos de acesso e utilização, mais concretamente no contexto da Internet das Coisas (Internet of Things (IoT)). Enquanto o Data Governance Act tem a função de definir a organização dos mecanismos de governo, o Data Act define as obrigações legais para que quem detém o controlo dos dados permita o acesso aos mesmos, fortalecendo uma visão contrária à proteção exaustiva de dados, do ponto de vista competitivo, bem como o isolamento em silos proprietários (Harmelink et al., 2024). Através desta visão, é possível identificar a complementaridade entre estes diplomas. Enquanto o Data Governance Act fornece o quadro de confiança e os intermediários certificados, o Data Act confere a base de direito jurídico no acesso aos dados.

Contudo, é de salientar que esta complexidade entre normas levanta novos desafios a diferentes níveis da partilha de dados, como desafios técnicos de transporte de dados. O sucesso da partilha de dados, através de infraestruturas como os espaços de dados, conceito a explorar em detalhe na 3.2, com análise do conceito e do conjunto de arquiteturas de referência Fierens (2024) and Casolari et al. (2023), depende da interoperabilidade semântica para assegurar a troca de dados sem perdas de valor e significado, estando esta dissertação centrada neste aspeto.

3.1.2 Ecossistemas de Dados: Definições, Atores e Dinâmicas de Valor

A operacionalização das normas europeias discutidas no ponto anterior, mais concretamente a materialização da soberania, numa economia baseada em dados, exige muito mais do que um enquadramento legal robusto. Surge a necessidade de estudar de forma aprofundada as estruturas que permitem a colaboração, bem como os mecanismos que irão permitir configurar estas estruturas, que de ponto de vista social, são complexas.

É neste contexto que surge o conceito de ecossistema de dados, assumindo desde logo um papel central no ambicioso objetivo europeu. Enquanto o Data Act e o Data Governance Act estabelecem as regras, o ecossistema representa o espaço onde decorre a interação entre múltiplos atores, com uma visão alinhada na transformação de dados em bruto em ativos de valor. Porém, a transição para novos modelos de governo envolve uma reformulação profunda das cadeias de valor tradicionais, na sua maioria forma lineares, para um conjunto de redes dinâmicas e multidimensionais focadas na colaboração (Susha et al., 2023). Com base na literatura, é possível definir um

ecossistema de dados como um conjunto de redes socio-técnicas composta por diferentes atores, com o objetivo de colaboração, garantindo a recolha, processamento e partilha de dados (Susha et al., 2023; Otto, 2022).

O conceito parte da analogia com os ecossistemas da biologia, nas relações entre os diferentes seres vivos, como animais e plantas, para assegurar o equilíbrio do ecossistema. Em linha com a visão de Susha et al. (2023), um ecossistema não está limitado à infraestrutura técnica, sendo definido por um conjunto de relações de governo, modelos de negócio e padrões técnicos que possibilitam a co-criação de valor. Lnenicka et al. (2024) identificam os ecossistemas como ambientes de informação distribuídos e abertos, que permitem reposicionar as cadeias de abastecimento, de uma visão tradicional e fechada, para uma colaboração onde atores autónomos podem entrar, interagir e partilhar recursos (Oliveira et al., 2019; Möller et al., 2024). Em simultâneo, o ecossistema assume um carácter evolutivo, uma vez que atua como um sistema dinâmico, no qual os elementos, atores e infraestruturas se adaptam e evoluem continuamente de acordo com as necessidades do mercado (Susha et al., 2023; Lnenicka et al., 2024). Portanto, numa definição clara, o ecossistema de dados é caracterizado por ser um sistema aberto e evolutivo que permite a interoperabilidade entre diversas fontes de dados.

O desenvolvimento destes ecossistemas está orientado à interação de diversos atores com papéis bem definidos. Möller et al. (2024) and Steinbuss; Turkmayali (2024) identificam no coração desta estrutura, os *Data Providers*, que detêm o controlo e propriedade sobre os dados e os *Data Consumers*, que utilizam os dados para gerar novas visões ou serviços. No entanto, para se fornecerem garantias de sustentabilidade dentro dos ecossistemas, o papel dos *Data Intermediaries* é crítico, sendo da responsabilidade destas entidades assegurar a confiança, a certificação e o cumprimento das regras de governo estabelecidas.

3.1.3 O Desafio da Fragmentação: Silos de Dados e Heterogeneidade

A fragmentação da informação constitui atualmente, um dos obstáculos mais persistentes para concretizar os objetivos de inovação e na economia de dados (Casolari et al., 2023; Harmelink et al., 2024). Este fenómeno manifesta-se predominantemente através da criação de silos de dados, que consistem em repositórios de informação isolados, onde a estrutura e o significado dos dados permanecem ocultos (Auer, 2022; Yan et al., 2024). Como consequência, o acesso e a reutilização de desses dados acabam por estar impedidos por barreiras técnicas, tanto pela heterogeneidade dos sistemas, como pela falta de semântica (Auer, 2022; Susha et al., 2023). Do ponto de vista estratégico, ainda existe uma forte ideia de protecionismo dos dados motivada por receios de apropriação indevida, perdas de vantagem competitiva, e incerteza quanto à sustentabilidade dos modelos de negócio desenvolvidos para a partilha (Harmelink et al., 2024; Jussen et al., 2024; Möller et al., 2024). Segundo Harmelink et al. (2024), a existência destes silos não surge de algo puramente tecnológico, mas frequentemente associado a uma decisão racional das organizações.

No entanto, mesmo quando as barreiras organizacionais são ultrapassadas, a fragmentação do ponto de vista técnico impõe desafios severos à interoperabilidade. Yan et al. (2024) sistematizam este desafio através da exploração do conceito de heterogeneidade, que se constitui em diferentes

níveis de complexidade, como a heterogeneidade sintática, a heterogeneidade estrutural e a heterogeneidade semântica. A heterogeneidade sintática foca-se na diversidade de formatos de ficheiro e protocolos de comunicação, enquanto a heterogeneidade estrutural diz respeito à variação na forma como os esquemas de dados são organizados, mesmo quando partilham a mesma sintaxe. Como é observado por Yan et al. (2024), esta diversidade de fontes e a ausência de normas de recolha uniformes tornam o alinhamento de dados um processo moroso, sem consistência e difícil de automatizar.

Todavia, a barreira mais profunda e complexa reside na heterogeneidade semântica. Como sublinha Auer (2022), a interoperabilidade tecnológica histórica focou-se na conectividade física e sintática, mas falhou frequentemente em garantir que o significado da informação fosse preservado na transição entre sistemas. A heterogeneidade semântica ocorre quando o mesmo conceito é descrito por termos diferentes, a sinonímia, ou quando o mesmo termo possui significados distintos dependendo do contexto de domínio, a polissemia (Auer, 2022; Yan et al., 2024). Tendo por exemplo a temperatura, diferentes sistemas de informação, consideram diferentes tipos de medida como os Celsius, o Fahrenheit, ou o Kelvin quando armazenam os dados. O mesmo quando se descreve ecossistema na ciência de computadores e na biologia. O nome é o mesmo, mas o significado é diferente. Metaforizando, esta "Torre de Babel digital" impede que as máquinas processem e integrem dados de forma autónoma, limitando o valor dos dados, do ponto de vista de extração de significado, a contextos específicos e isolados. De uma perspetiva crítica, com base na literatura analisada, a fragmentação não será resolvida unicamente através do ponto de vista legal de abertura de dados. Como referido anteriormente, enquanto o Data Act é focado na obrigatoriedade do acesso, a realidade do ponto de vista técnico descrito por Möller et al. (2024) demonstra que caso os dados não estejam inseridos num contexto ou tenham significado partilhado, é difícil qualquer extração de valor dos mesmos.

Portanto, a superação dos silos de informação obriga a uma transição para modelos de dados focados na semântica, onde o foco principal passa pela criação de um entendimento comum entre os diferentes atores, ao invés de uma simples troca de documentos, permitindo alcançar a soberania e o valor dos dados, com preservação ao longo de todo o ecossistema.

3.2 Espaços de Dados: Arquitetura, Componentes e Suporte

Partindo da necessidade de superar a fragmentação semântica identificada na secção anterior, esta secção estabelece as fundações do ponto de vista conceptual e de arquitetura dos espaços de dados enquanto infraestrutura de suporte aos ecossistemas de dados. É estabelecida a evolução histórica do conceito, partindo da visão original de integração de Franklin et al. (2005) e Halevy et al. (2006) até à redefinição sociotécnica promovida pela IDSA International Data Spaces Association (2025). São exploradas duas das principais arquiteturas de referência no contexto europeu, a IDS-RAM e a Gaia-X, com comparação das suas abordagens na visão de governo e interoperabilidade (Singh et al., 2025).

Num segundo momento, a análise desce ao nível dos componentes técnicos que operacionalizam a arquitectura IDS-RAM 4.0, aprofundando o papel do IDS Connector, do Metadata Broker e do Vocabulary Hub na viabilização da interoperabilidade semântica (International Data Spaces Association, 2025). A secção conclui com a definição dos atores e papéis funcionais que compõem a camada de negócio do espaço de dados, com clarificando a distinção entre espaço de dados e ecossistema de dados e estabelecendo a base conceptual necessária para a análise da interoperabilidade semântica que se segue.

3.2.1 Conceito e Arquiteturas de Referência (IDS e Gaia-X)

Partindo do pressuposto do ponto anterior, surge a necessidade de estabelecer, de forma clara e inequívoca, o conjunto de conceitos, atores, papéis e componentes, sendo vistos como a fundação para esta dissertação. Por outras palavras, este ponto parte da definição do conceito de "Espaço de Dados" (*Data Space*).

O presente conceito não é algo recente, mas sim o resultado de uma evolução conceptual que remonta à gestão de informação no inicio do presente século. Originalmente introduzido por Franklin et al. (2005), no artigo *From databases to dataspaces* e consolidado por Halevy et al. (2006), o termo surgiu como uma visão alternativa à integração de dados tradicional, baseada em armazéns de dados centralizados. Na altura, o objetivo inicial da infraestrutura de espaços de dados era permitir o acesso a fontes de dados heterogéneas sem a necessidade de uma integração física e semântica completa e imediata, sendo proposta uma abordagem de integração incremental e a pedido *pay-as-you-go*(Halevy et al., 2006).

No entanto, na última década, o conceito passou por uma redefinição profunda, impactada pela necessidade de alinhar confiança e autonomia no mesmo espectro da economia digital na indústria. Com base nos trabalhos de Otto (2022) e da International Data Spaces Association (2025), o espaço de dados evoluiu de um conceito de base de dados distribuída, da definição de Halevy et al. (2006), para uma visão de infraestrutura com foco predominantemente sociotécnico.

Em linha com Möller et al. (2024), a definição moderna de espaço de dados descreve uma infraestrutura projetada para garantir a soberania de dados, sendo operacionalizada por quatro propriedades fundamentais: a natureza distribuída, na qual os dados não são centralizados, permanecendo na sua fonte original; A abertura, que viabiliza a entrada livre de qualquer participante, desde que devidamente certificado; A segurança, que assegura a confiança no ecossistema através do cumprimento de requisitos técnicos rigorosos de comunicação e identidade (Pettenpohl et al., 2022). E por fim o controlo de utilização. Este último mecanismo atua de forma centrada nos dados, ultrapassando as simples restrições de acesso, através da imposição e especificação técnica em relação à forma como a informação deve ser operada, agregada ou reencaminhada após o acesso inicial ter sido concedido ao participante (Otto, 2022; Möller et al., 2024).

Em concordância com a definição de espaços de dados, através dos estudos identificados, é exigida uma maior perceção sobre o estado de maturidade dos espaços de dados, sendo esta evidenciada em dois modelos de referência com destaque no contexto europeu: o International Data Spaces (IDS) e o Gaia-X. Segundo a análise comparativa de Singh et al. (2025), através de

um mapeamento profundo das principais arquiteturas de referência, enquanto o IDS está focado essencialmente na soberania e no controlo de utilização durante uma transação direta de dados, o Gaia-X expande o conceito para uma visão de infraestrutura em nuvem, objetivando-se na federação de serviços, com destaque no armazenamento e processamento.

Feature	**IDS-RAM**	**Gaia-X**
Tipo de Arquitetura	Descentralizado	Federado
Standardização	Baseado em padrões abertos e focado na promoção da interoperabilidade	Quadro de conformidade (soberania dos dados, interoperabilidade)
Modelo de Governo	Execução da política IDS, perfis ODRL, mecanismo de política eDC	Federado, Serviço de conformidade, Registo Gaia-X
Escalabilidade	Altamente escalável, com uma arquitetura modular que se adapta a várias organizações	Alcançado através da arquitetura distribuída. Disponível para implementações na cloud/edge e serviços de containerização
Armazenamento de Dados	Descentralizado, mas com controlos rigorosos	Federado, dependente de fornecedores
Acesso aos Dados	Seguro, vinculado por contrato	Baseado em certificações, federado
Troca de Dados	Contratual, prioridade à conformidade	Orientado por metadados e com governança rigorosa

Table 3.1: Comparação entre IDS-RAM e Gaia-X

A comparação entre as arquiteturas IDS-RAM e Gaia-X, sintetizada na Tabela 3.1 e adaptada de Singh et al. (2025), revela duas abordagens complementares mas distintas para o problema da partilha de dados no contexto europeu. Enquanto a IDS-RAM adota um modelo descentralizado com foco na soberania ao nível da troca de dados, garantindo que cada troca é regulada através de contratos e políticas de utilização aplicadas automaticamente pelo *IDS Connector*, a Gaia-X expande esta visão para uma infraestrutura federada com maior escalabilidade, orientada para a gestão de serviços em nuvem e para a criação de um ecossistema de conformidade multi-ator (Singh et al., 2025). Esta distinção é particularmente relevante para a presente dissertação, já que a IDS-RAM opera ao nível da troca de dados entre participantes, sendo o modelo de referência mais utilizado para a instanciação de espaços de dados, onde o foco está na semântica e no governo da informação dentro de um domínio específico (International Data Spaces Association, 2025; Kovach et al., 2026). A Gaia-X, por sua vez, atua predominantemente a um nível de abstração mais alto, fornecendo o enquadramento da conformidade e federação entre ecossistemas, mas sem descer ao nível do conteúdo semântico das trocas.

Como observam Singh et al. (2025), a principal lacuna identificada em ambas as arquiteturas reside precisamente na ausência de mecanismos leves e acessíveis para pequenas e médias organizações, o que reforça a necessidade de um processo que governe a semântica ao nível do setor.

Esta distinção é fulcral para demonstrar a evolução do próprio conceito. Na perspetiva inicial de espaços de dados, introduzida por Franklin et al. (2005), era proposto um modelo descentralizado para a gestão e coexistência de fontes de dados heterogéneas, opondo-se à rigidez da integração tradicional. Contudo, a diferença mais contrastante face às conceptualizações atuais reside no objetivo proposto e nas garantias da infraestrutura, com a visão moderna a evoluir da resolução de problemas de integração técnica, para se assumir como uma infraestrutura sociotécnica estruturada em camadas assentes na confiança, governação e soberania de dados (Otto, 2022; Möller et al., 2024). Deste modo, a infraestrutura atual assegura a proteção e o controlo de utilização desde a fronteira técnica do conector até à extensão do ecossistema de partilha.

Conferindo profundidade na perspetiva da arquitetura, na evolução dos espaços de dados e na visão semântica a abordar no ponto 3.3.1, é crucial o conhecimento de forma detalhada da arquitetura *IDS Reference Architecture Model*, a IDS-RAM (International Data Spaces Association, 2025). Historicamente, a ocorrência da partilha entre organizações operacionalizada através de plataformas digitais centralizadas, o que desencadeava problemas como dependência e falta de soberania sobre a informação partilhada. Para colmatar esta falha, Pettenpohl et al. (2022), explicam que o IDS-RAM fornece uma arquitetura padronizada para que as organizações integradas num ecossistema de negócio possam trocar dados ponto a ponto. Neste contexto, o espaço de dados na visão contemporânea foi concebido como um ambiente virtual agnóstico em relação à tecnologia, no qual a confiança é assegurada do ponto de vista técnico, através da execução automática das políticas de controlo de utilização e pelas regras rígidas de certificação para os participantes. Neste modelo, os dados não são centralizados, com o fornecedor a manter o controlo em relação aos dados (Auer, 2022; Pettenpohl et al., 2022).

Para evitar ambiguidades conceptuais, é imperativo estabelecer uma distinção clara entre ecossistemas e espaços de dados. Por um lado, um ecossistema de dados é uma *"rede sociotécnica complexa de atores autónomos que colaboram e interagem para extrair valor e inovar a partir da partilha de dados"* (Möller et al., 2024; Curry et al., 2022). Por outro lado, um espaço de dados é a infraestrutura digital descentralizada que operacionaliza esse ecossistema, fornecendo os mecanismos técnicos e de governação necessários para garantir uma partilha de dados segura, confiável e soberana (Möller et al., 2024; Kernstock et al., 2025).

De uma visão crítica, a literatura aponta para o facto de o termo espaço de dados ter sofrido um fenómeno de saturação semântica, sendo frequentemente confundido com o próprio ecossistema (Möller et al., 2024). Esta imprecisão terminológica pode gerar expectativas irrealistas em termos de instanciação e implementação. Sem esta clareza conceptual, através da distinção entre a rede de negócio e a infraestrutura, a implementação de mecanismos de soberania para instanciar os espaços de dados corre o sério risco de passar para segundo plano, em prol de uma conectividade superficial que não resolve os problemas de confiança e de governo identificados anteriormente, perpetuando uma ambiguidade conceptual (Otto, 2022; Möller et al., 2024).

Em síntese, o conceito de espaço de dados é essencial para viabilizar e escalar a partilha de dados interorganizacional. Neste contexto, o suporte tecnológico fornecido por estas infraestruturas assume um papel relevante na garantia da confiança nas transações estratégicas e ao assegurar a

soberania de dados. Deste modo, é possível às organizações participarem no ecossistema de forma colaborativa e autónoma, sem perderem o controlo sobre a sua informação.

3.2.2 Componentes Técnicos Essenciais: O Papel do Connector, do Vocabulary Hub e do Metadata Broker

A definição da arquitetura IDS-RAM 4.0 estabelece a base para a definição de espaços de dados a adotar em específico. Para materializar a arquitetura, é crucial a perceção de todas as camadas que compõe o modelo de referência. Este ponto estará dedicado à camada do sistema, à qual correspondem os componentes que operacionalizam o sistema. Observemos esta camada como um conjunto de elementos integrados em sintonia num sistema para operacionalizar a informação e permitir a interação entre utilizadores. Esta camada representa o software desse sistema. Como principais componentes a abordar nesta dissertação, e tendo como base o artigo de Pettenpohl et al. (2022) e a International Data Spaces Association (2025), este ponto pretende aprofundar em específico o *IDS Connector*, o *Metadata Broker* e o *Vocabulary Hub*, componentes centrais da arquitetura IDS-RAM 4.0. Essencialmente, estes três componentes permitirão ultrapassar o desafio semântico, abordado no ponto 3.3.1.

Detalhadamente, o *IDS Connector* é o componente técnico central na arquitetura IDS-RAM 4.0, que garante a soberania dos dados, atuando como uma interface entre os sistemas internos de um participante e o espaço de dados. O *IDS Connector* é responsável por executar a troca de dados, aplicando políticas de utilização e garantindo a identidade através de certificados e tokens dinâmicos. Ao nível da arquitetura, o *IDS Connector* está estruturado pelos *Core Containers*, responsáveis pela comunicação e execução de políticas do sistema, os *App Store Containers* que alojam aplicações de dados certificadas e os *Custom Containers*. Estes últimos permitem que o próprio operador do conector implemente software proprietário que não se encontra na *App Store*, sendo essenciais para estender as capacidades do conector com a lógica de negócio específica da organização, garantindo que o tratamento da informação ocorre o mais próximo possível da sua fonte original (Pettenpohl et al., 2022).

Kirstein et al. (2022), considerando o movimento de dados abertos atual, destacam a vantagem de utilizar a capacidade nativa de fluxos *Push* dos *IDS Connectors*. Em vez do modelo tradicional de dados abertos, onde os portais centrais recolhem ativamente a informação de diversas fontes num modelo *Pull*, a adoção do conector permite aos fornecedores informarem automaticamente os *Brokers* sobre a disponibilidade ou alteração dos seus dados. Esta inversão de paradigma viabiliza atualizações em tempo real confere aos fornecedores de dados públicos um controlo soberano sobre a disponibilização da sua informação. Mais recentemente, em 2025, Singh et al. (2025) observa que as implementações atuais podem ser complexas, argumentando que existe uma necessidade de versões mais leves e fáceis de implementar para pequenas e médias empresas, garantindo, ao mesmo tempo, segurança e interoperabilidade. Já Hmimou et al. (2025) apontam que os *connectors* atuais têm o seu foco principalmente no controlo de acesso e pouco na validação da qualidade e conteúdo dos dados em si, propondo assim, a integração de *Policy Checkers* que geram, de

forma automática, um código para validar se o conteúdo dos dados cumpre as regras negociadas, conferindo assim transparência e rastreabilidade.

Outro componente central nesta arquitetura é o *Metadata Broker*, que atua como um catálogo central onde os *Data Providers* publicam as descrições dos seus dados e os *Data Consumers* pesquisam e descobrem os recursos disponíveis. O *Metadata Broker*, armazena apenas os metadados, não os dados propriamente ditos, facilitando, assim, o encontro entre a oferta e a procura. De acordo com a International Data Spaces Association (2025), nas definições da camada do sistema, o *Metadata Broker* é um conector IDS que contém um endpoint para registo, manutenção, publicação e consulta de autodescrições. Nas principais melhorias identificadas neste componente, foram identificadas diferentes perspetivas sobre o qual, o *Metadata Broker* poderá ser melhorado. Kirstein et al. (2022) identificaram o desenvolvimento de um *Open Data Broker*, considerando a integração com dados abertos, de forma a fornecer interfaces mais amigáveis na navegação na web por parte dos utilizadores, semelhantes a portais de dados abertos como o CKAN[1], mas que no *backend* seja aplicada a linguagem da IDS, permitindo, assim, uma ponte entre o público em geral e o espaço de dados industrial.

Nizamis et al. (2024) propõem a utilização do *broker* com uma semântica específica para a indústria. Os autores argumentam que os *brokers*, numa visão genérica, falham na captura de especificidades de domínios complexos como a manufatura. Propõem, assim, um *Metadata Broker* Evoluído, baseado numa nova ontologia como o modelo Data Sp4ce Metadata Registry (*DS4MR*), que permite descrever detalhadamente processos de qualidade, tipos de uso e formatos específicos, com uma interface de utilizador para registo semântico avançado. Já numa visão dentro do contexto dos metadados em si, a revisão sistemática de Tonnarelli et al. (2025), na exploração de princípios de catalogação, bem como as ferramentas de gestão de metadados, é importante para a exploração de melhorias dentro do *Metadata Broker*. Tonnarelli et al. (2025) identificaram nesta revisão lacunas significativas no processo de colaboração social.

Sem abordarem em específico o *Metadata Broker*, mas utilizando o exemplo dos catálogos de metadados sugerem a evolução e inclusão de funcionalidades de avaliação, revisão e comunicação direta entre os utilizadores, neste caso, os *Data Providers* e os *Data Consumers*. Além deste aspeto, os autores propõem a utilização de IA para automatizar o processo de extração de metadados e enriquecimento contextual, transformando o catálogo numa ferramenta proativa. Apesar de o *Metadata Broker* e o catálogo não terem o mesmo papel, Kirstein et al. (2022) estabelecem que num ecossistema de dados abertos baseado na arquitetura IDS, o *Metadata Broker* assume o papel de um tipo de catálogo de dados, agindo como uma entidade central para distribuição de dados. Sendo assim, ambos têm o foco em resolver o problema da fragmentação de dados, funcionando como ponto central no acesso aos metadados.

Por fim, importa destacar o *Vocabulary Hub*, um dos vários componentes intermediários que, em conjunto com os *connectors* e serviços, permitem a operacionalização dos espaços de dados. A essência deste componente foca-se estritamente na garantia da interoperabilidade semântica.

[1] https://ckan.org/

Na visão da Ciência da Informação, este serviço atua como um repositório e gestor de ativos semânticos como ontologias, fundamentado em normas da *Semantic Web*, nomeadamente RDF e OWL (David et al., 2025; Steinbuss; Turkmayali, 2024). A sua função nuclear do *Vocabulary Hub* visa garantir que os sistemas atribuam um significado exato e inequívoco às entidades de dados trocadas, independentemente do formato sintático. A base central para este componente na é o *IDS Information Model*, uma ontologia estruturada em RDFS/OWL que padroniza as descrições de recursos (Bader et al., 2020).

Este modelo é enriquecido no *Vocabulary Hub* com vocabulários de domínio específico, aplicando princípios de *Linked Data* que permitem não só a anotação rigorosa dos metadados, mas também o raciocínio automático e a descoberta avançada de dados por todos os participantes do ecossistema (Steinbuss; Turkmayali, 2024).

A um nível mais técnico, o *Vocabulary Hub* fornece uma plataforma para hospedar, manter, publicar e documentar os vocabulários adicionais, bem como os endpoints em conformidade com o IDS, para permitir uma comunicação com *IDS Connectors* e outros componentes da infraestrutura. Além disso, permite o acesso aos termos definidos e às descrições, apresentando as mudanças e diferentes versões. Atuam, assim, como plataformas de gestão para esquemas de dados que podem ser utilizadas em diferentes casos de uso (International Data Spaces Association, 2025). Numa visão mais atual, David et al. (2025) criticam a visão do *Vocabulary Hub* apenas como um armazém de ontologias, propondo uma elevação neste componente para um serviço de camada semântica *Semantic Layer Service*. Este serviço não apenas armazena o vocabulário, mas também estabelece uma vinculação dinâmica entre as descrições semânticas dos dados recebidos em tempo real. Isso permite harmonizar dados de diferentes fontes sem a necessidade de processos ETL (Extração, Transformação e Carregamento) complexos, utilizando tecnologias como OBDA (*(Ontology-Based Data Access*) e mapeamentos R2RML *(RDB to RDF Mapping Language)*, melhorando assim o *Vocabulary Hub* de um repositório passivo para uma camada semântica ativa.

Ainda neste artigo, David et al. (2025) incluem na proposta a capacidade de anotar automaticamente dados não estruturados (textos, relatórios) usando NLP (*Natural Language Processing*) com base em grafos, algo que a IDS-RAM 4.0, na descrição do componente, não detalha de forma aprofundada. Já com base nas versões 1.0 e 1.1 do IDSA Position Paper Steinbuss; Turkmayali (2024) and Steinbuss; Almeida, et al. (2025), os autores reforçam que, do ponto de vista dos modelos semânticos, além de se fornecer o modelo, é crucial estabelecer processos de governo para a evolução dos vocabulários dentro das instâncias dos espaços de dados, garantindo manutenção de *"Common Semantic Data Models"* que vão além do modelo base da arquitetura IDS-RAM 4.0.

3.2.3 Atores e Papéis Funcionais-Data Provider, Data Consumer, Data Intermediary e Vocabulary Provider

Na secção anterior, a principal abordagem esteve focada da camada de sistema da arquitetura IDS-RAM 4.0, através da definição dos principais componentes que operacionalizam o sistema, focando essencialmente nos componentes intrinsecamente associados ao contexto semântico de

interoperabilidade. A presente secção fará a transição para a camada de negócio e para a perspetiva de governação do ecossistema. Através da definição dos principais atores e dos seus respetivos papéis na arquitetura IDS-RAM 4.0, é possível verificar que a atribuição de papéis transcende o estabelecimento de modelos de negócio, constituindo o alicerce da governação do espaço de dados, essencial para garantir a confiança, a conformidade e a responsabilidade nas transações (Hmimou et al., 2025; Castellvi et al., 2025).

A operacionalização de um espaço de dados exige uma definição clara de papéis funcionais que transcende a garantia de descentralização e confiança. Mais do que assegurar uma infraestrutura segura, a delimitação formal de papéis, é o pilar sociotécnico que orquestra a complexidade do ecossistema e viabiliza a co-criação de valor através de novos modelos de negócio. Adicionalmente, esta estrutura de papéis é imperativa para suportar ativamente a interoperabilidade semântica e técnica dos dados, estabelecendo em simultâneo os limites de responsabilidade legal e a conformidade legal essencial às transações inter organizacionais.

Neste contexto, a arquitetura IDS-RAM (International Data Spaces Association, 2025) tem definido esse conjunto de elementos que permitem este funcionamento interoperável e com segurança. Segundo Pettenpohl et al. (2022) a arquitetura IDS-RAM 4.0 classifica os papéis e atores em quatro categorias principais. O presente trabalho foca-se essencialmente nas duas primeiras categorias, os Participantes Chave e os Intermediários. Esta delimitação justifica-se pelo facto de serem estes os atores ativamente envolvidos na execução operacional da troca de dados.

Nesta primeira categoria, os atores são representados na principal proposta de valor dos espaços de dados, a troca interoperável e soberana de dados entre diferentes participantes, com o intuito de fomentar a inovação e a co-criação de valor no ecossistema.

Iniciando pela perspetiva da propriedade e de quem fornece, são definidos o *Data Owner* e o *Data Provider*. O *Data Owner* é definido como o "dono dos dados", sendo, neste caso, e de forma mais concreta, a pessoa ou entidade que detém os direitos e o controlo dos dados do ponto de vista legal. Este ator é responsável por definir os termos de utilização dos dados, como as formas de pagamento, políticas de utilização e garantia de acesso. Em linha com o ator anterior, o *Data Provider* tem a função de estabelecer a troca de dados entre o *Data Owner* e o *Data Consumer*, através dos componentes de software abordados no ponto 3.2.2. O *Data Owner* pode em muitos casos assumir o papel de *Data Provider*.

Na perspetiva de quem utiliza e de quem recebe os dados, estão definidos o *Data User* e o *Data Consumer*. O *Data Consumer* é o responsável, dentro do processo, por receber os dados, funcionando como um ator-espelho ao *Data Provider*, no qual um fornece e o outro recebe. Este realiza atividades semelhantes ao *Data Provider*, sendo que, antes de ser estabelecida qualquer conexão, o *Data Consumer* pode pesquisar diferentes conjuntos de dados através do *Metadata Broker*. Seguindo esta perspetiva de espelho, o *Data User* pode ser visto com atividades semelhantes às do *Data Owner*. Neste caso, o *Data User* irá representar a entidade legal de utilização dos dados, com base na política de utilização, podendo assumir por vezes funções como *Data Consumer* no estabelecimento do processo de troca de dados. Esta visão permite estabelecer, de forma clara, a diferença entre os atores dentro dos espaços de dados, conferindo evolução à visão

do ponto de vista do ecossistema de dados.

Nesta categoria, ainda é essencial destacar o *App Provider*, que é responsável pelo desenvolvimento de aplicações para possibilitar a troca de dados. As aplicações têm de estar em conformidade com a arquitetura de referência, sendo essencial uma descrição de cada aplicação com metadados, no que respeita à semântica, funcionalidade e interface. Para aumentar a confiança, as aplicações podem ser certificadas por um mecanismo de certificação no próprio IDS (International Data Spaces Association, 2025).

Na segunda categoria, definida por Pettenpohl et al. (2022), podemos encontrar os participantes com papel de intermediários. Se na primeira categoria o foco era os atores que atuam à camada mais superficial do espaço de dados, que usufruem do espaço de dados para trocarem dados entre si, os intermediários representam os responsáveis por orquestrar o espaço de dados e permitir as trocas de dados entre os diferentes atores. Com o objetivo de alcançar a interoperabilidade semântica, o papel desempenhado por estes atores é considerado central, sendo desempenhado exclusivamente por entidades de confiança.

Pettenpohl et al. (2022) identifica, em detalhe nesta categoria, além dos componentes discutidos na secção anterior, o *Identity Provider* e o *Vocabulary Provider*. O *Identity Provider*, do ponto de vista do acesso, é o responsável pela manutenção, gestão e monitorização da identidade e da informação de cada participante no espaço de dados. Já o *Vocabulary Provider* faz a gestão dos vocabulários que podem ser utilizados para as descrições dos conjuntos de dados. As principais funções deste ator assentam no fornecimento de um repositório central para esquemas e informação de vocabulário, ferramentas de suporte para alterações colaborativas, criação e manutenção do arquivo de vocabulários, bem como em possibilitar a conexão entre a transferência de dados entre o conector e a informação do vocabulário. O papel deste ator, intermediário dentro do espaço de dados, é crucial para se estabelecer a interoperabilidade semântica.

A terceira categoria está essencialmente relacionada com as questões de software e serviços, compreendendo as empresas de TI que fornecem o suporte tecnológico e os serviços para facilitar a adoção do modelo IDS. Nesta categoria, temos o *Software Provider*, responsável por fornecer o software fundamental necessário para implementar as funcionalidades IDS em linha com a arquitetura. Já o *Service Provider* faz a hospedagem das infraestruturas técnicas para participantes que não as desejam manter localmente, bem como oferece serviços de valor acrescentado, como análise de dados, integração e limpeza de dados.

A quarta categoria representa as entidades de governo, com o papel de supervisionar e garantir a certificação de todo o espaço de dados. Aqui é possível encontrar a IDSA (International Data Spaces Association, 2025), que promove o ecossistema, orientando o desenvolvimento contínuo da arquitetura através de grupos de trabalho, acolhendo diferentes empresas e centros de investigação. Já as entidades de certificação são responsáveis por realizar a certificação rigorosa dos participantes e dos componentes técnicos do IDS, estabelecendo a base de confiança e segurança essencial para as transações no ecossistema.

Em síntese, a definição dos espaços de dados, em detalhe na arquitetura de referência IDS-RAM 4.0, aqui definida nas camadas de sistema e na camada de negócio, permite esclarecer e

clarificar as principais ambiguidades dentro deste tema, permitindo adicionalmente abrir caminho para a secção seguinte. Esta revisão servirá de base para abordar o problema central a que esta dissertação pretende responder, mais concretamente a interoperabilidade semântica.

3.3 Interoperabilidade Semântica: O Motor da Comunicação

Esta secção aborda o elemento nuclear desta dissertação, examinando a interoperabilidade semântica como condição necessária para a operacionalização plena dos espaços de dados. Partindo do modelo multidimensional de interoperabilidade definido pelo European Interoperability Framework e pela framework Technical, Semantical, Legal, Organizational (TSLO), é estabelecida a distinção entre os níveis técnico, semântico, legal e organizacional, com ênfase particular na relação de dependência entre a camada sintática e a semântica (Cimmino et al., 2024; Singh et al., 2025).

São analisados os artefactos semânticos de suporte, como ontologias, vocabulários controlados e catálogos, complementando com as tecnologias de suporte como *Knowledge Graphs* e *Linked Data*, que permitem formalizar o significado dos dados em ambientes distribuídos. A secção aprofunda ainda o problema da dissonância ontológica em ambientes federados, identificando as causas estruturais do conflito entre modelos semânticos de diferentes domínios, explorando as abordagens emergentes para a sua mitigação (Stäbler et al., 2023). As lacunas identificadas convergem na necessidade de um processo que governe a semântica ao nível setorial, fundamentando a questão de investigação desta dissertação.

3.3.1 Níveis de Interoperabilidade: Da conexão técnica ao entendimento semântico

A interoperabilidade num espaço de dados é um conceito multidimensional que se manifesta em diferentes camadas de abstração. Para que a colaboração ocorra, não basta que os sistemas estejam fisicamente ligados; é imperativo que o significado da informação trocada permaneça íntegro e inequívoco para todos os intervenientes. Segundo a framework *Technical, Semantical, Legal, Organizational* (TSLO) desenvolvida por Ali et al. (2024) e baseada no *European Interoperability Framework*(EIF), a interoperabilidade deve contemplar quatro dimensões interdependentes: a dimensão Técnica, focada nas aplicações, interfaces e protocolos seguros que ligam os sistemas; a dimensão Semântica, que assegura que o significado exato da informação trocada é preservado e inequivocamente compreendido por todas as partes; a dimensão Legal, que garante a colaboração entre organizações através da harmonização de políticas de dados, licenciamento e regulamentações; e, por fim, a dimensão Organizacional, referente ao alinhamento dos processos de negócio, estruturas de governação e responsabilidades para alcançar objetivos mutuamente benéficos. Neste modelo, a camada técnica fornece a infraestrutura de comunicação, enquanto a camada semântica garante que o conteúdo dos dados trocados são interpretados corretamente pelos sistemas recetores. É a articulação simultânea destas quatro camadas que pavimenta o caminho para a criação sustentável de ecossistemas de dados operacionais e de elevado valor.

Cimmino et al. (2024) and Ganzha et al. (2021) estabelece uma distinção clara entre a interoperabilidade sintática e a semântica. Como observam Ganzha et al. (2021), a interoperabilidade técnica ou sintática lida com a capacidade de dois sistemas trocarem dados num formato comum (como JSON, XML ou RDF). Contudo, a conformidade sintática apenas assegura que os bits e bytes são lidos sem garantir que a lógica do negócio ou o contexto dos dados seja preservado. É nesta lacuna que a interoperabilidade semântica atua, sendo definida como a capacidade dos sistemas interpretarem o significado dos dados de forma automática e precisa, permitindo a integração de recursos provenientes de fontes heterogéneas (Steinbuss; Turkmayali, 2024; Auer, 2022). A interoperabilidade semântica é o fenómeno que permite, ao longo de uma comunicação, manter a compreensão do significado, seja do ponto de vista social ou tecnológico. Diferentemente dos restantes níveis de interoperabilidade, a semântica é intrinsecamente ligada ao fator social, bem como, em grande parte, ao conceito de gestão da informação e à interpretação de significado em sociedade.

Informação, na mesma perspetiva dos dados, toma valor quando é comunicada e compreendida pelos diferentes atores. No contexto tecnológico atual, em que a partilha de dados ocorre frequentemente entre organizações de diferentes setores e países, a incapacidade de comunicar e interpretar corretamente o significado da informação continua a representar um obstáculo crítico. A literatura demonstra que a ausência de interoperabilidade semântica resulta em perdas financeiras e operacionais avultadas, a nível macroeconómico (Vogt et al., 2025).

Dentro dos espaços de dados, e de forma mais ampla dentro dos ecossistemas, o problema do ponto de vista comunicacional permanece. A União Europeia, tem definidas na sua estratégia diferentes níveis de interoperabilidade, referidos na framework de Ali et al. (2024). Na visão dos espaços de dados, é essencial estabelecer a distinção e a relação de dependência entre as dimensões técnica e semântica. Steinbuss; Almeida, et al. (2025) clarificam que a interoperabilidade sintática atua como um subcomponente do nível Técnico do EIF, tendo a função estrita de garantir que os dados são lidos num formato gramatical conhecido pelas máquinas. Contudo, o alinhamento sintático não resolve a dissonância de significados. É sobre esta base estrutural que atua o nível de interoperabilidade semântica, sendo o responsável exclusivo por garantir a interpretação e a compreensão unívoca da informação trocada entre os diferentes atores do espaço de dados (Steinbuss; Almeida, et al., 2025).

O trabalho de Cimmino et al. (2024) é crucial para esta distinção. A interoperabilidade sintática representa a estrutura dos dados, mais precisamente os protocolos de comunicação, formatos de arquivo e regras gramaticais, tendo como objetivo permitir uma troca física e lógica dos dados. Já a interoperabilidade semântica é responsável pelo significado, contexto e terminologias, tendo como objetivo permitir a interpretação de forma precisa do conteúdo, sem perda de significado. A principal lacuna que esta dissertação pretende resolver reside na dificuldade prática de alinhar estes significados num ambiente descentralizado. Abordando o problema no prisma da gestão de informação, este trabalho foca-se na dimensão sociotécnica da interoperabilidade, reconhecendo que a resolução dos conflitos semânticos exige não apenas o uso de infraestruturas tecnológicas padronizadas, mas também a implementação de modelos de governo e processos colaborativos

que alinhem os diferentes contextos organizacionais (Steinbuss; Almeida, et al., 2025).

A obtenção de um consenso absoluto sobre o significado da informação constitui um dos principais desafios práticos na implementação da interoperabilidade semântica. Esta dificuldade parte da elevada heterogeneidade inerente aos ecossistemas descentralizados, onde os participantes possuem visões distintas do mundo, motivadas por operarem de forma diferente do ponto de vista legal, dos setores de atividade e dos mercados, o que resulta em requisitos e processos de negócio incompatíveis entre si. Parte dos atores, dentro da literatura, aponta para os standards como artefacto central para a interoperabilidade (Steinbuss; Almeida, et al., 2025).

Adicionalmente, a literatura demonstra que o desenvolvimento de um vocabulário comum e universal é tipicamente inviável, uma vez que diferentes comunidades aplicam diferentes quadros de referência, enfatizando aspetos distintos da mesma realidade consoante o seu contexto de uso. Para contornar esta inevitável ausência de consenso inicial e a consequente heterogeneidade dos domínios, Cimmino et al. (2024) apontam a adoção de artefactos formais com um processo de como desenvolver e reutilizar as ontologias, bem como explorar os princípios de *Linked Data*.

No que concerne à gestão semântica do ecossistema, a estratégia não deve focar-se na criação de novos modelos de raiz, mas sim na reutilização e na melhoria contínua dos recursos existentes (Cimmino et al., 2024). Para promover a interoperabilidade nos espaços de dados, Cimmino et al. (2024) propõe que o desenvolvimento ontológico consista em três diretrizes práticas. Primeiro, deve priorizar a reutilização de ontologias padronizadas,ou seja, modelos que adquiriram o estatuto de padrão devido à sua ampla adoção pela comunidade ou por serem formalmente suportados e promovidos por organismos standard. Segundo, a modelação deve basear-se em definições conceptuais gerais, evitando impor restrições demasiado rígidas, de modo a atuar como um modelo de alto nível que permite aos participantes adaptarem os conceitos com base no seu contexto específico. Terceiro, deve focar-se num vocabulário mínimo e na utilização de termos de ligação evitando o excesso de especificidade no próprio modelo, o que possibilita o estabelecimento de pontes ágeis para outras ontologias externas sem impor dependência restritiva do ponto de vista tecnológico.

Num ponto importante, devem existir sugestões de modo a complementar o vocabulário, sendo que aqui, quem participa no ecossistema assume o papel de contribuir com o seu conhecimento de domínio, sugerindo a adição de novos termos e propriedades com base nas suas necessidades práticas de partilha. Para garantir que este enriquecimento semântico não comprometa a consistência dos modelos, o processo exige uma curadoria rigorosa suportada pelo *Vocabulary Hub*. A curadoria permite conferir coerência ao processo, evitando a dissonância ontológica e ao mesmo motivar a participação ativa dos diferentes atores ao evidenciar que os modelos evoluem para responder às suas necessidades reais (Bader et al., 2020; Steinbuss; Almeida, et al., 2025)

Outro ponto fundamental para garantir a compreensão inequívoca e a correta aplicação dos modelos é a documentação rigorosa das ontologias esta documentação deve assegurar uma dupla legibilidade: enquanto o código é processado por máquinas, deve ser gerada uma versão interpretável por humanos que detalhe as classes, as propriedades e as restrições lógicas (Cimmino et al., 2024). Além disso, as boas práticas recomendam a inclusão de exemplos práticos de in-

stanciação e a adição de metadados ricos que descrevam a proveniência, as licenças e o histórico da ontologia. Num espaço de dados, a gestão destas características é assegurada pelo *Vocabulary Hub*, que atua como um catálogo centralizado com o propósito não apenas de alojar os artefactos semânticos, mas de gerir as suas versões e garantir a sua descoberta, acesso e preservação a longo prazo por todos os participantes do ecossistema (Corcho et al., 2024; Chirvasuta et al., 2025)

Outros aspetos chave, têm por base demonstrações de aplicabilidade de ontologia, através de exemplos de utilização do RDF (*Resource Description Framework*) para guiar os utilizadores.

Um aspeto central na governação dos espaços de dados é a interoperabilidade semântica, operacionalizada através da relação estrita entre metadados e ontologias. Esta relação ocorre a dois níveis fundamentais: por um lado, as ontologias fornecem a estrutura lógica, como os predicados, classes, objetos e vocabulários controlados, necessária para que os metadados descrevam os dados de forma legível e inequívoca por máquinas; por outro lado, numa perspetiva de reutilização e partilha, as próprias ontologias necessitam de ser descritas por metadados. É a adição de metadados às ontologias que permite o seu registo em catálogos, como os *Linked Open Vocabularies* (Linked Open Vocabularies (LOV)), fornecendo pontos de contacto essenciais, declarações de licenças de uso e o histórico de versões. Deste modo, para garantir a eficácia no processo de publicação num ecossistema, os dados e os modelos semânticos que os suportam devem ser completos, consistentes, credíveis e devidamente atualizados (Hmimou et al., 2025; Steinbuss; Almeida, et al., 2025).

3.3.2 Artefactos Semânticos e Tecnologias de Suporte: Dos Vocabulários, Catálogos e Ontologias aos Knowledge Graphs e Linked Data

O impacto da interoperabilidade semântica nos espaços de dados e no estabelecimento de ecossistemas de dados está diretamente associado à utilização e operacionalização de artefactos. Na perspetiva semântica, os artefactos semânticos são definidos como formalizações legíveis e acionáveis através de máquinas de uma conceptualização, com o propósito de permitir a partilha e a reutilização de informação tanto por humanos como por sistemas informáticos (Corcho et al., 2024). O contributo de Corcho et al. (2024), através da avaliação de catálogos de artefactos semânticos, revelou-se fundamental para o entendimento destes artefactos, que assumem uma ampla variedade de formas, operando em diferentes níveis de abstração e de infraestrutura para assegurar a interoperabilidade semântica.

Portanto, este ponto teve o objetivo de explorar e identificar os artefactos e as tecnologias de suporte principais referenciados pelos autores, alinhando esta visão com os espaços de dados e com IDSA (International Data Spaces Association, 2025). Segundo a perspetiva de Cimmino et al. (2024), uma ontologia é definida como a especificação formal e explícita de uma conceptualização partilhada. As ontologias capturam a semântica e o significado dos dados, servindo de base para o estabelecimento de espaços de dados entre diferentes organizações. As ontologias podem ser categorizadas em *Lightweight*, focando-se em taxonomias compostas e hierarquias, *Heavyweight*, que representam uma extensão das *Lightweight*, com a inclusão formal de axiomas e restrições ao

nível lógico com o objetivo de inferir novo conhecimento e verificar a consistência dos modelos. A título de exemplo, o Modelo de Informação IDS, de acordo com Bader et al. (2020) e a International Data Spaces Association (2025), é representado por uma ontologia RDFS/OWL, com o objetivo de descrever as interações entre papéis, atores e componentes dentro do espaço de dados. Nos espaços de dados, o RDF e o OWL são fundamentais por não exigirem que todos partilhem o mesmo esquema de dados, o que permite um representação de dados alinhada de uma perspetiva holística.

Neste universo dos artefactos, para além das ontologias, os vocabulários e terminologias são concebidos para serem formalizações *machine readable* que permitem a partilha e reutilização de dados (Corcho et al., 2024), sendo compostos por definições legíveis por humanos e especificações de diferentes significados. Os vocabulários e ontologias podem ser altamente especializados, sendo destacados por Steinbuss; Almeida, et al. (2025) pela especificidade que conferem a um determinado domínio na captura de conhecimento e pela importância do ponto de vista legal e contratual para aplicação das políticas de utilização de forma automática.

Para que a reutilização dos vocabulários e terminologias seja verdadeiramente eficaz nas trocas de dados nos espaços de dados, eles necessitam de ser armazenados em infraestruturas específicas, como é o papel do *Vocabulary Hub*, e mais essencialmente dos catálogos. Os catálogos como artefactos semânticos, segundo Corcho et al. (2024), são definidos como *"sistemas web com o objetivo de promover a disponibilidade, a descoberta, a preservação e a manutenção de artefactos semânticos"*

Na perspetiva semântica, os catálogos são considerados elementos cruciais para estabelecer a interoperabilidade, por maximizarem a disponibilidade de modelos de dados numa infraestrutura, bem como por facilitarem o acesso e a compreensão dos diferentes artefactos. Considerando os espaços de dados, no qual o foco é agrupar e disponibilizar diferentes conjuntos de dados, autores como Steinbuss; Almeida, et al. (2025) destacam a utilização do vocabulário padrão Data Catalog Vocabulary (DCAT) *(Data Catalog Vocabulary)*, adotado Modelo de Informação IDS no *Vocabulary Hub*. O DCATv3 (considerando a versão atual é uma norma desenvolvida pelo W3C *(World Wide Web Consortium)* com o objetivo de descrever conjuntos de dados e serviços de catálogos de modo a permitir agregar catálogos distribuídos, melhorando, assim a interoperabilidade entre os metadados. Steinbuss; Almeida, et al. (2025) argumentam que, para integrar o DCAT nos espaços de dados, e exigida uma transição de um modelo genérico para um modelo mais específico e operacionalizável, concretizando esta transição através de determinados passos. Uma das propostas está centrada na criação e utilização de perfis de aplicação, proposta essa considerando que as restrições mínimas de adoção possuídas pelo DCATv3, o que requer a adoção de perfis de aplicação. Estes perfis atuam como "pontes semânticas" que estendem e especializam o vocabulário de acordo com as necessidades do ecossistema. Para concretizar esta estratégia, Steinbuss; Almeida, et al. (2025) apontam como recomendação a definição de campos obrigatórios e recomendados para cumprimentos das normas do ecossistema, a imposição de regras de validação para assegurar que a estrutura dos metadados partilhados cumpre a estrutura exigida, e a utilização de vocabulários controlados como o *Eurovoc* de modo a evitar a ocorrência de ambiguidades.

Enquanto os artefactos semânticos fornecem a formalização acionável por máquina para um determinado domínio de conhecimento, as tecnologias de suporte fornecem a infraestrutura e os mecanismos computacionais necessários para criar, alojar, processar, e aplicar esses artefactos na prática. Neste aspeto, identificam-se os *Knowledge Graphs* e os *Linked Data*. Os *Knowledge Graphs*, tendo em conta a visão de Auer (2022), são estruturas semânticas concebidas como uma malha de conceitos, classes, propriedades, relações e descrições de entidades, com o objetivo fundamental de acumular e transmitir o conhecimento do mundo real, e na qual, cada nó da malha, representa as entidades de interesse e as arestas representam as diferentes relações entre entidades.

As maiores vantagens desta tecnologia, segundo Auer (2022) and Hmimou et al. (2025) residem na capacidade de fornecer uma representação que unifica as instâncias dos dados, os vocabulários, as ontologias e os metadados de forma integrada na mesma estrutura, operando simultaneamente ao nível terminológico e ao nível assertivo. Tendo em evidência a capacidade dos *Knowledge Graphs* para armazenar conhecimento factual rico e explícito, bem como a capacidade de acomodar ligações e mapeamentos de fontes heterogéneas, estes estabelecem a fundação tecnológica ideal para integração semântica em espaços de dados. Permitem, assim, operacionalizar fluxos de dados, aplicar políticas de governo de forma automatizada e garantir que a informação preserva o seu significado ao interagir com diferentes sistemas. (Auer, 2022; Hmimou et al., 2025).

Ainda nas tecnologias de suporte, os *Linked Data* são um método de publicação de dados estruturados que permitem que estes sejam interligados e se tornem mais úteis através da realização de consultas semânticas. Do ponto de vista prático, é uma abordagem que permite a conexão e a integração eficiente de dados heterogéneos provenientes de sistemas distintos, ajudando a criar um amplo espaço global de dados distribuídos, à semelhança da Web na representação de documentos. Ao utilizar tecnologias e normas da Web Semântica, como a RDF, a visão de*Linked Data* ajuda a ultrapassar os desafios de interoperabilidade ao nível sintático, o que permite publicar dados de forma flexível.

De acordo com Auer (2022), a visão *Linked Data* assenta em quatro princípios fundamentais, destacando a utilização de Uniform Resource Identifier (URI)s (Identificadores Uniformes de Recursos), que permitem nomear de forma inequívoca os diferentes elementos dos dados, como entidades, conceitos abstratos ou propriedades, e a inclusão de links para outras entidades e dados relacionados, o que facilita a descoberta de novos dados, encoraja a reutilização de dados e ajuda a estabelecer uma cultura de colaboração, visão estabelecida para os espaços de dados.

Considerando o estado da arte dos artefactos, diferentes autores como Kovach et al. (2026) and Stäbler et al. (2023) identificaram lacunas relacionadas com a gestão semântica, o governo e a integração. Stäbler et al. (2023) identificaram que a distância semântica real entre domínios não é possível de resolver através de alinhamentos puramente estruturais, faltando ferramentas operacionais capazes de harmonizar e descrever automaticamente modelos de dados em *Data Space Mesh*. Uma *Data Space Mesh* consiste numa infraestrutura descentralizada constituída por Espaços de Dados interligados, com o seu desenvolvimento a surgir da necessidade estratégica de evitar a criação de ecossistemas de partilha isolados que iriam representar novos silos de infor-

mação. Esta permite utilizar de forma integrada dados provenientes de fontes que pertencem a diferentes domínios, cruzando por exemplo dados de mobilidade com dados de saúde(Stäbler et al., 2023).

Num mapeamento sistemático sobre espaços de dados mais recente Kovach et al. (2026) apontam essencialmente para a falta de integração holística e de gestão de proveniência subdesenvolvida. No primeiro aspeto, de acordo com Kovach et al. (2026), menos de 35% dos trabalhos identificados integram modelos semânticos de interoperabilidade, soberania e criação de valor simultaneamente em arquiteturas coesas, resultando em soluções isoladas, o problema dos silos de informação. Num outro ponto, a análise revela que a proveniência e rastreabilidade nos espaços de dados, sãoo aspeto da interoperabilidade mais subdesenvolvido de acordo com a literatura, estando presente em 18,7% dos estudos identificados. Há, assim, uma lacuna de mecanismos que utilizam a semântica de forma integrada com as tecnologias operacionais.

3.3.3 O Problema da "Dissonância Ontológica" em Ambientes Federados

A dissonância ontológica pode ser definida como a heterogeneidade semântica e os conflitos (sintáticos e semânticos) que ocorrem quando diferentes domínios, organizações ou sistemas utilizam modelos de dados, vocabulários e ontologias distintos para representar informações (Stäbler et al., 2023). Parte, assim do conflito sobre a natureza do que existe em comparação com a realidade.

Esta dissonância surge porque a compreensão é inconsistente, derivada de diferentes interpretações das partes de uma ontologia ao descrever dados no mundo real, algo que é inevitável na prática. Observando a perspetiva dos espaços de dados, e tendo em conta uma visão na qual vários atores diferentes participam de forma não centralizada, esta dissonância, segundo Stäbler et al. (2023), surge porque a compreensão é inconsistente, derivada de diferentes interpretações das partes de uma ontologia. No que confere às iniciativas de partilha de dados, como a IDSA, GAIA-X ou Catena-X, estas partem, na sua maioria das vezes, focadas em problemas específicos de domínio, o que converge, no momento do desenvolvimento das ontologias, em algo muito específico para um determinado setor específico, com descrição optimizada e estrita.

Se na fase de conceção as ontologias são muito específicas, as mesmas não terão flexibilidade para evoluir, e é neste aspeto que surge a dissonância, nesta tentativa de evolução para uma "*Data Space Mesh*", onde o objetivo é permitir um fluxo de informação entre domínios para gerar semanticamente os artefactos.

O problema maior neste aspeto está na elevada especificidade dos artefactos para um determinado domínio, tendo como consequência a partilha de poucas semelhanças entre diferentes domínios. Assim sendo, a alta especificidade a nível de domínio tende a causar um choque entre representações de diferentes setores, gerando, assim, heterogeneidade semântica e sintática. Este problema é mais impactado quando a utilização é intersetorial, na qual o conflito ocorre ao nível dos modelos semânticos de cada Data Space individual. É de salientar que este problema pode ocorrer também quando uma empresa tenta conectar as suas fontes de dados a um Data Space

existente, sendo, assim forçada a lidar com grandes diferenças de domínio. Compreender a dissonância ontológica é uma condição básica para que a estratégias de espaços de dados seja bem sucedida e escalável, considerando a visão de integração de atores como uma *Data Space Mesh*.

Stäbler et al. (2023), Ganzha et al. (2021), and Hoseini et al. (2025), em diferentes trabalhos mencionam um conjuntos de aspetos que podem influenciar a ocorrência da dissonância ontológica nos ecossistemas de dados. Iniciando pela impossibilidade de impor único modelos, como estamos perante um ecossistema distribuído, não existe a possibilidade de simplesmente forçarmos todos os participantes a alterar a sua semântica interna para um padrão único, o que torna a realidade da heterogeneidade inevitável. Esta realidade, associada aos dados, impacta a ocorrência da dissonância, pois a adoção de diferentes tecnologias de informação leva a uma a existir uma diversidade de formatos e modelos distribuídos, seguindo diferentes estratégias de governo. Um último aspeto, referido por Stäbler et al. (2023) centra-se na especificidade de domínio através da análise de grafos. Embora alguns atributos sejam partilhados entre domínios, a grande maioria dos setores utiliza vocabulários e atributos muito específicos que contrastam fortemente entre si.

Neste aspeto, e tendo por base os pontos de melhoria identificados pelos autores, destacam-se as visões de Hoseini et al. (2025) e Chirvasuta et al. (2025) para colmatar os principais problemas de dissonância que ocorrem ao nível dos espaços de dados. Hoseini et al. (2025) propõem a utilização de Large Language Model (LLM) *(Large Language Models)* para automatizar a anotação de aspetos semânticos e a criação de modelos, compreendendo, assim, o contexto das fontes de dados, mesmo sendo heterogéneas, minimizando o esforço manual exigido no alinhamento dos modelos semânticos. Por outro lado, Chirvasuta et al. (2025) argumentam que, em vez de se procurarem semelhanças diretas entre duas ontologias dissonantes, deve ocorrer a fusão das mesmas, focando principalmente nas suas complementaridades através de uma ontologia superior. O argumento destes autores foi testado num caso específico de harmonização de ontologias sobre baterias, na monitorização da economia circular.

No entanto, existe um conjunto de lacunas necessárias a resolver, sendo um dos objetivos desta investigação fornecer um entendimento mais claro sobre este fenómeno. A primeira lacuna está essencialmente relacionada com a falta de estudos neste aspeto, em particular em relação aos espaços de dados, o que confere a esta dissertação uma oportunidade de explorar um aspeto intimamente relacionado com a semântica e por vezes negligenciado por não ter uma resolução definitiva. Do ponto de vista dos estudos, Stäbler et al. (2023) apontam duas lacunas principais, destacando a falta de processos automatizados e resilientes, pois não existe uma descrição formal dos requisitos de modelos semânticos específicos para construir uma Data Space Mesh, nem um processo automatizado maduro o suficiente para garantir a sua escalabilidade, sendo a maioria dos mapeamentos muito complexa, de forma isolada e manual. A outra lacuna prende-se essencialmente ao alinhamento superficial, pois existe uma necessidade de abordagens de alinhamento de ontologias não apenas ao nível da estrutura global das classes, mas detalhadamente ao nível dos elementos e de instâncias individuais.

3.4 Instanciação e Governação de Espaços de Dados

A secção final da revisão de literatura examina as dimensões de governação e instanciação dos espaços de dados, articulando os modelos teóricos com as lacunas práticas identificadas na literatura, em linha com as boas práticas de gestão de dados do DAMA-DMBOK. Partindo da análise do mesmo no 2.2, são analisados os três modelos de governo de espaços de dados, nomeadamente centralizado, descentralizado e federado, identificando as suas vantagens, limitações e implicações para a gestão da interoperabilidade semântica, com particular atenção ao modelo adotado pela arquitetura International Data Spaces Reference Architecture Model (IDS-RAM) 4.0 (International Data Spaces Association, 2025).

De seguida, é examinado o processo de integração de novos participantes nos espaços de dados, identificando as tensões entre acessibilidade técnica e rigor semântico, demonstrando que a integração técnica de um conector é condição necessária mas não suficiente sem vocabulários partilhados. É ainda aprofundada a gestão do ciclo de vida dos artefactos semânticos no contexto do Vocabulary Hub, incluindo as fases de *Design Time* e *Runtime* e as lacunas de maturidade identificadas nos catálogos existentes (Corcho et al., 2024; Steinbuss; Almeida, et al., 2025).

A secção culmina com a sistematização das lacunas na literatura, distinguindo entre o nível intersetorial no qual a investigação está mais estabelecida, e o nível setorial, onde persiste uma lacuna estrutural que justifica o contributo desta dissertação (Kovach et al., 2026).

3.4.1 Governação de Espaços de dados: Modelos Centralizados, Descentralizados e Federados

O último ponto desta secção de revisão de literatura incidiu principalmente na visão de governo do espaço de dados, infraestrutura discutida nesta dissertação. De modo a clarificar a forma de governar o ecossistema de dados, este ponto incidiu na exploração dos modelos de governo existente identificando desde a sua atuação, ao funcionamento das regras e estratégias adotadas segundo cada visão de governo. Como sustentam Brechtel (2025), reforçados por Kernstock et al. (2025), num modelo de governo centralizado, a tomada de decisão, o controlo do ecossistema e o armazenamento de dados estão concentrados numa única entidade ou proprietário. Este formato de organização centralizada orquestra as interações e impõe regras de funcionamento a todos os utilizadores.

Desta forma, este modelo permite gerir tensões operacionais de forma muito decisiva, impondo eficiência, compatibilidade técnica e controlo de qualidade de forma estrita. Todavia, o principal problema deste modelo, reside na criação de grandes diferenças de poder, levando a decisões que podem servir apenas interesses dos donos dos dados em contrapartida com o objetivo de benefício comum destas infraestruturas. Numa visão diferente, suportada por Kernstock et al. (2025) e com o contributo de Dominik Lis et al. (2023) , surge a reflexão sobre o modelo descentralizado. Assim, no modelo descentralizado, a autoridade, as decisões, e os dados, estão dispersos

e distribuídos por vários atores de rede, sem a dependência de uma entidade central. Nesta abordagem de governação partilhada, as organizações operam em igualdade de circunstâncias, isto é, ao mesmo nível, assentando no consenso entre partes.

A principal vantagem neste modelo reside na promoção de um ambiente de partilha justo, garantindo que cada ator mantém a soberania dos seus dados, mitigando, assim, receios de monopolização do espaço de dados. No entanto, se a autoridade estiver demasiado dispersa, a capacidade de gerar ação coletiva, criar normas rápidas e de chegar a consensos em ecossistemas de grande escala diminui, gerando uma elevada complexidade operacional.

Ainda na ótica de Kernstock et al. (2025), em conformidade com Kovach et al. (2026), surge a visão do modelo federado ou híbrido. Esta proposta aparece como um passo essencial quando o ecossistema de dados amadurece, procurando equilibrar as duas visões anteriores, da exigência de descentralização à necessidade de uma orquestração centralizada.

Daí que, neste cenário, a infraestrutura e a partilha de dados podem permanecer distribuídas nas origens, assegurando a soberania, mas o governo, regras e políticas de utilização e certificação são orquestradas através de uma entidade de governo central. Tendo em consideração a arquitetura IDS-RAM 4.0, o modelo de base, de modo conceptual, localiza-se numa visão descentralizada, com governação prática através de uma visão federada. Com vista a esclarecer estes aspetos, como sustentam Kernstock et al. (2025), a IDSA segue ideais de descentralização e comunicação *peer-to-peer* nos ecossistemas, com os dados a nunca serem armazenados centralmente e o fluxo a ocorrer diretamente entre *connectors*, de quem fornece e de quem consome, sendo crucial para assegurar a soberania de dados. Para clarificar a parte da governação, Nagel et al. (2022), em consonância com Lis et al. (2021) e reforçado por Steinbuss; Almeida, et al. (2025) para garantir o funcionamento da rede descentralizada, a IDSA International Data Spaces Association (2025) assume características centralizadas na forma de gerir o ecossistema. Existe uma entidade central, sem fins lucrativos, uma autoridade dentro do espaço de dados que impõe regras e padroniza o sistema, institucionalizando standards técnicos, certificação de participantes, identidades e vocabulários comuns.

Em síntese, nesta visão de governo existe um ecossistema federado onde a partilha é essencialmente descentralizada, mas as normas que asseguram a confiança na infraestrutura são geridas de modo central.

3.4.2 Processos de Instanciação e Onboarding: Como integrar novos participantes?

Para além da exploração dos modelos de governo, com o objetivo de averiguar as estratégias seguidas pelos principais instanciadores de espaços de dados, o ponto seguinte analisa o processo de integração de atores nessas infraestruturas. Na arquitetura de referência considerada neste projeto, o processo de integração é mapeado como o primeiro dos processos centrais para o funcionamento do espaço, sendo, fundamentalmente, baseado na integração de participantes com verificação de identidade (Kovach et al., 2026). Ao nível técnico, o modelo original da IDSA International Data Spaces Association (2025) é baseado em infraestruturas centralizadas

de gestão de identidades e acessos. No entanto, a evolução atual da arquitetura e das suas implementações na perspetiva descentralizada culminou em alterações no processo. Este "novo" processo de integração de participantes recorre a identificadores descentralizados e credenciais verificáveis, permitindo que as organizações atestem a sua identidade e os direitos de acesso de forma criptograficamente segura, sem a dependência de um ponto central de autoridade (Kovach et al., 2026). Em concordância com Steinbuss; Almeida, et al. (2025), no processo de integração, em simultâneo com os dados, a integração envolve uma fase de *"Design Time"*. Durante esta fase, o novo participante cria autodescrições dos seus recursos e conecta-os a vocabulários semânticos alojados no *Vocabulary Hub*.

Tendo em consideração as visões de diferentes investigadores relativamente a este processo, é de salientar as diferentes perspetivas com base nos diferentes modelos de arquitetura de espaços de dados. Heinz et al. (2022) sublinham as limitações de capacidade das PME (pequenas e médias empresas), defendendo que quem orquestra os espaços de dados deve desenhar processos de partilha altamente acessíveis, recorrendo a provas de conceito para atenuar a hesitação das empresas. Brechtel (2025), com base no exemplo do Catena-X, focam-se na motivação e nos incentivos práticos, argumentando que a complexidade da TI dificulta a entrada, destacando uma abordagem com foco nos componentes de suporte (como SDKS e serviços geridos para aumentar o comportamento de partilha). Já Kernstock et al. (2025) observam o processo de integração sob a ótica das tensões de governo, demonstrando que, à medida que os ecossistemas amadurecem, surgem entidades que operam de forma central para orquestrar a entrada de participantes de forma eficiente. Referem também a utilização da pressão das cadeias de abastecimento, através de termos contratuais, para "forçar" empresas fornecedoras de dados a aderirem. Kovach et al. (2026), através do mapeamento sistemático para compreensão dos espaços de dados, argumentam a necessidade de uma transição necessária de sistemas centralizados de gestão de identidade para tecnologias de identidade de forma auto soberana, para sustentar a confiança numa visão descentralizada para os espaços de dados.

Apesar de o processo de integração apontar muito para as questões de gestão de governo, o aspeto semântico é um processo interdependente da integração. É fundamental salientar que a transferência de dados através de um conector pode ser perfeitamente bem-sucedida do ponto de vista técnico e sintático. No entanto, se os atores da rede não partilharem o mesmo significado para os dados, a colaboração falha na sua finalidade.

Sem esta clareza semântica, os sistemas recetores podem interpretar a informação fora de contexto, gerando *outputs* e conclusões incorretas, mesmo que a transferência técnica dos dados não tenha tido qualquer erro. Por conseguinte, a adoção de modelos semânticos comuns e vocabulários partilhados é precisamente o que permite reduzir a complexidade de integração de tecnologias de informação para novos participantes. Sem um significado partilhado, as empresas seriam obrigadas a despender recursos e tempo significativos para mapear manualmente dados heterogéneos para as estruturas exigidas pelos seus sistemas internos a cada nova colaboração. (Steinbuss; Almeida, et al., 2025)

O uso de uma infraestrutura semântica comum diminui as barreiras de entrada e o esforço de

engenharia, facilitando e automatizando a integração operacional de novos atores no ecossistema de dados (Steinbuss; Almeida, et al., 2025; Brechtel, 2025).

3.4.3 Ciclo de Vida dos Vocabulários: Gestão, Manutenção e Versão em Ecossistemas

Este ponto pretende clarificar em que termos ocorre a gestão do ciclo de vida dos artefactos semânticos, alinhado dentro da arquitetura IDS-RAM 4.0 para destacar a importância do *Vocabulary Hub* neste contexto. A gestão e manutenção dos artefactos semânticos na arquitetura IDS-RAM 4.0 é orquestrada, principalmente, através do *Vocabulary Hub*, que, como mencionado no ponto de definição dos componentes principais da arquitetura, possibilita o alojamento e a gestão de ontologias, modelos de dados de referência e elementos dos metadados (Steinbuss; Almeida, et al., 2025). Para que essa orquestração seja possibilitada, e como mencionado por Steinbuss; Almeida, et al. (2025), o*Vocabulary Hub* tem definido um processo, constituído por duas fases operacionais, sendo concretamente:

Na Fase de Design (Design Time), durante a criação de uma oferta de dados, o fornecedor reutiliza normas existentes ou cria novas descrições semânticas para os seus dados, publicando os vocabulários no *Vocabulary Hub*, conectando-os posteriormente a uma autodescrição do recurso fornecido, neste caso, a oferta de dados (Steinbuss; Almeida, et al., 2025).

Em seguida, na Fase de Execução (Runtime), quando um conector de um consumidor de dados procura um conjunto de dados através do Metadata Broker, este recebe uma referência para o vocabulário aplicável e, se necessário, para o *Vocabulary Hub*. Assim, o consumidor de dados recupera o vocabulário, valida o esquema para assegurar que compreende a estrutura e, só depois, avança para a negociação do contrato e para o consumo dos dados (Steinbuss; Almeida, et al., 2025). Um exemplo de caso de uso fornecido dentro deste capítulo aborda as questões portuárias para exemplificar o funcionamento do *vocabulary hub*.

Para garantir a gestão do ciclo de vida dos artefactos semânticos, cada um dos atores acima definidos terá de desempenhar papéis específicos. Vários investigadores apresentam diferentes sugestões de como gerir os artefactos semânticos, salientando que os modelos semânticos não são estáticos, sendo que a sua manutenção deve basear-se em catálogos devidamente governados ao longo do ciclo de vida.

Iniciando a abordagem por Corcho et al. (2024), ainda que não diretamente com uma abordagem nos ecossistemas de dados, propõe um modelo de avaliação de maturidade e gestão de catálogos de artefactos semânticos com base em 12 dimensões, como Metadados, Abertura, Governação, Qualidade, Sustentabilidade e Transparência. Steinbuss; Almeida, et al. (2025) apontam para um cenário de atualização, argumentando que o ciclo de vida exige a capacidade de executar operações CRUD (Criar, Ler, Atualizar, Eliminar) baseadas em modelos cruzados. Por outras palavras, isto significa que, quando um artefacto semântico muda, o sistema deve preservar a integridade relacional e propagar as atualizações aos elementos de dados dependentes para manter a consistência.

Ainda em Steinbuss; Almeida, et al. (2025), é apontada a utilização do DCAT v3 para gestão de versões, dado que propriedades explícitas como *DCAT:version*, *DCAT:previousVersion* e *DCAT:hasCurrentVersion* o que permite acompanhar de forma exata a evolução dos diferentes conjuntos de dados.

Apesar de uma visão teórica robusta nesta perspetiva, do ponto de vista prático, existe um conjunto de lacunas ainda por resolver. Considerando o ponto de vista dos catálogos de artefactos, Corcho et al. (2024) salientam a falta de transparência e de gestão de versões. No seu estudo, e dentro dos catálogos explorados, identificaram que cerca de 38% dos catálogos documentam de forma clara os fluxos de trabalho de curadoria de metadados e de suporte a versões anteriores dos recursos ainda falha em muitos sistemas, prejudicando a compatibilidade com versões mais antigas de sistema. Ainda neste estudo, foi identificado o uso residual de Identificadores Persistentes (PIDs), no qual existe uma grande lacuna na atribuição de identificadores próprios aos artefactos semânticos (praticada por cerca de 11,5% dos catálogos avaliados no estudo), o que dificulta a rastreabilidade desses modelos ao longo do tempo.

Em linha com a rastreabilidade, o mapeamento sistemático de Kovach et al. (2026), adicionando o aspeto da proveniência, destaca que a capacidade de rastrear a linha temporal de transformação dos dados e vocabulários continua a ser o bloco de interoperabilidade menos desenvolvido na prática, carecendo assim de soluções holísticas integradas. Hmimou et al. (2025), com vista à mitigação destas lacunas, propõem automatizar a governação através de *knowledge graphs*, que retêm as regras e oferecem transparência no registo de alterações. Considerando o ponto de vista semântico, ponto central nesta dissertação, se não existir um bom governo dos artefactos semânticos, não é possível conferir persistência, disponibilidade e manutenção dos mesmos a longo prazo. Torna-se, assim, um grande desafio estabelecer a confiança entre os diferentes participantes dentro do espaço de dados. Além disso, quando a semântica de um modelo evolui de forma descontrolada e não padronizada, as relações semânticas podem perder-se no momento da troca de dados, o que pode levar a que os sistemas recetores interpretem mal o contexto da informação, mesmo que a transferência técnica tenha ocorrido sem erros.

3.4.4 Lacunas na Literatura: A falta de um processo unificado de governação semântica.

Neste ponto, pretende-se concluir e fundamentar as lacunas abordadas ao longo desta revisão, bem como deixar em claro a que esta dissertação pretende responder.

As questões de descentralização impactam a gestão dos artefactos e o alcance da interoperabilidade semântica através dos espaços de dados. Nos estudos identificados, existe uma visão do ponto de vista intersetor, na qual domínios diferentes, como a saúde e a indústria, podem interagir, comunicar e partilhar dados sem ambiguidades. A exploração por parte dos diferentes investigadores tem sido maioritariamente centrada nesta última, o que gera uma lacuna na literatura do ponto de vista setorial. Para este efeito, partimos do estudo de Kovach et al. (2026) através do mapeamento completo dos espaços de dados para uma compreensão aprofundada sobre o estado da arte em que estes se encontram.

3.4.4.1 Lacunas no Intersetor

Daí que, para concluir esta revisão da literatura, surja a necessidade de condensar as diferentes lacunas identificadas. O primeiro ponto a discutir parte da tensão existente ao nível das políticas entre a semântica formal e a expressividade. Hmimou et al. (2025) apontam que os vocabulários como o ODRL são abertos e altamente expressivos para acomodar requisitos legais e de negócio, mas carecem de semântica formal por serem baseados em descrições de alto nível.

Esta flexibilidade tem como consequência o elevado custo da ambiguidade, na qual diferentes sistemas interpretam a mesma política de formas distintas, o que bloqueia a automatização e a interoperabilidade entre organizações. A literatura atual Hmimou et al. (2025), reforçada com Steinbuss; Almeida, et al. (2025), reconhece a urgência de desenvolver semânticas declarativas, bem como a utilização de *Knowledge Graphs* para traduzir o conjunto de regras até então ambíguas em código executável e inequívoco.

Outro aspeto essencial para a lacuna em relação aos limites de padronização, contrastando com a visão dos espaços de dados numa abordagem *Pay-as-you go*, ao nível do ecossistema. Do ponto de vista da interoperabilidade semântica, existe um limite para o nível em que esta pode ser alcançada. Como cada membro de um ecossistema, como um determinado setor de atividade, tem visões diferentes do mundo (do ponto de vista de processos, legal ou de mercado), a imposição de uniformidade no ecossistema é algo irrealista.

Em concordância, os espaços de dados adotam frequentemente uma abordagem de integração *"pay-as-you-go"*, na qual os dados são registados apenas com metadados básicos, mantendo os seus formatos de origem. Neste cenário, apesar da visão dos espaços de dados se basear na criação e promoção de modelos de dados comuns, a prática revela um cenário muito fragmentado, condensando-se numa visão utópica de estabelecer um padrão universal único. Assim sendo, o principal desafio do ponto de vista semântico não é forçar todos os participantes a utilizar a mesma ontologia, mas sim gerir a heterogeneidade contínua. Na literatura atual, o foco está demasiadamente apontado para a troca de dados, negligenciando os modelos de dados e a proveniência dos mesmos (Kovach et al., 2026).

3.4.4.2 Diferenças entre Intersetorial e Setorial

A perspetiva intersectorial parte da capacidade de trocar e reutilizar dados entre diferentes indústrias e domínios de negócio, com o objetivo de evitar o desenvolvimento de silos de informação (Steinbuss; Almeida, et al., 2025). Esta visão confere à economia e à sociedade o benefício de cruzar dados entre diferentes indústrias de forma transversal, sendo possível, através da implementação de *soft infrastructures* comuns e de especificações neutras de metadados padronizadas, permitindo a harmonização semântica e a pesquisa de dados independentemente da origem (Steinbuss; Almeida, et al., 2025; Kovach et al., 2026)

No entanto, para que esta visão seja implementada de modo eficaz, é essencial a resolução de questões de visão mais profunda dentro de cada setor, numa perspetiva setorial. Esta visão é caracterizada pela colaboração e partilha de dados dentro de um mesmo setor ou domínio, com

fronteiras claramente definidas, situando-se a um nível mais baixo de abstração. Neste nível, a arquitetura, como a IDS-RAM 4.0, é optimizada ao nível da cadeia de abastecimento, contemplando a realidade, legislação e processos específicos, exigindo frequentemente a utilização de vocabulários específicos de domínio e ontologias que capturam o conhecimento especializado, bem como as terminologias apropriadas para aquele setor em específico. (Steinbuss; Almeida, et al., 2025)

As diretrizes políticas europeias e as entidades coordenadoras têm como visão final a criação de um ambiente interoperável que permita a reutilização de dados cruzando as fronteiras de diferentes setores da economia. Na literatura científica, esta visão já tem um peso muito considerável: segundo Kovach et al. (2026), as abordagens e estudos intersetoriais representam hoje cerca de 23,5% da investigação sobre espaços de dados. Este volume indica um esforço contínuo no desenvolvimento de tecnologias fundamentais que são essenciais para garantir a futura interoperabilidade entre diferentes domínios.

Porém, apesar de grande parte da visão ser intersetorial, a construção prática da interoperabilidade semântica tem raízes profundas nas necessidades específicas de cada indústria. A literatura demonstra que a adoção é liderada pela verticalidade de alguns setores, com a manufatura a dominar amplamente as implementações (com cerca de 22,8% a 24,7% de representação, seguida de perto pela mobilidade, saúde e energia).

Esta predominância setorial justifica-se pela necessidade de extensões específicas e pelos receios de partilha e de concorrência. Na prática, as implementações setoriais adotam arquiteturas genéricas de espaços de dados, mas necessitam inevitavelmente de criar extensões semânticas, vocabulários e ontologias estritamente desenhados para o seu setor. Devido a isto, a integração e a troca de dados entre diferentes setores continuam muito limitadas na vida real, embora estejam a começar a emergir. Um outro aspeto prende-se com a maioria das iniciativas atuais em espaços de dados ser de colaboração vertical, dentro da mesma cadeia de abastecimento de um setor, porque as organizações ainda têm receios em partilhar dados devido a preocupações com a concorrência. Estando a investigação muito centrada nos aspetos intersetoriais, as necessidades específicas de cada setor acabam por ser negligenciadas.

3.4.4.3 Lacunas no Setor

Um primeiro aspeto específico é, essencialmente, a diversidade interna dentro das cadeias de abastecimento, que acaba por ser negligenciada, já que os modelos de espaços de dados são frequentemente desenhados com uma visão de alto nível, que pressupõe que as empresas são entidades homogéneas que simplesmente trocam dados. Este aspeto, ao longo da cadeia de abastecimento, carece de uma compreensão detalhada dos diferentes motivos e requisitos. Brechtel (2025) dá o exemplo do setor automóvel, na arquitetura Catena-X para mostrar que um setor não é um bloco único, e que dentro do mesmo atuam fornecedores de níveis diferentes, bem como a equipa de reciclagem e a de manufatura. Argumenta ainda que cada um destes subgrupos tem assimetrias

de poder, bem como motivos diferentes para participar no ecossistema, e que precisam de incentivos diferentes para partilhar dados. Neste caso, uma visão intersectorial falha ao não contemplar a complexidade interna.

Num segundo aspeto, a literatura demonstra que o facto da tecnologia estar a ser desenhada numa perspetiva comum tem como consequência a não correspondência de acordo com objetivos de cada setor específico, sem grande capacidade de adaptação. Nagel et al. (2022), apontam que os blocos de construção técnicos dos espaços de dados são desenhados para serem agnósticos ao setor, resultando na limitação de implementações, na sua maioria, através da adaptação de arquiteturas genéricas com extensões específicas para um setor, em vez de desenvolverem uma visão profunda da semântica própria da indústria. No entanto, a capacidade de escalar de alcançar a interoperabilidade, não se conseguem apenas com protocolos de troca de dados, mas sim com vocabulários específicos de domínio que captem o conhecimento especializado e a terminologia de setores particulares, como é salientado porSteinbuss; Almeida, et al. (2025). Sem a criação de perfis de aplicação maduros que adaptem modelos genéricos à realidade do setor, a visão intersectorial cai por terra devido à falta de alicerces semânticos no próprio setor.

Num terceiro aspeto, a literatura refere uma abordagem mais abrangente na interligação entre setores, revelando que persistem certas disparidades entre os setores dominantes e os domínios sub-representados. Os dados da revisão sistemática de Kovach et al. (2026) provam que a suposta maturidade dos espaços de dados é uma ilusão criada pelo sob desenvolvimento de setores como Manufatura e a Mobilidade, refletindo, assim, as suas origens industriais. No entanto, uma lacuna importante nesta visão, segundo Kovach et al. (2026), reside na falta de implementação e visão específica para alguns setores específicos, como as finanças ou os media, que representam ao mesmo tempo desafios e oportunidades de investigação. Tendo em conta estes aspetos, a tentativa de forçar a comunicação entre setores quando ainda falta uma organização robusta ao nível da semântica e da arquitetura setorial acaba por ser prematura. Esta precipitação é sustentada pelas diretrizes da IDSA International Data Spaces Association (2025), que alertam que ecossistemas com elevada variedade de atores intersetorial) exigem padrões semânticos flexíveis, resultando num esforço contínuo de integração manual, enquanto a automação real só é alcançável em ambientes de baixa variedade, como os setoriais, através de padrões estritos (Steinbuss; Almeida, et al., 2025). Consequentemente, a literatura evidencia que as implementações atuais ainda dependem de extensões profundamente específicas de cada domínio, sendo a integração cruzada entre setores ainda muito limitada na prática. Adicionalmente, estudos recentes confirmam que o campo de investigação se encontra ainda numa fase de maturidade onde a novidade técnica é elevada ao nível de arquitetura e conceptual Kovach et al. (2026). Isto justifica que a visão atual dos espaços de dados permanece posicionada numa perspetiva de alto nível, existindo ainda profundas tensões de operacionalização ao transitar destas arquiteturas conceptuais para ecossistemas de dados reais e funcionalmente maduros(Kernstock et al., 2025).

Num último aspeto, é referida a tensão existente entre a política nos espaços de dados e o que realmente existe, tendo em conta a visão para que os espaços de dados não operem em silos e estejam ligados de forma intersetorial desde o primeiro momento (Nagel et al., 2022). Nesta

situação, a visão política contrasta com a realidade, porque do ponto de vista prático, cada ecossistema tem as suas próprias características, moldadas por diferentes processos, leis e mercados (Steinbuss; Almeida, et al., 2025). É possível interpretar, com base nesta grande diversidade de características, que qualquer padrão semântico estabelecido exigirá alguma flexibilidade e esforço de integração manual. Sendo assim, e de acordo com o Steinbuss; Almeida, et al. (2025), é crucial consolidar e aprofundar as estratégias de governo, bem como os modelos semânticos dentro dos próprios setores de atividade, como passo anterior à implementação da ambição europeia, no seu objetivo de cruzar dados entre diferentes setores. As atuais arquiteturas focadas em interoperabilidade, como é o caso da IDS-RAM, fornecem a *soft infrastructure* segundo Nagel et al. (2022), mas não o conteúdo semântico. Sendo assim, é possível afirmar que o foco até então está centrado na comunicação intersetorial e nos protocolos de comunicação, descurando a necessidade de definir claramente a especificidade do conteúdo que circula nas redes de comunicação, como a especificidade, a governação e as diferentes taxonomias de cada setor individual.

Em continuação, reforçando este ponto, justifica-se a adoção da arquitetura IDS-RAM 4.0 como base devido ao seu nível de maturidade do ponto de vista teórico. De acordo com Kovach et al. (2026), a arquitetura IDS é a mais madura e dominante no desenvolvimento de espaços de dados europeus, dominando cerca de 75% das implementações de acordo com a literatura científica. É a arquitetura que define de forma mais clara as fundações para a interoperabilidade semântica, na qual estão integrados os diferentes *IDS Connectors* e os *Vocabulary Hubs*, pois mapeia ativamente o vocabulário DCAT v3 e a linguagem ODRL (Steinbuss; Almeida, et al., 2025). Apesar de muito aplicada do ponto de vista teórico e estando na base de aplicações práticas, como a Catena-X, a principal lacuna na IDS-RAM 4.0 reside na falta de validação empírica de estudos e prática de longo prazo que avaliem verdadeiramente o impacto e a eficácia da implementação desta arquitetura no terreno (Kovach et al., 2026). Esta lacuna é suportada por Kovach et al. (2026) and Heinz et al. (2022), identificando como aspetos centrais, a falta de avaliação concreta, com 64,4% da literatura a concentrar-se apenas na proposição de soluções e validações iniciais sem avaliação concreta e o foco excessivo na arquitetura conceptual, ficando as implementações limitadas ao nível da própria arquitetura e de provas de conceito sem avaliações a longo prazo.

As lacunas identificadas permitem concluir que, para ultrapassar os desafios da interoperabilidade semântica em espaços de dados, que se assentam numa visão federada, é essencial a conceção de soluções orientadas para o interior de cada setor e estabelecer um processo que assegure uma gestão eficaz da informação ao longo de todo o ecossistema. Em particular, permanece por consolidar uma perspetiva setorial que contrarie a ideia, frequentemente assumida, de que as organizações do mesmo setor "falam a mesma língua" e utilizam, de forma homogénea, os mesmos sistemas e modelos de dados. De forma análoga, torna-se pouco consistente ambicionar a comunicação semântica entre setores distintos quando, no seio do próprio setor, subsistem ainda problemas relevantes de interpretação, sentido e significado.

Neste contexto, e como síntese da revisão de literatura realizada, a solução proposta tem como objetivo definir um processo que permita às empresas de um mesmo setor comunicarem entre si, no âmbito da infraestrutura dos espaços de dados, minimizando ambiguidades e evitando perda

de significado. O processo a ser desenvolvido ataca diretamente as raízes de várias lacunas identificadas ao longo desta revisão de literatura, desde a ambiguidade semântica até a complexidade dos processos de governo dentro de cada setor. Com base em Steinbuss; Almeida, et al. (2025) and Kernstock et al. (2025), o desenvolvimento deste processo permitirá capturar conhecimento de modo especializado, bem como permitir padrões mais estritos e com menos variedade.

Esta abordagem de forma uniformizada, desenvolvida para dentro de um único setor permite impor modelos semânticos mais rigorosos, o que reduz a ambiguidade e remove a necessidade de adaptar manualmente os padrões semânticos, permitindo automatizar a validação e a troca de dados. (Steinbuss; Almeida, et al., 2025). Indo mais além, considerando uma solução ao nível de governo do setor, o desenvolvimento deste processo permite selecionar ou adaptar vocabulários específicos de domínio de acordo com as necessidades de cada setor. Este processo revela-se essencial por garantir que a seleção de vocabulário reflete exatamente os regulamentos e requisitos operacionais que cada setor tem em específico, sendo flexível ao ponto de poder ser adaptado (Kernstock et al., 2025; Steinbuss; Almeida, et al., 2025).

Chapter 4

Metodologia

4.1 Introdução

A presente investigação adota uma abordagem DSR (Design Science Research) para que, através da revisão de literatura sobre a instanciação de espaços de dados, seja possível concretizar o objetivo de desenvolver um processo de suporte à interoperabilidade semântica em espaços de dados setoriais.

Esta abordagem metodológica, vista como um paradigma na área científica dos sistemas de informação (Gregor et al., 2013), foi selecionada com o intuito de que, desta investigação seja produzido um artefacto com foco em resolver um dos problemas centrais na instanciação de espaços de dados ao nível setorial, bem como contribuir com conhecimento científico valioso na área da interoperabilidade semântica. Tendo em conta a adoção do DSR, esta investigação seguirá a framework de Peffers et al. (2007), que estabelece as fases necessárias com objetivo de garantir rigor e relevância ao artefacto a ser produzido. De forma enquadrada com a framework DSR, são estas as fases que irão guiar o desenvolvimento do processo:

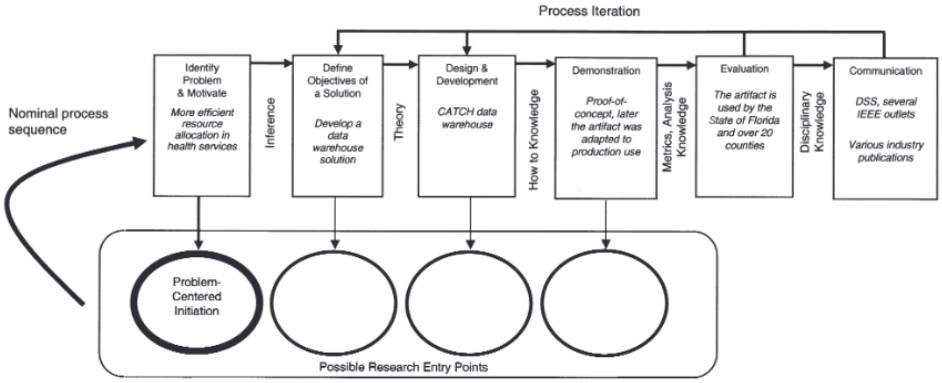


Figure 4.1: Design Science Research, Peffers et al. (2007)

1. Revisão de literatura e identificação de lacunas ao nível da interoperabilidade semântica na instanciação de espaços de dados setoriais, ontologias e vocabulários de domínio.

2. Desenvolvimento das questões de investigação e definição do objetivo de desenvolver um processo de suporte à interoperabilidade semântica, para quem gere o ecossistema de dados setorial.

3. Desenho do processo de suporte, com foco em elementos de interoperabilidade semântica, como ontologias e vocabulários de domínio e revisão de arquiteturas de referência como a IDS-RAM 4.0 (International Data Spaces Association, 2025), e os componentes de software para troca de dados.

4. Demonstração do processo, com construção de diagramas, construção de cenários, demonstração passo a passo de que forma o processo orienta o gestor do ecossistema setorial na seleção de ontologias, vocabulários e outros elementos de interoperabilidade semântica.

5. Avaliação do processo desenvolvido ex ante e artificial, para testar a completude lógica, a robustez procedimental e a aplicabilidade do artefacto antes da sua instanciação em ambientes de produção (Venable et al., 2012)

6. Defesa da Dissertação com apresentação do processo de suporte e resposta clara às questões de investigação.

A metodologia encontra-se organizada em seis fases, condensadas em três fases principais: (1) Planeamento e diagnóstico do problema, (2) Concepção e construção do artefacto, e (3) Avaliação e Reflexão. Estas fases não são obrigatoriamente sequenciais, sendo possível regressar a fases anteriores em função dos resultados obtidos.

4.2 Desenvolvimento da Investigação-Design Science Research

4.2.1 Ciclos de Investigação

Para o desenvolvimento desta investigação, a abordagem DSR envolve três ciclos, conforme identificado por Hevner (2007) que garantem que o processo a ser desenvolvido é aplicável na prática e é teoricamente sólido.

4.2.1.1 Ciclo de Relevância

Este ciclo localiza a investigação nas necessidades do mundo real, sendo identificados requisitos para a interoperabilidade semântica ao nível específico do setor ou domínio, com o objetivo de garantir que o artefacto responde aos problemas concretos de silos de informação, falta de semântica partilhada entre participantes de espaços de dados, gestão eficaz da informação em ecossistemas de dados.

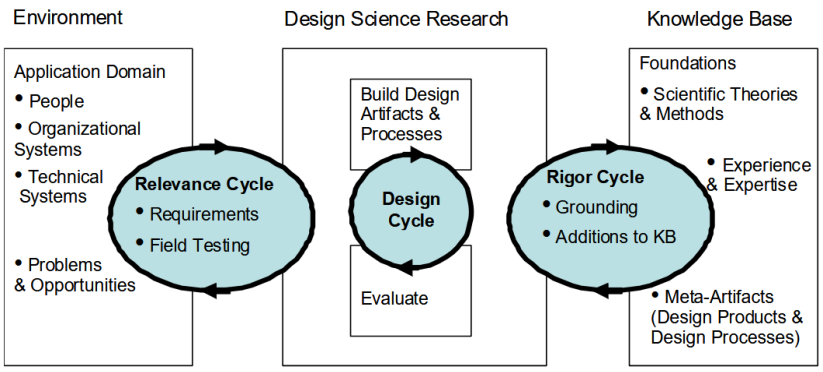


Figure 4.2: Design Science Research Cycles, Hevner (2007)

4.2.1.2 Ciclo de Rigor

O ciclo de rigor fornece as bases científicas, bem como a fundamentação teórica desta dissertação. Aqui, a investigação é fundamentada na exploração do estado da arte sobre as arquiteturas de Espaços de Dados e interoperabilidade semântica para assegurar que a solução proposta é inovadora, por ter uma visão setorial, específica para entender a comunicação entre diferentes participantes de um mesmo setor, estando a um nível abaixo da visão intersetorial.

4.2.1.3 Ciclo de Design

No centro da DSR, o ciclo de design é o processo interativo de construção e avaliação do processo de suporte à interoperabilidade semântica, com melhorias sucessivas para encontrar a melhor solução. O processo será testado quanto à capacidade de representar a realidade de um domínio específico (como a manufatura), com a possibilidade de ser adaptado a diferentes setores, permitindo estabelecer a interoperabilidade semântica. Através da seleção de artefactos semânticos (ontologias, vocabulários) adequados a uma determinada troca de dados entre participantes de um mesmo setor, o processo permitirá a uma determinada entidade gerir o espaço de dados neste aspeto, avaliando a maturidade dos artefactos, permitindo melhorar o sentido e o significado dos mesmos com base nas interações.

4.2.1.4 Problema e Motivação

O estabelecimento da interoperabilidade semântica em espaços de dados é possível através de estruturas de suporte ao governo como a DSSC e de arquiteturas de referência como a IDS-RAM 4.0, que permite instanciar os espaços. Esta arquitetura tem flexibilidade até certo ponto na instanciação dos espaços de dados, se a visão do ecossistema for de alto nível. No entanto, quando a visão passa a estar num domínio específico, de que forma é garantida a partilha de dados, com entendimento por todos os atores do espaço de dados, sem ambiguidades ou perdas de sentido ou

significado? Não existindo ferramentas, uma estrutura ou mesmo um processo que possibilite instanciar espaços de dados ao nível dos setores, os objetivos europeus de estabelecer uma economia focada na partilha e troca de dados acabam por ficar comprometidos.

Um dos maiores desafios à instanciação de espaços de dados dentro de setores, centra-se precisamente na interoperabilidade semântica, com o objetivo de "garantir uma compreensão adequada dos dados e possibilitar a sua utilização de forma eficaz em diferentes contextos e aplicações". (BAIDATA, [n.d.]; Steinbuss; Almeida, et al., 2025) A visão de iniciativas como a IDSA ou a GAIA-X partem da visão de alto nível de colocar diferentes setores a comunicar entre si. No entanto, esta visão de alto nível acaba por ser redutora, dando o entendimento que os atores dentro do mesmo setor são homogéneos na comunicação e nos sistemas utilizados para comunicar.

4.2.1.5 Objetivos para a Solução

A construção de um processo de suporte à interoperabilidade semântica para espaços de dados setoriais, compreende como objetivos desde o estado da arte sobre implementações, levantamento de requisitos funcionais e não funcionais, até ao desenho do processo em si através de modelação conceptual e mapeamento do setor. Os principais objetivos a desenvolver nesta dissertação, compreendem como pontos desde logo, analisar como a informação circula em ecossistemas de dados, estabelecidos através de espaços de dados ao nível setorial, identificando os atores, as tipologias de dados e os mecanismos de governo. Considerando a arquitetura IDS-RAM 4.0 como referência, é objetivo o estudo do papel de componentes como *Vocabulary Hubs, Connectors e Metadata Brokers,* na forma como operam no espaço de dados e no seu contributo para a interoperabilidade semântica.

4.2.1.6 Desenho e desenvolvimento

Após a revisão de literatura nos aspetos mencionados, a identificação de requisitos funcionais e não funcionais nos setores representa o passo seguinte para o planeamento do processo de suporte a interoperabilidade semântica e o governo da informação. Neste passo, o conjunto de tarefas essenciais compreende a identificação de fluxos de informação, procedimentos de governo e artefactos semânticos utilizados nas diferentes trocas de dados.

Realizado o levantamento de requisitos, alinhados com a Blueprint 3.0 DSSC, o DAMA-DMBOK e a arquitetura IDS-RAM 4.0, será então desenhada uma proposta de processo para a gestão de espaços de dados setoriais, focado em interoperabilidade semântica, através da gestão de artefactos semânticos como vocabulários e ontologias.

4.2.1.7 Demonstração e Avaliação

Estando a proposta desenhada, é necessário validar a proposta de processo. Para esse efeito será desenhado um caso hipotético-dedutivo de forma descritiva, para demonstrar a viabilidade do processo na prática, comparando a evolução na gestão dos artefactos semânticos antes e após

a existência do processo, bem como observar possíveis melhorias, de forma a aumentar a sua maturidade para implementação no mercado.

4.2.1.8 Comunicação

Por fim, o último passo desta dissertação, de acordo com a DSR, tem o objetivo de comunicar o artefacto produzido, contribuindo academicamente para o estado da arte sobre os principais desafios de interoperabilidade semântica no desenho e instanciação de espaços de dados setoriais. A presente dissertação de mestrado terá uma defesa pública, seguindo-se a sua publicação no repositório da Universidade do Porto.

4.3 Concretização da Investigação

4.3.1 Questões de Investigação

A investigação desenvolvida no âmbito da interoperabilidade semântica em espaços de dados setoriais, teve como objetivo responder às seguintes questões de investigação:

Questão Principal *De que modo podemos auxiliar as entidades de governo em espaços de dados setoriais para garantir a interoperabilidade semântica?*

Questões complementares

- Que elementos-chave de suporte à interoperabilidade semântica deverão estar contemplados neste processo?
- De que modo os espaços de dados resolvem a dissonância ontológica e a interoperabilidade semântica entre sistemas heterogéneos ao nível setorial?

Estas questões permitiram a construção da base de conhecimento necessária para o ciclo de rigor da DSR, a qual foi concretizada através de uma abordagem exploratória, combinada com um conjunto de passos selecionados do protocolo PRISMA, numa revisão de literatura. O alinhamento com este protocolo revela-se essencial para garantir transparência e rigor científico na seleção dos documentos a sustentar o artefacto.

A primeira fase consistiu na identificação de fontes para estabelecer uma base de conhecimento abrangente sobre interoperabilidade semântica, espaços de dados e modelos de governo.

4.3.1.1 Expressão de Pesquisa

A estratégia de busca foi consolidada através de um refinamento da *query* de pesquisa, recorrendo a tesauro para expandir o vocabulário e garantir a recuperação de um conjunto documental robusto. O finalidade da *query* abrange de forma integrada os conceitos de ecossistemas e espaços de dados, as suas dinâmicas de governação, intersectados com o pilar central da interoperabilidade semântica. Tecnicamente, a sintaxe foi adaptada às especificidades de cada base de dados, utilizando operadores booleanos (AND, OR), aspas para termos compostos e parênteses para a estruturação

lógica da heurística. Este rigor na formulação permitiu uma recolha exaustiva e precisa, essencial para mapear o estado da arte e os mecanismos de governo em ambientes descentralizados.

((“Semantic Interoperability” OR “Data Interoperability” AND
“Data Space*” OR “Data-Space*” OR “Data Ecosystem*” AND
OR “Data Governance” OR “Governance” OR “Ecosystem Governance”))

Figure 4.3: Expressão de Pesquisa

4.3.1.2 Bases de Dados

A temática dos espaços de dados ainda carece de um número relevante estudos publicados em artigo de revista, sendo um tema amplamente discutido em conferências e workshops. Assim sendo, foi selecionada a SpringerNature por conter tanto livros como capítulos, concentrando a maioria da informação disponível sobre espaços de dados. Outras duas bases como a Scopus e a Web of Science foram selecionadas com base no conhecimento prévio do seu funcionamento, bem como no rigor que conferem na investigação, por disporem na sua maioria de artigos de revista revistos por pares. As restantes fontes de informação partem desde literatura cinzenta, como é o caso da documentação IDSA (International Data Spaces Association, 2025), Position Papers (Steinbuss; Turkmayali, 2024; Steinbuss; Almeida, et al., 2025), de documentação decorrente de projetos e workshops em desenvolvimento e de fontes anteriormente selecionadas através de pesquisa exploratória.

4.3.2 Análise da Literatura: Fases de triagem

4.3.2.1 Restrições-Bases de Dados

Após a execução da expressão de pesquisa nas diferentes bases de dados, foram recolhidos 519 documentos, com 32 duplicados. No entanto, com o objetivo de explorar o que existe de mais recente dentro do universo dos espaços de dados, surgiu a necessidade de aplicar filtros às bases de dados, de modo a convergir para o objetivo de explorar a semântica e o governo em espaços de dados. Assim, na aplicação de filtros foram selecionados estudos compreendidos entre 2015 e 2026, bem como seleção por áreas como *Computer Science, Engineering, Business and Management, Economics, Science, Humanities and Social Sciences, Multidisplinary*. Isto permitiu reduzir o conjunto de documentos para 370.

4.3.2.2 Critérios de Seleção

Para garantir que a seleção de fontes fosse objetiva e alinhada com os objetivos da investigação, com foco em interoperabilidade semântica e governo de espaços de dados, foram definidos critérios de elegibilidade. Estes critérios permitem filtrar o conhecimento essencial, relevante para o estabelecimento do ciclo de rigor, sendo definidos com pertinência temática, acessibilidade e rigor metodológico das fontes.

4.3.2.3 Critérios de Inclusão:

IC1: Relevância e Contexto: Estudos que abordam diretamente os conceitos fundamentais da investigação e fornecem base conceptual ou histórica necessária.

IC2: Contribuição em Processos e Modelos: Documentos que apresentam artefactos semânticos específicos, modelos de governo, arquiteturas técnicas ou aplicações práticas no contexto organizacional.

IC3: Foco Setorial: Estudos com foco em colaboração inter-organizacional e redes de partilha de dados.

4.3.2.4 Critérios de Exclusão:

EC1: Desalinhamento Temático: Artigos com foco exclusivo em hardware e componente técnica, sem considerar aspetos semânticos dos dados.

EC2: Limitações de Qualidade e Formato: Inclui documentos que não apresentem metodologia, pouco rigor, texto em idiomas não considerados (ex: língua alemã), bem como literatura não validada como artigos de opinião ou resumos de conferência sem revisão por pares.

EC3: Redundância: Registos com nomes diferentes e que apresentem os mesmos resultados de estudos já incluídos.

EC4: Acesso ao documento: Fontes onde não foi possível obter o texto integral, documentos sem link ativo para consulta.

4.3.2.5 Secções de análise

- **Titulo e Abstract:** Nesta fase de triagem foram analisados 370 artigos por título e abstract, aplicando o conjunto de critérios para convergir numa amostra sólida. A partir deste procedimento resultou a exclusão de 182 documentos. A maioria dos documentos excluídos estava concentrada nos primeiros três critérios definidos.

- **Introdução e Conclusão** Nesta segunda fase de análise, foram analisados 188 documentos incluídos na fase anterior. A análise nesta fase foi centrada na introdução e conclusão dos artigos, com o objetivo de perceber se o título e abstract efetivamente correspondiam à proposta do artigo, replicando novamente os critérios. Desta análise resultou a exclusão de 96 documentos. A análise avançou para a fase seguinte, com a leitura dos 92 documentos selecionados para análise na íntegra, com utilização da ferramenta de Inteligência Artificial Google Notebook LM para aprofundar a extração de dados desses documentos. Neste conjunto encontram-se os 92 documentos recuperados pelas bases de dados através do protocolo PRISMA, bem como 35 documentos recolhidos anteriormente no âmbito da análise e no decorrer do projeto.

- **Leitura Completa e Inteligência Artificial:** Esta última fase de análise, a fase mais importante da definição da base teórica a suportar a presente dissertação, os documentos foram analisados em primeira instância com recurso ao Notebook LM. A utilização deste recurso

permitiu selecionar, validar e identificar com precisão o conjunto de documentos que iria necessitar de uma leitura mais atenta e aprofundada para extração de conceitos e entendimento amplo do estado da arte dos espaços de dados na perspetiva semântica e de gestão de governo. Inicialmente, todos os documentos selecionados nesta fase, foram carregados num único *notebook*, e utilizado o seguinte *prompt* adaptado para extração de dados:

"Analyse the top 20 papers related to the topic of Semantic Interoperability in data ecosystems and data spaces contexts, considering the focus in setorial ecosystems and a possible process development for ecosystem governance and produce a structured synthesis. For each article, present the information in the following format:

- *Full reference (Author, Year)*
- *Analytical summary (2–4 sentences explaining the aim, methods, and key findings)*
- *Relevant direct quote #1 (include quotation marks and page number if available)*
- *Relevant direct quote #2 (optional, if appropriate)*
- *Critical comment or link to other work (optional but desirable; e.g. how it relates to other papers)*

At the end, create a comparative summary table with the following columns: | Author/Year | Aim | Methodology | Main Contribution | Limitations |

Important:

- *Use academic language.*
- *Include quotes from the article that illustrate core claims or definitions.*
- *Always mention methodology and context (e.g. lab study, field study, case study).*
- *Keep the summaries concise but informative."*

A utilização deste *prompt* foi repetida até todos os documentos serem analisados pelo *Notebook LM* e seguidamente escrutinados para assegurar o rigor científico da investigação. Estes processos permitiram definir o conjunto de documentos que requeriam uma análise mais aprofundada para situar o estado da arte nos espaços de dados, nas perspetivas de governo, de interoperabilidade semântica e de validação empírica da instanciação dos espaços de dados.

4.4 Construção do Artefacto

A construção de um processo unificado de suporte à interoperabilidade semântica em espaços de dados, a ser desenvolvido ao longo desta dissertação, é centrado no Ciclo de Design da DSR. Como mencionado anteriormente, este ciclo é focado na construção e avaliação de artefactos semânticos através de um processo iterativo de refinamento. Na presente investigação, o foco está centrado

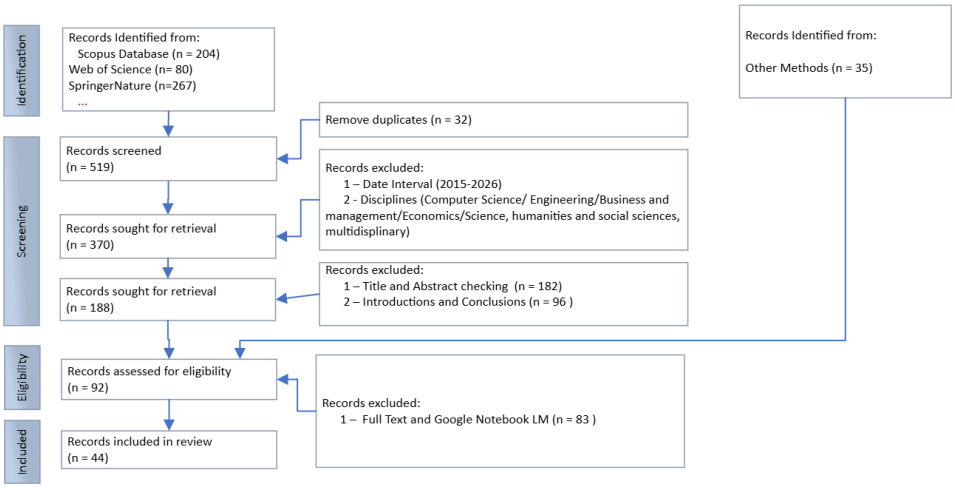


Figure 4.4: Diagrama PRISMA para Revisão de Literatura

numa visão setorial, garantindo que os diferentes artefactos semânticos como vocabulários e ontologias partilhados sejam adaptados às especificidades técnicas de um setor em concreto, como o caso da manufatura ou saúde. A construção do processo nesta dissertação fundamenta-se nos seguintes pontos:

- **Representação Conceptual:** Com base na identificação de conceitos e relações extraídos da literatura e do DSSC Blueprint 3.0, foram desenvolvidos os modelos que representam a governação no domínio, decompostos em sete subprocessos formais em UML 2.5.1 que satisfazem os requisitos de interoperabilidade semântica ao nível setorial.

- **Demonstração e Validação Descritiva (*Walkthrough*):** Em conformidade com o enquadramento metodológico de DSR (Peffers et al., 2007; Hevner, 2007) e a *framework* FEDS (Venable et al., 2012) para avaliação ex-ante e artificial, a utilidade, completude lógica e operacionalidade do artefacto são demonstradas através de um cenário ilustrativo dedutivo walkthrough centrado no Passaporte Digital do Produto (DPP) para a indústria têxtil, articulando o eBIZ, TRICK e a especificação setorial CEN CWA 18291:2025 (*CWA 18291*, [n.d.]).

- **Avaliação de Qualidade e Governação:** A robustez do processo é formalizada através da incorporação de fichas operacionais de atividade ancoradas nas normas ISO/IEC 11179 e standards W3C, complementadas pelo instrumento métrico do *Data Model Scorecard* para assegurar auditorias objetivas antes da publicação no *Vocabulary Hub*.

- **Delimitação de Trabalho Futuro:** Reconhecendo a natureza *ex-ante* do teste conceptual em fase de desenho (*Design-Time*), a validação empírica através de entrevistas semi-estruturadas ou painéis de especialistas em espaços de dados fica formalmente delimitada como etapa prioritária de investigação futura.

Chapter 5

Construção do Artefacto

5.1 Posicionamento do Artefacto no Ciclo de Vida de Espaços de Dados

A conceção do processo proposto parte do enquadramento metodológico preconizado pelo DSSC para o ciclo de vida de implementação de ecossistemas federados. A presente secção delimita o posicionamento exato do artefacto na etapa de desenho (*Design-Time*), identificando a sua incidência direta sobre o bloco técnico de *Data Models* do Blueprint 3.0 e estabelecendo as fronteiras entre o governo do significado dos dados e a implementação da infraestrutura técnica.

5.1.1 Enquadramento no Blueprint 3.0 e no Modelo de Co-Criação

A conceção do artefacto proposto nesta dissertação partiu de um mapeamento e revisão da intervenção no ciclo de desenho e definição de espaços de dados, tal como estruturado pelo Data Space Support Centre (DSSC) no *Blueprint* 3.0(Data Spaces Support Centre, 2026). A documentação estabelece que a criação de um espaço de dados obedece a um ciclo composto por três fases essenciais - exploratória, preparatória e de implementação. A análise desta estrutura permitiu determinar em que fase se deveria localizar o artefacto desenvolvido, com destaque para a ligação entre a fase preparatória e a fase de implementação.

De acordo com o modelo de co-criação do *Blueprint*, o artefacto desenvolvido situa-se na **fase de desenho (*Design*)**, na qual os requisitos do espaço de dados já se encontram estabelecidos, os casos de uso definidos, bem como os papéis e responsabilidades dos atores envolvidos. Não se trata, portanto, da fase de instanciação - nesta, o espaço de dados ainda não existe enquanto infraestrutura operacional - mas sim da fase em que é necessário definir, com detalhe, os elementos que hão de reger o funcionamento do espaço antes da sua materialização, como as regras que disciplinam a comunicação sobre a qual o espaço de dados vai operar.

Especificamente, o *Blueprint* define, através da sua *framework*, um conjunto de *building blocks* necessários à materialização de um espaço de dados. O artefacto aqui desenvolvido incide diretamente sobre o *building block* "Data Models", para o qual o *Blueprint* estabelece a necessidade

de um bloco técnico, mas não define o fluxo operacional, o passo a passo, segundo o qual a entidade de governo deve avaliar, monitorizar e, em última instância, posicionar-se e definir o modelo de dados a adotar. É precisamente neste ponto identificado no método de co-criação como o pontoB.4.Step-3, dentro do *building block* "Data Models" que se posiciona o artefacto desenvolvido: um processo estruturado que define o percurso completo de um modelo de dados, desde a sua seleção até à publicação, governo e manutenção dentro do espaço de dados (Data Spaces Support Centre, 2026).

5.1.2 Âmbito do Processo: Início, Fim e Tarefas Essenciais

Como referido na revisão de literatura (3.3.1), importa reforçar a distinção entre interoperabilidade técnica e interoperabilidade semântica, na medida em que o artefacto desenvolvido não tem como objetivo a construção de conectores técnicos de troca de dados, mas assentando no governo do significado dos dados, através de vocabulários, ontologias, *profile applications* e *data schemas*, enquanto elementos centrais do modelo de dados, e em ligação direta com a Ciência da Informação, na sua vertente de organização e representação do conhecimento. Dentro do método de co-criação e no passo específico em que se posiciona, este processo tem, assim, o seu foco claramente estabelecido na interoperabilidade semântica.

Para estabelecer claramente o início e o fim do processo desenvolvido, foram identificados os pontos de partida e de chegada de acordo com os *standards* existentes e as boas práticas do DAMA-DMBOK (Henderson et al., 2017). O processo tem início na seleção de uma ontologia, vocabulário ou outro elemento semântico complementar ao modelo de dados, a partir dos requisitos pré-existentes do setor selecionado. Segue-se um percurso de avaliação e verificação de aspetos como adequação, extensão e qualidade do elemento selecionado, culminando no registo e implementação desse elemento no espaço de dados, onde passa a ser governado nas suas tarefas subsequentes de atualização, versões e gestão do processo.

De acordo com as boas práticas do DSSC para os modelos de dados, a entidade de governo do espaço de dados deverá privilegiar sempre que possível a reutilização de modelos base pré-existentes, consolidados e disponíveis no mercado em contrapartida à criação de um modelo de raiz.

5.2 Fundamentação da Conceção do Processo

A estruturação procedimental do processo apoia-se nas melhores práticas internacionais de administração e curadoria de ativos informacionais. Esta secção fundamenta as bases normativas que colmatam as omissões operacionais do DSSC através da mobilização do referencial DAMA-DMBOK, derivando as atividades a partir das questões-chave de co-criação e formalizando a segregação estrita de funções entre a entidade proponente e a autoridade auditora de governo.

5.2.1 Base Normativa e Lacuna Identificada

A conceção do artefacto tem como base a literatura de referência e a documentação do DSSC, partindo do método de co-criação descrito no *Blueprint*, com foco específico no *building block* "Data Models". Do ponto de vista da *framework*, este bloco de construção técnico é definido como necessário para assegurar a interoperabilidade no espaço de dados, estabelecendo os requisitos, os casos de uso e a estrutura geral do que deve ser implementado, deixando, no entanto, um conjunto de questões orientadoras de co-criação em aberto, a resolver por cada entidade de governo na fase de desenho.

É precisamente aqui que se situa a lacuna que motiva o desenvolvimento deste artefacto: o *Blueprint* responde à questão do *"o quê"*, que elementos deve conter um *Data Model*, que papéis existem, mas não explicita o *"como"*, de que forma, com que meios, através de que fluxos e, especificamente, com que repartição de responsabilidades entre atores se concretiza esta fase de desenho do espaço de dados.

Para colmatar esta lacuna, a conceção do processo apoiou-se no guia de referência do DAMA-DMBOK (Henderson et al., 2017), enquanto corpo de conhecimento consolidado em gestão de dados, para estruturar não apenas as atividades do processo mas também os pontos de decisão e os registos que devem ficar formalmente documentados ao longo do ciclo de vida do modelo de dados. A análise ao DAMA-DMBOK permitiu identificar quais as áreas essenciais para o desenvolvimento de um processo de governo semântico, bem como selecionar os referenciais para o suportar o processo, desde a definição de papéis, artefactos a produzir e fluxos para uma gestão de dados eficaz.

A leitura do DAMA-DMBOK teve a sua base em quatro capítulos essenciais que permitiram estabelecer as áreas de atuação do processo em linha com a documentação de referência DSSC.

A primeira análise partiu no Capitulo 3-Data Governance, essencial para a definição de papéis e responsabilidades inerentes à gestão do espaço de dados, aqui diretamente nos modelos de dados. O capítulo 5-Data Modeling and Design para um entendimento sobre a forma é que os dados são modelados, para entender o conceito mais amplo do que é um modelo de dados, como é construído, que elementos estão presentes e sobretudo a ferramenta do Data Model Scorecard para avaliar a qualidade dos modelos de dados, a referir mais adiante no esclarecimento sobre o processo. O capítulo 9-Document and Content Management foi essencial para o entendimento claro e definição dos elementos que compõem o modelo de dados, como os vocabulários, ontologias e esquemas, o conceito de modelação semântica, bem como os diferentes formatos. Ainda neste capítulo estão presentes recomendações aplicáveis à gestão de documentos e registos, como o controlo de versões e gestão de metadados.

A compreensão de gestão de metadados resultou também na análise do capítulo 12, especificamente dedicado a este tema, no qual estão estabelecidas recomendações para a gestão de metadados e identificado o conjunto de normas ISO/IEC de suporte, como o caso da ISO/IEC 11179 (*ISO/IEC 11179-1*, [n.d.]), norma de referencia para o registo de metadados.

A descoberta de modelos de dados em catálogo depende da forma como os metadados são registados. A definição clara de regras no momento da descrição dos produtos de dados torna-

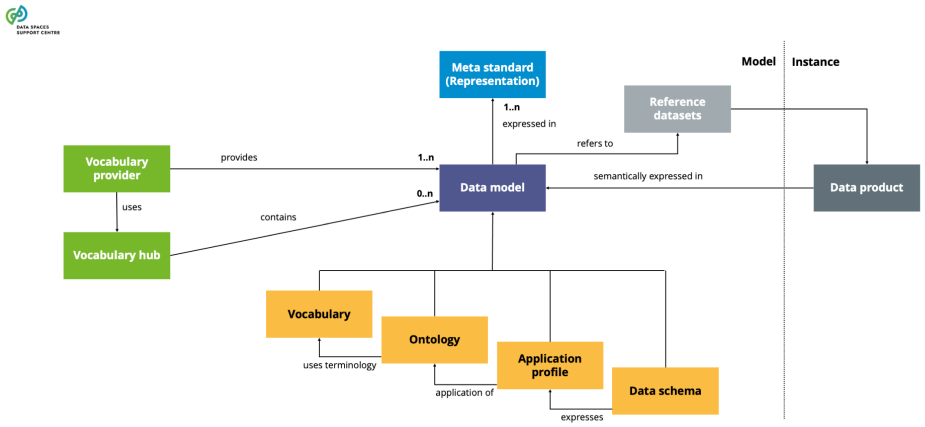


Figure 5.1: Data Models Technical Building Block (Data Spaces Support Centre, 2026)

se essencial no objetivo de alcançar a interoperabilidade semântica ao facilitar a descoberta de modelos.

5.2.2 Perguntas de Co-Criação como Estrutura de Derivação do Processo

A resposta às perguntas de co-criação documentadas no método do *Blueprint* constituiu o ponto de partida para a definição das atividades e dos diferentes subprocessos que compõem o artefacto. A reflexão foi estruturada em torno de questões nucleares, cada uma delas correspondente a uma fase distinta do ciclo de vida do modelo de dados enquanto documento vivo dentro do espaço de dados setorial.

Do conceito abstrato de domínio ao esquema. A primeira questão de co-criação prende-se com a forma como um conceito abstrato de domínio, como próprio do vocabulário de negócio de um setor é traduzido num esquema formal e implementável. Esta transição é ilustrada na *framework* do *Blueprint* (Figura5.1) e corresponde, no artefacto desenvolvido, à fase inicial de seleção e triagem de modelos candidatos: em vez de partir de uma modelação de raiz do domínio, o processo privilegia a identificação de modelos semânticos já existentes como vocabulários, ontologias ou *profile applications*, que possam ser diretamente adotados ou estendidos para cobrir o domínio em causa.

Descoberta e seleção do modelo de dados pelo fornecedor do modelo de dados A segunda questão de co-criação compreende como é que um ator dentro da entidade de governo descobre e decide utilizar um determinado modelo de dados, correspondendo diretamente ao subprocesso de seleção do modelo de dados detalhado na arquitetura Unified Modeling Language (UML) do artefacto (5.3.3, subprocessos 1 e 2). A resposta a esta questão materializa-se na disponibilização de um *Vocabulary Hub*, onde os modelos aprovados e publicados ficam registados

com metadados de descoberta (ex. cobertura de domínio, licença, maturidade), permitindo ao fornecedor do modelo de dados do espaço pesquisar e avaliar a adequação de um modelo antes da sua adoção e implementação no *Vocabulary Hub* (*IDSA Knowledge Base*, [n.d.]; International Data Spaces Association, 2025).

Evolução do modelo perante a mudança do negócio. A terceira questão de co-criação identifica o que acontece quando o negócio evolui, reconhecendo que, embora o setor e o domínio de negócio se mantenham globalmente estáveis, a sua concretização operacional está sujeita a alterações pontuais que o modelo de dados tem de acomodar. Por esta razão, o artefacto trata os modelos semânticos geridos não como artefactos estáticos, mas como documentos vivos, sujeitos a um processo formal de controlo de versões e revisão sempre que uma alteração ao negócio exija a extensão, alteração ou criação de novos elementos no modelo, o que pode implicar, nos casos mais profundos de mudança, a revisão da própria ontologia subjacente.

Papéis e responsabilidades na governação do modelo A quarta questão de co-criação, aponta para quem, dentro da entidade de governo do espaço de dados, é responsável pela seleção, validação e manutenção dos modelos semânticos, reconhecendo que a governação semântica não pode ser atribuída a um único ator indiferenciado. O artefacto desenvolvido responde a esta questão desdobrando a entidade de governo do espaço de dados em dois papéis distintos e complementares: o **Fornecedor de Modelo de Dados**, responsável pela seleção, extensão e submissão dos modelos semânticos, e a **Autoridade de Governo**, responsável pela avaliação e validação da sua qualidade (5.2.3). Esta separação assegura que a proposta e a validação de um modelo nunca são da responsabilidade do mesmo ator, estabelecendo desde logo, ao nível da co-criação, o princípio de segregação de funções que estrutura toda a arquitetura do processo.

Esta visão de documento vivo estrutura-se em três fases do ciclo de vida do modelo semântico, formalizadas no artefacto:

- **Fase de Seleção e Triagem** - identificação e prospeção de modelos candidatos em repositórios abertos, sujeitos a filtros iniciais de cobertura de domínio e compatibilidade de licença;
- **Fase de Análise e Modelação** - avaliação da expressividade semântica e maturidade do modelo, que orienta o processo para a adoção direta, para a estruturação e refinamento de uma extensão, ou para a especificação de raiz;
- **Fase de Publicação e Preservação** - registo do modelo aprovado pela entidade de governo no *Vocabulary Hub*, com versionamento formal e ligação ao *Service Catalog* via a propriedade `dcterms:conformsTo`, alinhada com a norma DCATv3.

5.2.3 Papéis e Responsabilidades no Processo de Governo

Ao contrário da distinção genérica entre consumidores e fornecedores identificada na revisão de literatura 3.2.3), o processo aqui desenvolvido desdobra a entidade de governo em dois papéis distintos, cuja separação é central ao rigor do artefacto: o **Fornecedor do Modelo de Dados**, responsável pela seleção, extensão e submissão do modelo semântico; e a **Autoridade de Governo**, que atua num papel próximo do de um auditor, avaliando a qualidade do modelo submetido.

Ambos os atores são, formalmente, entidades de governo do espaço de dados, sendo que a distinção não é, portanto, entre uma entidade de governo e um agente externo, mas sim uma separação funcional dentro da própria estrutura de governação, justificada pela necessidade de assegurar que a entidade que propõe ou adapta um modelo não é também aquela que valida a sua conformidade e qualidade.

Para este efeito, num ecossistema em que o governo semântico é fundamental para garantir a interoperabilidade e a interpretação correta dos dados pelos múltiplos atores, torna-se necessário desdobrar tanto o Fornecedor de Modelo de Dados como a Autoridade de Governo no seu núcleo funcional. Em linha com o Capítulo 3 (Data Governance) do DAMA-DMBOK (Henderson et al., 2017) e com as suas recomendações para o desempenho da função de governo, procedeu-se à análise do papel do Data Steward enquanto elemento comum às duas entidades, ainda que com funções distintas e devidamente delimitadas em cada uma.

O Data Steward é definido como o responsável pela qualidade, definição, controlo e uso eficaz dos ativos de dados(Henderson et al., 2017), atuando neste processo dentro das duas entidades de governo do espaço de dados com funções diferenciadas: desde a execução e o desenvolvimento técnico, no Fornecedor de Modelo de Dados, até à supervisão e à definição de políticas, na Autoridade de Governo.

O Fornecedor de Modelo de Dados é responsável por criar, publicar e gerir o ciclo de vida dos modelos de dados (vocabulários, ontologias, Application Programming Interface (API) e esquemas) que viabilizam a comunicação semântica no espaço de dados. Nesta entidade, o papel do Data Steward consiste em definir os conceitos e termos que representam a especificidade de um domínio, assegurando simultaneamente a integridade dos ativos de dados, desde as bases de dados até aos próprios metadados que os representam. Este papel materializa-se através de atividades como a tradução dos termos de negócio em estruturas formais legíveis por máquina, recorrendo aos meta-padrões existentes no mercado; a gestão do ciclo de vida dos modelos, através do controlo da sua criação, evolução, publicação, versões e descontinuação; e a garantia da ponte entre a interoperabilidade semântica e técnica, através da tradução correta dos modelos para esquemas físicos de troca de dados, integrados nos protocolos de comunicação dos conectores do espaço. O Data Steward assume, assim, um duplo papel de atuação, relacionado tanto com o negócio como com a componente técnica da curadoria dos dados.

A Autoridade de Governo é o órgão máximo do espaço de dados, responsável por definir as regras através do Rulebook (Steinbuss, 2020; Data Spaces Support Centre, 2026), gerir o enquadramento operacional e assegurar a sustentabilidade e a confiança no espaço de dados. Nesta

entidade, o Data Steward assume um papel distinto, mais próximo do de um Enterprise Data Steward(Henderson et al., 2017), com uma visão focada no governo global, na supervisão transversal dos domínios de dados do espaço e na sua coordenação. As funções deste Data Steward, dentro da Autoridade de Governo, passam pela definição do Rulebook do ponto de vista semântico, com diretivas gerais, e pela recomendação ou exigência de normas internacionais, como as do W3C, ou de perfis europeus, como o DCAT-AP. É também responsável pela especificação dos requisitos mínimos de metadados que qualquer participante deve fornecer para que os seus produtos de dados sejam descritos, catalogados e descobertos no Vocabulary Hub, bem como pelo estabelecimento dos mecanismos formais para auditar se os produtos de dados partilhados cumprem as regras semânticas estabelecidas, recorrendo a ferramentas de validação. Cabe-lhe ainda a gestão de incoerências terminológicas, desvios ou conflitos de interpretação semântica entre os diferentes participantes ou fornecedores de modelos de dados. A distinção entre estas duas formas de curadoria de dados, aplicada à seleção de modelos de dados, fundamenta-se na divisão proposta pelo DAMA-DMBOK entre uma visão de governo, o fazer as coisas certas, e uma visão de gestão, o fazer bem as coisas(Henderson et al., 2017).

Desta forma, enquanto o Data Steward do Fornecedor de Modelo de Dados garante a qualidade e a usabilidade do modelo de dados, o Data Steward da Autoridade de Governo assegura a conformidade jurídica e o governo global desse mesmo modelo. Esta distinção reforça, ao nível operacional, o princípio de segregação de funções já estabelecido, aproximando-se do princípio da neutralidade consagrado no Artigo 12.º do DGA que impõe uma separação estrita de funções aos prestadores de serviços de intermediação de dados sem que tal implique, nesta fase, a conformidade formal do artefacto com esse artigo.

5.3 Arquitetura do Processo

A operacionalização do processo exige uma representação formal que assegure rigor visual, capacidade de auditoria e reprodutibilidade técnica por entidades gestoras. A presente secção detalha a arquitetura do processo modelada em UML 2.5.1, descrevendo a coordenação geral entre as raias de responsabilidade (*swimlanes*), os blocos de decisão e ciclos de iteração com limiar mínimo de qualidade, e mapeando de forma hierárquica os sete subprocessos que estruturam o ciclo de vida do modelo de dados.

5.3.1 Visão Geral da Arquitetura

A arquitetura do artefacto desenvolvido é formalizada através do padrão UML2.5.1(Eriksson et al., 2000; White et al., 2008), com o objetivo de assegurar rigor notacional, auditabilidade e capacidade de implementação por parte de uma entidade de governo de um espaço de dados. O processo global, intitulado *"Selecionar e Publicar Modelo de Dados do Espaço de Dados"*, encontra-se modelado a alto nível num diagrama de atividades organizado em raias de responsabilidade (*swimlanes*), decomposto hierarquicamente em sete subprocessos de detalhe (`CallBehaviorAction`) e suportado por uma legenda notacional própria (Figura5.4).

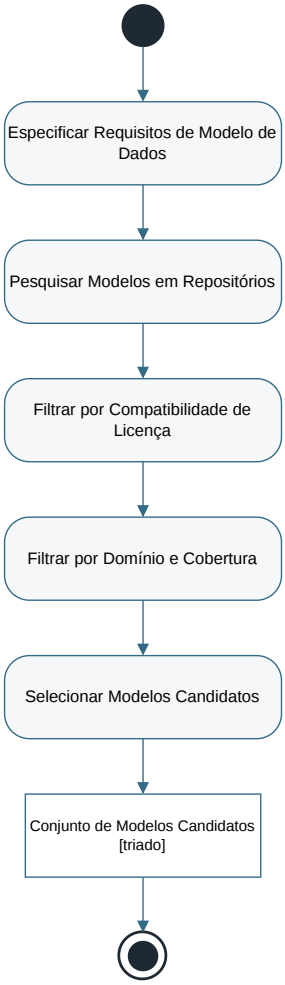


Figure 5.2: Selecionar Modelo Candidato

Os Subprocessos 2 a 7 encontram-se detalhados no Anexo B seguindo a mesma notação da figura 5.4.

O fluxo principal está organizado em duas partições operacionais, que traduzem a separação de deveres (*segregation of duties*) entre os dois atores de governo identificados no processo: a partição superior, correspondente ao **Fornecedor de Modelo de Dados**, responsável pela seleção, extensão e publicação do modelo; e a partição inferior, correspondente à **Autoridade de Governo do Espaço de Dados**, responsável pela avaliação de qualidade, validação e registo formal do modelo no catálogo do espaço de dados. Esta separação resulta da necessidade de garantir que a entidade que propõe ou adapta um modelo de dados não seja também aquela que valida a sua qualidade e aprova a sua publicação (5.2.3).

5.3.2 Fundamentação da Notação e Processo de Desenvolvimento

A escolha da notação UML2.5.1 decorre da necessidade de representar, com rigor formal, um processo de decisão distribuído entre dois atores organizacionais distintos, algo que a norma disponibiliza através do mecanismo de *swimlanes* e da tipologia completa de nós de controlo. Face a alternativas mais informais, a UML permite ainda a decomposição hierárquica do processo através de `CallBehaviorAction`, essencial para gerir a complexidade do processo sem sobrecarregar um único diagrama.

O desenvolvimento seguiu uma abordagem iterativa: as atividades foram identificadas a partir das questões de co-criação (secção5.2.2), organizadas segundo uma estrutura sujeito-predicado-objeto e traduzidas num esboço preliminar de fluxograma, posteriormente refinado em ciclos sucessivos - decomposição em subprocessos, redução de informação por elemento, distinção por cor das fases, e introdução de raias específicas para os dois atores de governo, disponível em anexo C.1. A validação técnica da notação foi complementada por um processo de revisão assistido por inteligência artificial generativa (Claude, Anthropic) (prompt utilizado em anexo A.1), utilizado para identificar inconsistências de notação UML, seguido de validação e decisão final pelo autor sobre cada alteração proposta.

5.3.3 Lógica de Decisão, Validação e Ciclos de Iteração

O processo segue a orientação estabelecida na *framework* do *Blueprint* para a fase de desenho do espaço de dados: não parte do desenvolvimento de um modelo de dados de raiz através de abordagens *top-down* ou *bottom-up* clássicas, mas antes da seleção de um modelo de dados candidato para reutilização e, se necessário, extensão.

O processo estrutura-se em três blocos de decisão:

1. **Avaliação de candidatos e adequação.** Após a atividade *Selecionar Modelo Candidato* (Subprocesso1), o nó `<Existem candidatos viáveis?>` verifica a presença de modelos disponíveis. Na ausência de candidatos, o controlo é desviado para `(Merge am0)`, ativando a atividade *Especificar Modelo de Raiz* (Subprocesso3). Existindo candidatos,

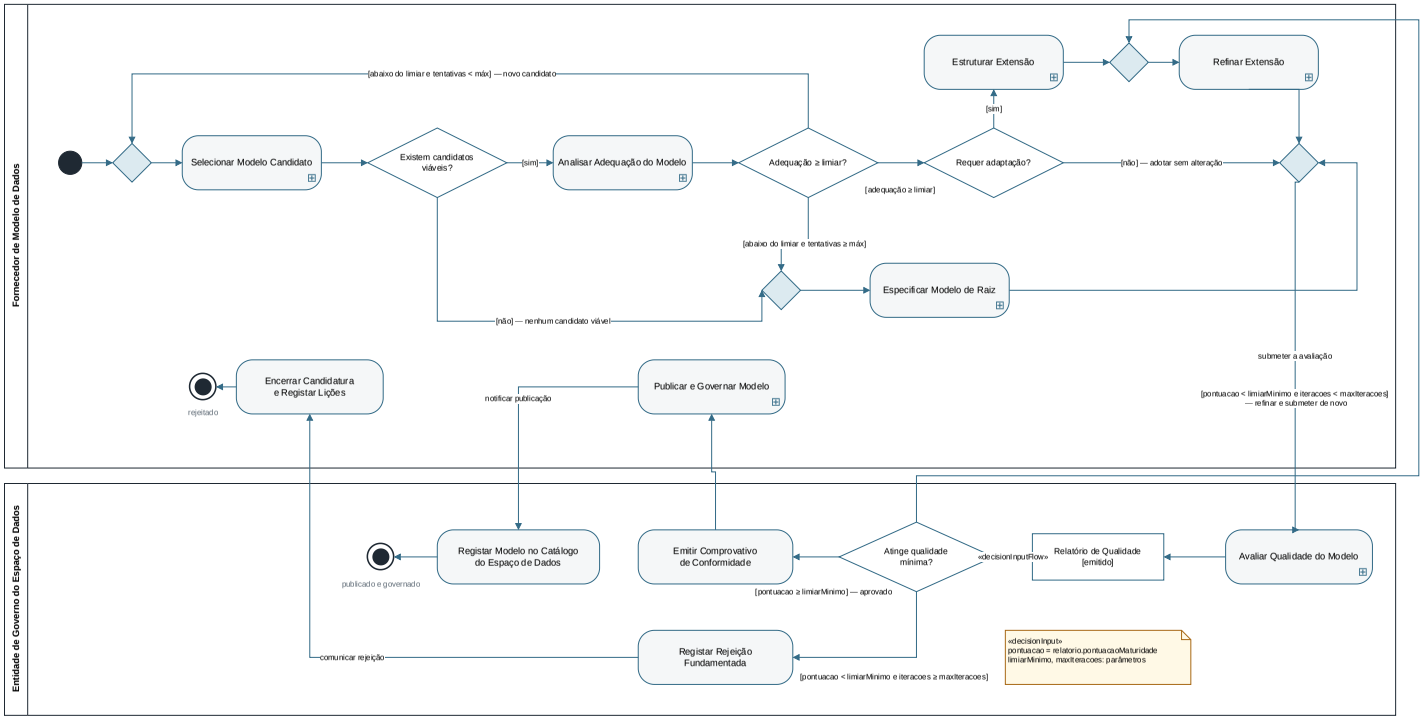


Figure 5.3: processo de atividades de alto nível do processo *Selecionar e Publicar Modelo de Dados do Espaço de Dados*, com as duas partições de responsabilidade.

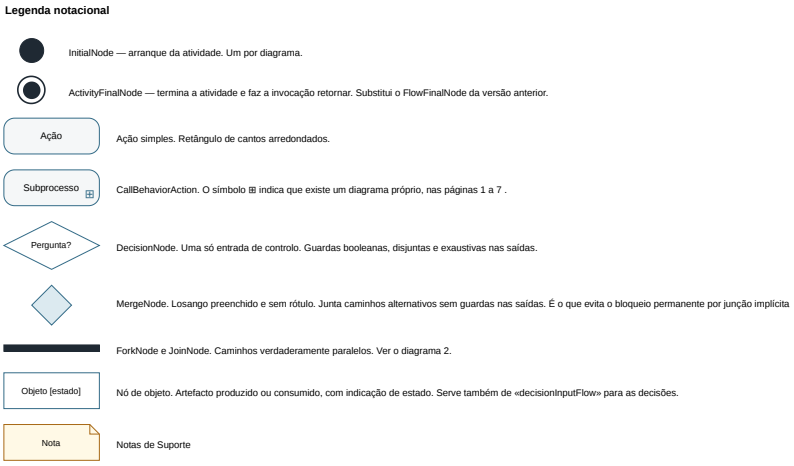


Figure 5.4: Legenda notacional utilizada nos diagramas UML do artefacto.

executa-se *Analisar Adequação do Modelo* (Subprocesso2), com um ciclo de retroalimentação até ao limite máximo de tentativas configurado.

2. **Ramificação de adaptação - extensão vs. adoção direta.** Estando o modelo acima do limiar de adequação, o nó `<Requer adaptação?>` determina o caminho: adoção direta via `(Merge am3)`, ou extensão através de *Estruturar Extensão* (Subprocesso4) e *Refinar Extensão* (Subprocesso5).

3. **Avaliação de qualidade e governação concorrente.** Todos os caminhos convergem em `(Merge am3)`, cruzam a fronteira de partição e são submetidos à Autoridade de Governo em *Avaliar Qualidade do Modelo* (Subprocesso6), que gera um *Relatório de Qualidade* [emitido], alimentando o nó `<Atinge qualidade mínima?>` através de um fluxo de dados de decisão (`«decisionInputFlow»`).

    - *Se aprovado*: a Entidade de Governo executa *Emitir Comprovativo de Conformidade*, devolvendo o controlo ao Fornecedor para *Publicar e Governar Modelo* (Subprocesso7), culminando no registo DCAT e no estado final *publicado e governado*.
    - *Se reprovado*: dentro do limite de iterações (`maxIteracoes`), o controlo regressa a `(Merge am2)` para novo refinamento; excedido o limite, a Entidade de Governo executa *Registar Rejeição Fundamentada*, conduzindo a *Encerrar Candidatura e Registar Lições* e ao estado final *rejeitado*.

Este desenho garante que nenhum modelo de dados entra em produção no espaço de dados sem validação independente da sua qualidade, ao mesmo tempo que evita ciclos de refinamento indefinidos através da limitação explícita do número de iterações permitidas.

Como referência de utilização neste processo do DAMA-DMBOK, permite dar visibilidade à aplicação do artefacto no ambiente dos espaços de dados. Como fase e consequente subprocesso

#	Category	Total score	Model score	%	Comments
1	How well does the model capture the requirements?	15			
2	How complete is the model?	15			
3	How well does the model match its scheme?	10			
4	How structurally sound is the model?	15			
5	How well does the model leverage generic structures?	10			
6	How well does the model follow naming standards?	5			
7	How well has the model been arranged for readability?	5			
8	How good are the definitions?	10			
9	How consistent is the model with the enterprise?	5			
10	How well does the Metadata match the data?	10			
	TOTAL SCORE	100			

Figure 5.5: Data Model Scorecard (Hoberman, 2015; Henderson et al., 2017)

de avaliação de qualidade do modelo, a utilização do Data Model Score Card torna-se essencial para validar metricamente se o modelo tem qualidade suficiente para integrar o espaço de dados. Esta ferramenta, criada por Steve Hoberman (Hoberman, 2015; Henderson et al., 2017), fornece uma metodologia estruturada através de uma avaliação que consiste em 10 categorias individuais e uma pontuação global acumulada de 100 pontos, sendo aplicada em modelos conceptuais, lógicos e físicos. O mesma poderá ser consultado no DAMA-DMBOK, e o conjunto de fichas de atividade serve como apontamento de onde utilizar este modelo.

Cada categoria tem uma pontuação máxima que pode ser atribuída, a pontuação total da a qualidade do modelo, percentagem de conformidade e ideias onde o mesmo pode ser melhorado, na secção de comentários. Os critérios que compõem os objetivos de avaliação são, de acordo com o DAMA-DMBOK:

O resultado final do Data Model Scorecard fornece uma visão quantitativa e qualitativa do estado real de maturidade e conformidade do modelo de dados, servindo como um mecanismo de segurança antes de ser publicado no Vocabulary Hub do espaço de dados e acessível à pesquisa pelos diferentes utilizadores.

5.3.4 Importância do limite de iterações nas decisões do processo

Nas decisões do processo e para garantir conformidade dentro do espaço de dados, é essencial definir e configurar um número limite de iterações. Definir este limite produz consequências diretas no fluxo de execução e na governação do espaço de dados, com destaque para a prevenção de retificações excessivas e controlo de custos, avaliação de capacidade técnica do proponente e agilidade no ecossistema. A modelação de dados é por natureza um processo iterativo. Contudo, o DAMA-DMBOK alerta que ciclos indeterminados de revisão geram custos indiretos elevados e desperdício de recursos.

Fixar um limite de 3 iterações concede ao Fornecedor de Modelo de Dados margem suficiente para, numa primeira iteração, receber o diagnóstico do relatório de qualidade como falhas no *Data Model Scorecard* ou omissões de metadados, numa segunda iteração aplicar correções estruturais e semânticas principais, e finalizar com ajustes residuais. Superar este limite sem atingir a qualidade mínima obriga a equipa a despender tempo a auditar novas submissões do mesmo modelo, sem que no entanto esteja a ocorrer uma evolução real.

Do ponto de vista de capacidade técnica, se após 3 tentativas formais, o modelo continuar a falhar na sua validação conceptual e de qualidade, conclui-se que o modelo carece de qualidade para ser publicado no *Vocabulary Hub* do espaço de dados, possuindo, assim, inconformidades de fundo que não são corrigíveis por via de pequenos ajustes, exigindo a seleção de um novo modelo ou uma reestruturação de raiz em última instância.

Do ponto de vista da engenharia de processos, o parâmetro (`maxIteracoes`) atua como uma condição de guarda determinante, que impede a ocorrência de ciclos infinitos, conjugando o esgotamento dos procedimentos destas iterações com uma pontuação inferior ao limiar do *Data Model Scorecard*. O fluxo interrompe automaticamente novas tentativas de refinamento e força a transição para a atividade Registar Rejeição Fundamentada (Ficha 8). Assim, são fundamentados os motivos de reprovação, terminando com o encerramento da candidatura e registo de falhas identificadas no repositório para conhecimento futuro.

Por último, salvaguarda-se a agilidade e a fluidez do processo, porque a permissão de iterações ilimitadas congestionaria a auditoria da Entidade de Governo, atrasando a publicação e versionamento de novos modelos de dados no *Vocabulary Hub*, comprometendo a prontidão operacional exigida para as transações em *Runtime*.

5.3.5 Mapeamento dos Subprocessos Decompostos

A Tabela 5.1 sintetiza os sete subprocessos de detalhe que decompõem o processo global, identificando para cada um o ator responsável, o objetivo funcional e os principais artefactos de entrada e saída.

Table 5.1: Mapeamento dos subprocessos decompostos do artefacto.

#	Subprocesso	Ator	Objetivo	Inputs	Outputs
1	Selecionar Modelo Candidato	Fornecedor	Pesquisar e identificar modelos de dados candidatos em catálogos semânticos federados	Requisitos de Partilha de Dados (necessidade de estabelecimento de nova transação ou interface no Espaço de Dados)	Conjunto de Modelos Candidatos [triado]
2	Analisar Adequação do Modelo	Fornecedor	Avaliar o alinhamento semântico e a maturidade dos elementos do modelo candidato selecionado em relação às necessidades de negócio, determinando a herança, compatibilidade e conformidade	Conjunto de Modelos Candidatos [triado]	Relatório de Adequação [emitido] (Decisão formal: "Requer Adaptação?" - Sim/Não)
3	Especificar Modelo Raiz	Fornecedor	Conceber, modelar e documentar um novo modelo de dados completo do zero sempre que o catálogo semântico não possua candidatos viáveis ou adaptáveis para responder aos requisitos de partilha de dados	Requisitos de Partilha de Dados (com transição originada pelo fluxo decisório: "Nenhum candidato viável encontrado")	Esquema de Raiz [especificado]
4	Estruturar Extensão	Fornecedor	Projetar as novas classes, atributos e relações necessários para estender o modelo candidato existente, assegurando a herança e as relações lógicas com o modelo base do Espaço de Dados	Relatório de Adequação [emitido] (com decisão "Requer Adaptação = Sim")	Esquema Estendido [estruturado]

continua na página seguinte

(continuação da Tabela 5.1)

#	Subprocesso	Ator	Objetivo	Inputs	Outputs
5	Refinar Extensão	Fornecedor	Aplicar regras de qualidade semântica, convenções linguísticas e princípios de desenho de metadados sobre a estrutura estendida, enriquecendo-a com metadados administrativos e exceções para submissão	Esquema Estendido [estruturado]	Esquema Estendido [refinado]
6	Avaliar Qualidade do Modelo	Entidade de Governo	Realizar a auditoria semântica e técnico-estrutural das propostas de modelo submetidas, aplicando testes formais e calculando métricas e score de maturidade	Esquema Estendido [refinado] OU Esquema de Raiz [especificado]	Relatório de Qualidade [emitido]
7	Publicar e Governar Modelo	Fornecedor / Entidade	Verificar alterações ao modelo, emitir comprovativos de conformidade, registar rejeição fundamentada e publicar o modelo no Vocabulary Hub	Nó de decisão: Atinge qualidade mínima? (Sim/Não)	Registo de Histórico / Publicação no Vocabulary Hub / DCAT

Para suportar cada um destes subprocessos (disponíveis em Anexo) e garantir robustez e acompanhamento adequado a cada uma das etapas de desenvolvimento, foram desenvolvidas um conjunto de "Fichas de atividade".

O ponto de partida para a construção de fichas de atividade foi a análise do processo de topo **"Selecionar e Publicar Modelo de Dados"**. O fluxo UML original tem mapeada a coreografia de alto nível entre duas entidades no Espaço de Dados. No entanto, para que este processo fosse reprodutível, auditável e estivesse em conformidade com as diretrizes europeias, foi necessário converter cada atividade abstrata num procedimento detalhado. A estrutura de Fichas de Atividade foi o instrumento escolhido para documentar de forma inequívoca o **Responsável**, o **Objetivo Geral**, as **Normas de Suporte**, os **Artefactos de Entrada** e os **Artefactos de Saída** de cada etapa.

Seguindo o mesmo principio da tabela anterior, as fichas reforçam com maior detalhe o objetivo de cada atividade dentro dos subprocessos, sendo acompanhadas pelo conjunto de standards, normas internacionais e legislação na garantia da realização de cada etapa de processo de forma adequada. Foram desenvolvidas onze fichas de atividade (ver Apêndice D para concretizar cada um dos sete subprocessos existentes com o maior detalhe possível.

Ficha de Atividade 1: Selecionar Modelo Candidato

- **Papel:** Fornecedor de Modelo de Dados (Submitter / Steward).
- **Objetivo:** Pesquisar e identificar modelos de dados candidatos no catálogo semântico federado (Vocabulary Hub) com base em requisitos e compatibilidade de licença, mitigando redundâncias e promovendo a reutilização de esquemas.
- **Normas:** ISO/IEC 11179-1:2023, ISO/IEC 11179-30:2023, ISO/IEC 11179-35:2023, ISO/IEC 11179-3:2023, ISO/IEC 11179-34:2024, ISO/IEC 11179-33:2023, ISO/IEC 11179-32:2023, ISO/IEC 11179-6:2023, ISO/IEC TR 19763-9, W3C DCAT v3, Data Act (Artigo 33.º, n.º 1, alínea b).
- **Artefacto Entrada:** Requisitos de Partilha de Dados (necessidade de estabelecimento de nova transação ou interface no Espaço de Dados).
- **Artefacto Saída:** Conjunto de Modelos Candidatos triado (identificado de forma unívoca através de URIs/IRDIs ativos no repositório de metadados).

Figure 5.6: Exemplo de Ficha de Atividade Simples

5.4 Alinhamento Normativo e Governação

Para assegurar que o processo não assenta em critérios arbitrários e garante a conformidade com as diretivas europeias, cada atividade foi ancorada em padrões internacionais de gestão de metadados e da Web Semântica. Esta secção explicita o papel orientador da família de normas ISO/IEC 11179, a articulação com os standards W3C (DCAT v3, SHACL, OWL, PROV-O) para suporte em tempo de execução, e a conformidade direta com os requisitos de interoperabilidade estipulados pelo Artigo 33.º do *Data Act*.

5.4.1 Normas Internacionais (International Standard Organization)

A utilização e recomendação de um conjunto de normas International Organization for Standardization (ISO), aplicadas nas fichas de atividade, colocam rigor ao processo, garantindo transparência na seleção e garante a utilidade de governo das práticas e seleção das famílias de normas partiu essencialmente das referências dentro das instruções do DAMA-DMBOK (com uma análise mais aprofundada a cada norma, com discussão e análise através do Notebook LM A.2).

A família de normas base adotada, partindo do capítulo de gestão de metadados, considera as instruções da família de normas ISO/IEC 11179, selecionadas e adquiridas para a sua análise a um nível mais profundo de conhecimento e para integração nas fichas de atividade como recomendação à entidade de Governo. De acordo com o que está descrito, a família de normas ISO/IEC 11179 é responsável pelo registo de descrições de dados, incluindo tanto a semântica quanto a representação desses dados. Sendo por meio destas descrições que se obtém uma compreensão precisa da finalidade e um referencial útil para os dados e estando cada uma das partes desta norma alinhada com os aspetos de avaliação, organização e registo de metadados, para descreverem os diferentes conjuntos de dados, referidos ao longo da dissertação como modelos de

dados. Cada uma das normas desta família é aplicada às fichas de atividade, de modo a dar suporte orientador nos diferentes aspetos e fases do processo (*ISO/IEC 11179-1*, [n.d.]).

Neste conjunto, surge a relação com outros standards para reforçar as recomendações para execução das diferentes atividades, para os diferentes elementos que compõem os modelos de dados.

A série de normas ISO/IEC 19763, com o objetivo de facilitar a interoperabilidade entre sistemas e pessoas, fornece um conjunto de especificações que permitem registar os modelos. De acordo com a mesma "a interoperabilidade neste contexto assume um sentido amplo, através da capacidade de comunicar, executar programas ou transferir dados entre várias unidades funcionais, sem que o utilizador necessite de grande conhecimento das características dessas unidades" (*ISO/IEC 19763-1*, [n.d.]).

A série ISO/IEC 21838 especifica os requisitos para um dos elementos centrais dos modelos de dados, neste caso as ontologias de nível superior. A parte 1 desta norma identifica e especifica as características necessárias de uma ontologia de nível superior neutra, em relação ao domínio que pode ser utilizado em conjunto com ontologias de nível inferior para apoiar as transações, recuperação, descoberta, integração e a análise de dados. O objetivo de garantir estes aspetos a quem irá utilizar o espaço de dados, viabiliza a necessidade da entidade de governo utilizar a recomendação desta norma na seleção do modelo de dados, seja no estado de reutilização e extensão, seja em último caso, na criação do modelo de raiz (*ISO/IEC 21838-1*, [n.d.])

As normas aplicadas permitem alinhar as atividades inerentes ao processo com as recomendações internacionais, garantindo a capacidade de atualização rápida caso seja necessário proceder a alterações, de forma auditável.

5.4.2 Padrões W3C

O estabelecimento de um ecossistema federado de Espaços de Dados que cumpra as exigências de interoperabilidade e soberania exige uma transição metodológica clara. Não basta modelar a governação e o ciclo de vida dos dados num plano abstrato e conceptual, sendo imperativo dotar o sistema de capacidade de execução em tempo de execução em Runtime. Esta operacionalização física é concretizada através da adoção em conjunto com as normas, dos standards da Web Semântica desenvolvidos pelo World Wide Web Consortium (W3C).

A integração destes padrões nas Fichas de Atividade do processo "Selecionar e Publicar Modelo de Dados" permite criar uma conexão de suporte entre as recomendações conferidas pelas normas, diretamente com os standards de troca de dados, organização e circulação na Web Semântica. Cada standard desempenha um papel específico, atuando nas diferentes etapas do ciclo de vida do processo para garantir que as trocas de informação automáticas, seguras e alinhadas com os objetivos do artigo 33 do Data Act.

Neste sentido, foram utilizados os seguintes standards de forma a reforçar as Fichas de Atividade:

W3C Resource Description Framework (RDF) (Resource Description Framework) e **RDFS (RDF Schema)** - O Resource Description Framework (RDF) atua como a linguagem universal

para a troca de dados e metadados no ecossistema. Ao estruturar a informação sob a forma de grafos direcionados compostos na forma de sujeito, predicado e objeto, o RDF unifica a representação de dados heterogéneos, permitindo a sua fusão lógica mesmo quando as origens físicas utilizam esquemas distintos. O RDF Schema (RDFS) atua como uma extensão direta, fornecendo o vocabulário básico de modelação de dados para descrever classes de recursos e propriedades, bem como relações de herança taxonómica através de propriedades como rdfs:subClassOf e rdfs:subPropertyOf (*RDF - Semantic Web Standards*, [n.d.]).

World Wide Web Consortium (W3C) Web Ontology Language (OWL) (Web Ontology Language) - Partindo de uma extensão RDFS, o OWL é um standard baseado na lógica descritiva desenvolvido especificamente para representar conhecimento rico e complexo sobre coisas, grupos de coisas e as relações entre elas, permitindo que os computadores processem o conteúdo de forma ativa, verificando a consistência lógica do modelo e deduzindo através de motores de inferência ou reasoners o conhecimento implícito que não estava originalmente presente de forma explícita (*OWL - Semantic Web Standards*, [n.d.]).

W3C Simple Knowledge Organization System (SKOS) (Simple Knowledge Organization System) - é um vocabulário RDF projetado especificamente para representar de forma intuitiva e partilhar sistemas de organização do conhecimento como tesauros, taxonomias, glossários e esquemas de classificação. Ao contrário da rigidez lógica do OWL, o SKOS foca-se na facilidade de mapeamento informal de termos de negócio através de relações de hierarquia e associação flexíveis (*SKOS Simple Knowledge Organization System Reference*, [n.d.]).

W3C SHACL (Shapes Constraint Language) - é o standard para descrever e validar grafos RDF contra um conjunto rigoroso de condições e formas lógicas. O SHACL distingue formalmente o grafo que contém as regras de validação do grafo de dados que se pretende auditar. O processamento do validador SHACL gera um relatório formal legível por máquina, com a indicação explícita se o grafo de dados está em conformidade (sh:conforms true/false) e aponta as falhas específicas através de instâncias de sh:ValidationResultm(*Shapes Constraint Language (SHACL)*, [n.d.]).

W3C DCAT (Data Catalog Vocabulary) v3 - é um vocabulário RDF estruturado para facilitar a interoperabilidade e a descoberta de catálogos de dados distribuídos na Web. O DCAT v3 assenta em três classes fundamentais: o catálogo (dcat:Catalog), que descreve e lista recursos; o conjunto de dados (dcat:Dataset), que descreve a coleção de dados conceptual; e a distribuição (dcat:Distribution), que descreve a forma física e o formato de acesso a esses dados (como o endpoint de uma API ou um ficheiro para download). Crucialmente, a versão v3 do DCAT introduz suporte oficial para representar o ciclo de vida e a gestão de versões de dados e metadados, utilizando propriedades explícitas como dcat:version, dcat:previousVersion e dcat:hasCurrentVersion (Steinbuss; Almeida, et al., 2025).

W3C PROV-O - é a especificação do W3C para representar e intercambiar informação sobre a origem e proveniência de de recursos gerados em sistemas heterogéneos. O modelo lógico do PROV-O assenta em três conceitos centrais: Entidades (prov:Entity, as coisas físicas ou digitais), Atividades (prov:Activity, os processos que modificam ou criam entidades) e Agentes (prov:Agent,

as pessoas ou sistemas responsáveis pela atividade) (*PROV-O: The PROV Ontology*, [n.d.]).

W3C Open Digital Rights Language (ODRL) (Open Digital Rights Language) - uma linguagem declarativa baseada em políticas semânticas que fornece um modelo de informação interoperável para exprimir e representar de forma rigorosa as condições de utilização, limites de direitos e licenças (compostas por permissões, proibições e obrigações) sobre ativos digitais e serviços (*ODRL Information Model 2.2*, [n.d.]).

Complementares Para além dos standards identificados anteriormente, foram identificados esquemas complementares a utilizar de modo a permitir uma maior integração nas fichas de atividade, tais como:

JSON Schema: Fornece o standard legível por máquina para estruturar, tipificar e impor validações sobre mensagens em JSON (*JSON Schema*, [n.d.]).

W3C Extensible Markup Language (XML) Schema (XSD): Define as regras gramaticais, restrições lógicas e tipos de dados estruturados para mensagens XML (*W3C XML Schema*, [n.d.]).

W3C CSVW (CSV on the Web) - Standard que fornece metadados estruturados em JSON-LD sobre ficheiros de dados tabulares (CSV/TSV), permitindo declarar namespaces, tipos de dados de colunas e regras de integridade referencial diretamente sobre ficheiros planos (*CSVW - CSV on the Web*, [n.d.]).

Os standards do W3C não foram introduzidos nesta arquitetura como meros formatos de ficheiro alternativos, mas sim como a infraestrutura de suporte que torna os metamodelos da ISO/IEC 11179 funcionais em Runtime. Ao fazê-lo, a solução de governo desenvolvida consolidada de forma robusta e integrada as barreiras de heterogeneidade semântica identificadas. Crucialmente, esta harmonização confere aos conectores e o espaço de dados as capacidades necessárias para cumprir eficazmente os requisitos de interoperabilidade estipulados pelo Artigo 33.º do EU Data Act (Casolari et al., 2023).

5.4.3 Cumprimento do Artigo 33º do Data Act

O Regulamento (UE) 2023/2854, relativo a regras harmonizadas sobre o acesso equitativo e a utilização de dados (*Data Act*)Casolari et al., 2023, estabelece no seu **Artigo 33º** os requisitos essenciais de interoperabilidade a que devem obedecer os participantes em espaços de dados. Em particular, a alínea (b) do n.º1 determina que as estruturas de dados, formatos, **vocabulários, esquemas de classificação, taxonomias e listas de códigos** sejam descritos de forma pública e consistente, requisito ao qual o processo de seleção, extensão e publicação de modelos de dados no *Vocabulary Hub* dá resposta operacional direta. A alínea (a) do mesmo número exige ainda que o conteúdo dos conjuntos de dados, restrições de utilização e licenças sejam descritos em formato passível de leitura automática, endereçado pela fase de triagem de licenças e pelo Relatório de Adequação (5.3.3).

5.5 Síntese de Resultados do Artefacto

Concluída a especificação técnica e normativa da arquitetura, importa consolidar a visão global dos componentes concebidos. Esta secção sintetiza os principais resultados alcançados pelo artefacto, evidenciando como a integração entre os sete subprocessos modelados, o referencial métrico do *Data Model Scorecard* e as onze fichas operacionais de atividade estabelece um quadro metodológico completo para a governação semântica setorial.

5.5.1 Síntese do Artefacto Desenvolvido

O desenvolvimento deste capítulo teve como base a implementação de um processo derivado da *framework* do Data Space Support Centre, através do qual foram identificados os aspetos essenciais para assegurar a interoperabilidade semântica na instanciação de espaços de dados setoriais, respeitando as especificidades próprias de cada domínio. O processo materializa e formaliza as decisões e os fluxos necessários à seleção de modelos de dados, contemplando as perspetivas de extensão e de reutilização, e considerando como elementos centrais os vocabulários, as ontologias e os *profile applications*, bem como os *data schemas*, na sua função de transição para a interoperabilidade técnica - evidenciada, mas não foco deste desenvolvimento.

A opção por definir este processo na etapa de desenho do espaço de dados, antes do estabelecimento de protocolos técnicos de troca de dados, estabelece uma ponte entre a interoperabilidade legal, como ponto de partida, e a interoperabilidade técnica, como ponto de chegada.

O processo desenvolvido, apoiado num conjunto de fichas de atividade, contempla de forma sequencial as fases de seleção, decisão sobre reutilização ou extensão, análise de adequação e de qualidade, culminando na publicação no *Vocabulary Hub* e no subsequente processo de governo.

O capítulo seguinte descreve o processo de validação do artefacto de forma descritiva, através de um caso de uso hipotético-dedutivo.

Chapter 6

Validação e Discussão do Artefacto

Tendo sido apresentada a conceção e a arquitetura do artefacto, o seguinte capítulo apresenta o processo de validação a que o mesmo foi submetido.

6.1 Caracterização e Cenário de Caso de Uso Hipotético

A demonstração do artefacto desenvolvido recorre a uma abordagem de demonstração e avaliação descritiva baseada em cenários de uso (illustrative scenario walkthrough), amplamente reconhecida na literatura metodológica de Design Science Research (Peffers et al., 2007; Hevner, 2007; Venable et al., 2012). Segundo o enquadramento de avaliação FEDS (Framework for Evaluation in Design Science), proposto por (Venable et al., 2012), esta opção configura uma avaliação do tipo ex ante e artificial, orientada para testar a completude lógica, a robustez procedimental e a aplicabilidade de artefactos sociotécnicos antes da sua instanciação em ambientes de produção.

Dado o estado de maturidade ainda residual dos espaços de dados setoriais e a consequente ausência de infraestruturas federadas plenamente operacionais no domínio em estudo a simulação de um percurso de governação sobre especificações industriais abertas (eBIZ, TRICK e CEN CWA 18291:2025) (*CWA 18291*, [n.d.]) constitui, nesta fase, o mecanismo de avaliação mais adequado disponível para demonstrar a capacidade do processo em mitigar a dissonância ontológica e assegurar a rastreabilidade normativa exigida pelo Artigo 33.º do Data Act.

6.1.1 Caso de uso exemplo

O caso de uso para aplicar o artefacto vai partir de uma comparação entre o "antes" (com as recomendações gerais que o DSSC aborda de forma geral) com o cenário "depois", aplicando o processo de governação semântica. Com base na documentação previamente selecionada como referencial, para o enquadramento do presente caso de uso é trabalhado com o desafio da rastreabilidade no Digital Product Passport (DPP) (Digital Product Passport).

6.1.2 Cenários

Cenário Um consórcio europeu de vestuário quer implementar o Passaporte Digital de Produto para rastrear o ciclo de vida de um lote de lã reciclada. Eles precisam de partilhar de forma federada dados como: identificação do lote, percentagem de conteúdo reciclado com certificado Global Recycled Standard (GRS) (Global Recycled Standard), pegada de carbono e identificação do fornecedor.

6.1.3 Cenário 1 — Só com as diretrizes da Blueprint 3.0

O DSSC Blueprint fornece um excelente referencial de alto nível. O consórcio, quando lê o bloco técnico de *Data Models*, vê referido "devem co-criar modelos de dados, dar prioridade à reutilização de padrões existentes, definir perfis de aplicação e publicar os metadados num catálogo". No entanto, a *blueprint* não diz como fazer isto de forma operacional e como fluxo estruturado, sendo que, sem um processo, o consórcio tenta aplicar as diretrizes de forma direta e *ad-hoc*, resultando no seguinte cenário:

1. Seleção por intuição — O consórcio sabe que o eBIZ é o padrão da indústria. Decidem usá-lo. Contudo não há uma análise formal de adequação. Mais tarde, a meio do desenvolvimento, apercebem-se de que o eBIZ não tem campos para pegada de CO2 nem para o certificado GRS.

Consequência: Perda de tempo e decisões tomadas por intuição técnica, sem um relatório de adequação que justifique a escolha aos parceiros de negócio.

2. Modelação caótica e desarticulada (sem Subprocessos 4 e 5) Cada departamento técnico das marcas do consórcio tenta estender o XML do eBIZ à sua maneira: a Marca A cria um campo chamado `CO2_Footprint` como texto livre; a Marca B cria um campo chamado `Emissoes` como valor decimal numa estrutura aninhada. Não há regras de nomenclatura uniformes (com a ausência de normas ISO/IEC 11179-5) nem dicionário de dados centralizado.

Consequência: Dissonância semântica extrema — Os sistemas informáticos das PME fornecedoras não conseguem ler as instruções de forma uniforme, gerando erros de interpretação e perdas de informação.

3. Falta de Controlo de Qualidade As extensões XML são colocadas diretamente num repositório partilhado (como o GitHub) assim que os programadores acabam de escrever o código. Não há validação de consistência semântica ou estrutural.

Consequência: Esquemas de dados instáveis, com erros de sintaxe ou contradições lógicas, entram em produção, gerando falhas nos motores de ERP das empresas e quebras de pipeline operacional.

4. **Rutura na Cadeia de Abastecimento** Uma grande marca decide que o campo de certificação GRS passa a ser obrigatório no XML e atualiza o esquema sem aviso estruturado. De repente, os sistemas de validação de produção automática das PME começam a rejeitar encomendas comerciais porque os seus ERPs antigos não reconhecem aquela tag obrigatória.

Consequência: Rutura operacional de transmissão de dados e exclusão de PMEs que não têm capacidade de adaptação imediata.

6.1.4 Cenário 2 — Aplicação do Processo

Neste cenário, o mesmo consórcio aplica o processo de governo para implementar o DPP de lã reciclada, operacionalizando as instruções do DSSC Blueprint:

- **Subprocesso 1 –** Identificar eBIZ como candidato
- **Subprocesso 2 –** Relatório de adequação "eBIZ cobre 80% do modelo de negócio - requer extensão de sustentabilidade"
- **Subprocesso 4 –** Core - subclasses Recycled Wool
- **Subprocesso 5 –** Adicionar atributos GRS e CO2 seguindo o core
- **Subprocesso 6 –** Validação de qualidade do modelo - Data Model Scorecard
- **Subprocesso 7 –** Publicação governada no Vocabulary Hub (versão minor retrocompatível)

1. Seleção e Adequação Científica (Subprocessos 1 e 2)

- O consórcio segue o Subprocesso 1 e filtra os modelos candidatos por licença e domínio, elegendo o eBIZ 2018.
- No Subprocesso 2, gera-se um relatório de adequação semântica. O consórcio comprova documentalmente que o eBIZ cobre a parte comercial, mas precisa de uma extensão estruturada para acomodar o DPP (alinhando-se com a especificação setorial CEN CEN Workshop Agreement (CWA) 18291:2025 do TRICK) (*CWA 18291*, [n.d.]).

2. Extensão Semântica Uniforme (Subprocessos 4 e 5)

- Aplicando as duas fases de Estruturar (Subprocesso 4) e Refinar (Subprocesso 5), os engenheiros criam uma extensão controlada: os campos `GRS_Certification` e `CO2_Footprint` são normalizados segundo as regras de formulação de nomes da ISO/IEC 11179-5.
- É documentado um dicionário de dados único e regras para validar valores das Pequenas e Médias Empresas (PME)s.
- *Vantagens:* Diminui-se a dissonância semântica. Todos os parceiros falam exatamente a mesma língua digital.

3. Garantia de Maturidade

- Antes do processo ser publicado, o esquema estendido é submetido à avaliação de qualidade do Subprocesso 6. É aplicada a grelha de métricas de adequação e maturidade. O modelo só avança quando atinge o limiar mínimo e recebe o comprovativo de conformidade.
- *Vantagens:* Os esquemas não são corrompidos ou mal desenhados chegando aos sistemas de produção.

4. Transição e Inclusão

No processo, a alteração é classificada como retrocompatível minor (os novos campos ecológicos são opcionais para quem ainda não emite DPP, mas estruturados para quem já utiliza).

- O modelo é publicado de forma governada no *Vocabulary Hub*. O consórcio garante a coexistência de versões e notifica formalmente todos os participantes com um plano de transição.
- *Vantagem:* Nenhuma PME é excluída do espaço de dados por incompatibilidade técnica repentina, nem a cadeia de abastecimento sofre sobressaltos operacionais.

Reconhece-se que, por ter sido construído pelo autor para ilustrar o funcionamento do processo, o Cenário 2 não constitui prova independente de eficácia, mas sim demonstração de coerência lógica interna entre o artefacto e os problemas identificados no Cenário 1.

6.2 Limites de demonstração e validação do artefacto

A construção/avaliação do artefacto, como mencionado, foi feita a partir de um caso de uso hipotético-dedutivo, a partir de um caso real aplicado ao modelo eBIZ através da extensão TRICK, que exemplifica o objetivo de extensão do modelo de dados para acomodar novos elementos de negócio.

O exemplo ilustra com exatidão a vantagem de se proceder às extensões, ao invés de criar um novo modelo de raiz, que levaria mais tempo a estar disponível no mercado. O desenvolvimento deste processo, traz uma visão estrutural e aplicável ao espaço de dados, ainda em desenvolvimento, que necessitam de regras claras antes de ser instanciados e entrarem em funcionamento. A construção deste processo envolveu uma profunda exploração das normas, técnicas de governo de dados, seleção de standards e adoção de boas práticas para limitar ao máximo as lacunas, dúvidas no momento de ação/instanciação, bem como na garantia de conformidade com os regulamentos europeus em vigor, que têm sido alterados nos últimos anos.

Com o processo a apontar para normas e standards de organizações internacionais, é possível um bom grau de atualização e acompanhar o desenvolvimento de diferentes tecnologias, bem como adaptar rapidamente sempre que surjam alterações processuais nos regulamentos daí a sua viabilidade de não ser um artefacto estático. O grau de construção do processo, ao ser pensado na sua adaptabilidade a diferentes negócios e setor de atividade, permite colocar empresas dentro do mesmo setor a comunicar de forma interoperável. Ao validar através de um caso de uso, é possível alterar e testar diferentes aspetos dos setores, ainda em fase de desenho, permitindo fazer alterações que vão de encontro às necessidades do negócio.

Para este aspeto, e como principais limitações a este desenvolvimento deste artefacto, faltaram claramente um conjunto de validações através de provas de conceito, bem como uma validação empírica com um ecossistema real que permitisse obter dados concretos, bem como uma amostra real da forma como este processo seria aplicado. Este processo fica também limitado ainda ao

que de forma propositada e definida a priori na implementação em espaços de dados intersetoriais. Outra limitação na elaboração deste processo, centra-se na potencialidade no momento de *runtime*, a partir do momento em que o catálogo é disponibilizado no Vocabulary Hub.

O processo está devidamente delimitado na interoperabilidade semântica, e num dos passos delimitados para o desenho dos espaços de dados. A instanciação do espaço de dados acaba por ser uma limitação apesar de questões de co-criação importantes ao desenvolvimento. Este processo acaba por fornecer o ponto final do passo anterior e dar o início de como o passo 3 para as trocas de dados, na definição dos protocolos para as trocas de dados, para suportar a interoperabilidade técnica. Este processo é inovador ao alinhar uma arquitetura de referência num estado mais maduro, por trabalhar diretamente as estratégias da gestão e governo de dados, e aplicar o conjunto de standards que permite trocar dados dotados de significado de forma interoperável.

O processo segue fielmente as recomendações do DSSC, com exploração em detalhe das normas a estabelecer no processo o governo e a gestão do ciclo de vida, bem como o desenho dos elementos de metadados para facilitar a pesquisa em catálogos, com recomendações assentes em regulamentos em mudança. As regras para a gestão dos modelos de dados, consideram as diferentes situações, bem como os diferentes níveis de abstração dentro do modelo.

6.3 Discussão: Resposta às Questões de Investigação

A validação concetual e procedimental do artefacto permite refletir de forma fundamentada sobre os objetivos formulados no início desta dissertação. A presente secção sintetiza as respostas às questões de investigação principal e complementares, demonstrando o valor operacional gerado para as entidades de governo de espaços de dados e evidenciando o contributo científico trazido para o domínio da Ciência da Informação.

6.3.1 Síntese das Respostas às Questões Formuladas

O desenvolvimento desta dissertação definiu como objetivo responder a um conjunto de questões de investigação, que funcionaram como linha condutora para o desenvolvimento do artefacto e para a sua posterior validação, seguindo as etapas da metodologia DSRM adotada para este efeito. Importa, para este propósito, recapitular as questões de investigação estabelecidas na secção4.3.1:

Questão Principal *De que modo podemos auxiliar as entidades de governo em espaços de dados setoriais para garantir a interoperabilidade semântica?*

Questões Complementares

- Que elementos-chave de suporte à interoperabilidade semântica deverão estar contemplados neste processo?
- De que modo os espaços de dados resolvem a dissonância ontológica e a interoperabilidade semântica entre sistemas heterogéneos ao nível setorial?

Em resposta à **questão principal**, o processo de governo semântico desenvolvido no Capítulo 5 suporta a autoridade de governo do espaço de dados ao formalizar, num fluxo auditável, o passo

B.4.Step-3 do modelo de co-criação do Blueprint 3.0, transformando uma diretriz genérica de reutilização e publicação de modelos de dados num processo operacional, com papéis segregados (Fornecedor de Modelo de Dados e Autoridade de Governo) e critérios de decisão explícitos. Em resposta à **primeira questão complementar**, os elementos-chave de suporte à interoperabilidade semântica contemplados no processo incluem: os vocabulários, ontologias e *profile applications* enquanto artefactos semânticos centrais; o *Vocabulary Hub* enquanto infraestrutura de descoberta e registo; os papéis de Data Steward, diferenciados entre as duas entidades de governo (5.2.3); e o conjunto de normas ISO/IEC e W3C que suportam as fichas de atividade (5.2.1 e 5.4.2).

Em resposta à **segunda questão complementar**, a resolução da dissonância ontológica foi demonstrada através da validação por caso de uso: o Cenário 1 6.1.3 evidenciou a ocorrência de dissonância semântica extrema quando o processo de extensão de um modelo de dados decorre de forma *ad-hoc*, ilustrada pela criação, sem regras de nomenclatura uniformes, de campos distintos (`CO2_Footprint` vs. `Emissoes`) para representar o mesmo conceito, enquanto o Cenário 2 6.1.4 demonstrou como a aplicação estruturada do processo, respetivamente através das regras de nomenclatura da ISO/IEC 11179-5 e da validação de qualidade do modelo, elimina essa dissonância antes da publicação. Assim, as questões de investigação foram respondidas através da construção do artefacto, das fichas de atividade e da sua demonstração por caso de uso. Ainda assim, esta demonstração teve natureza ilustrativa e hipotético-dedutiva, não empírica, pelo que se estabelece como trabalho futuro a validação do processo num contexto real, com dados efetivos de um ecossistema setorial em funcionamento.

Adicionalmente, e seguindo a visão da DSSC, será possível no futuro aprofundar esta linha de investigação através do desenvolvimento de processos de suporte às fases posteriores do modelo de co-criação, uma vez que o presente processo parte de um conjunto de recomendações para o desenho de espaços de dados que, embora aqui formalizadas, permanecem, nas etapas seguintes, definidas apenas de forma *ad-hoc* e sem estrutura processual própria. Em concreto, o trabalho seguinte poderia incidir sobre o desenvolvimento de um processo equivalente para o passo relativo ao estabelecimento dos protocolos de comunicação e troca de dados, etapa que fica propositadamente, fora do âmbito deste artefacto, cujo foco esteve inteiramente centrado na interoperabilidade semântica. Esta delimitação estabelece, de forma consistente, a fronteira face à interoperabilidade técnica, ainda que ambas as dimensões constituam, respetivamente, uma camada mais profunda e uma camada mais superficial de interoperabilidade na framework TSLO Singh et al., 2025.

6.3.2 Valor para a Governação Semântica Setorial

A necessidade de selecionar um modelo de dados adequado permite às organizações estabelecer mecanismos que possibilitam cumprir os objetivos europeus de interoperabilidade, assegurando a interoperabilidade semântica das suas comunicações. O processo reúne o conjunto de fluxos necessários - desde a seleção até à implementação - para que a entidade de governo, através do fornecedor do modelo de dados, possa assegurar a interoperabilidade semântica nas transações entre organizações, em alinhamento com os procedimentos do DAMA-DMBOK, com as normas ISO relevantes e com os requisitos do Artigo33.º do *Data Act* (5.4.3).

6.3.3 Contributo Científico para a Ciência da Informação

O desenvolvimento deste processo alinha a gestão de dados e de metadados, a organização e a recuperação de informação, sob uma perspetiva inter e transdisciplinar, plenamente alinhada com a evolução do ciclo de estudos em Ciência da Informação. Esta infraestrutura foi concebida para responder a uma das lacunas que a área procura resolver há décadas: a fragmentação da informação em silos dispersos, sem integração nem partilha efetivas.

A utilização de ontologias, vocabulários e *profile applications* constitui um aspeto essencial da organização de dados, informação, conhecimento e sabedoria, seguindo a hierarquia clássica DIKW (*Data-Information-Knowledge-Wisdom*) da Ciência da Informação, clarificando o modo como o processo desenvolvido enriquece a área, permitindo ultrapassar uma visão puramente centrada na engenharia de processos.

O capítulo seguinte fará o balanço de desenvolvimento desta dissertação, apresentando o conjunto de conclusões que a que o processo permitiu chegar, bem como o trabalho futuro a realizar na instanciação de espaços de dados setoriais, em diferentes aspetos de interoperabilidade e funcionalidades técnicas.

Chapter 7

Conclusões e Trabalho Futuro

7.1 Síntese do Trabalho Desenvolvido

A concretização da economia de dados europeia e a interoperabilidade sustentável nos espaços de dados dependem da preservação do significado da informação ao longo das trocas entre organizações. A presente dissertação respondeu a esta necessidade através do desenvolvimento de um artefacto metodológico original, com um processo formal de suporte à governação semântica para entidades gestoras de espaços de dados setoriais em fase de desenho (*Design Time*). Ancorado no método de co-criação do DSSC Blueprint3.0 e estruturado segundo as boas práticas do DAMA-DMBOK2 e na norma ISO/IEC 11179 e derivadas da mesma, o processo operacionaliza a seleção, extensão, auditoria de qualidade e publicação de modelos de dados através de sete subprocessos modelados em UML2.5.1, detalhados em fichas de atividade operacionais e no instrumento de avaliação como o *Data Model Scorecard*.

7.2 Contributos Principais

O trabalho apresenta contributos relevantes em duas vertentes complementares:

- **Contributo Teórico:** Contribui para a Ciência da Informação ao formalizar o ciclo de vida dos modelos de dados enquanto sistemas de organização e representação do conhecimento em arquiteturas federadas, demonstrando como a resolução da dissonância ontológica e a gestão de metadados atuam como pilares estruturantes na transição da hierarquia DIKW.
- **Contributo Prático:** Disponibiliza às autoridades de governo de espaços de dados um processo auditável que garante a segregação de funções entre quem fornece e quem valida o modelo de dados e operacionaliza os requisitos legais do Artigo 33º do *EU Data Act*, assegurando a compatibilidade entre normas e novos requisitos que possam surgir com o evoluir do negócio.

7.3 Limitações do Estudo

Reconhecem-se duas limitações metodológicas decorrentes da delimitação do âmbito:

- **Foco Estrito em *Design-Time*:** O processo encerra a sua intervenção na homologação e catálogo do modelo no *Vocabulary Hub*, não cobrindo a orquestração dinâmica de contratos ou o mapeamento em tempo real de execução (*Runtime*) executado pelos conectores técnicos.
- **Validação Descritiva *Ex-Ante*:** Face ao âmbito do estudo, que esteve centrado na parte do desenho de regras para os espaços de dados setoriais antes da instanciação e por alterações ao plano previsto inicialmente, a avaliação sustentou-se na simulação lógica de cenários (*walkthrough*), de forma hipotético-dedutiva sobre especificações industriais eBIZ/TRICK, carecendo de observação empírica sobre o comportamento real de equipas, bem como na validação com especialistas de domínio.

7.4 Linhas de Investigação Futura

Como continuidade deste trabalho, apontam-se duas direções prioritárias:

1. Execução empírica de entrevistas semi-estruturadas com especialistas de domínio, gestores e *Data Stewards* de consórcios industriais europeus para validação de usabilidade.
2. Replicação e teste real da capacidade de adaptação do processo noutros domínios caracterizados por cadeias de abastecimento fragmentadas, designadamente os setores do calçado e do mobiliário.

References

ALI, M; PAPAGEORGIOU, G; AZIZ, A; LOUKIS, E; CHARALABIDIS, Y; ALEXOPOULOS, C; PELLICER, FJL, 2024. A Framework for the Multi-Dimensional Assessment of Interoperability for Open Data Ecosystems Development. *INFORMATION POLITY*. Vol. 29, no. 4, pp. 439–466. ISSN 1570-1255. Available from DOI: 10.1177/15701255241297172.

AUER, Sören, 2022. Semantic Integration and Interoperability. In: OTTO, Boris; TEN HOMPEL, Michael; WROBEL, Stefan (eds.). *Designing Data Spaces* [online]. Cham: Springer International Publishing, pp. 195–210 [visited on 2025-12-02]. ISBN 978-3-030-93974-8 978-3-030-93975-5. Available from DOI: 10.1007/978-3-030-93975-5_12.

BADER, Sebastian; PULLMANN, Jaroslav; MADER, Christian; TRAMP, Sebastian; QUIX, Christoph; MÜLLER, Andreas W.; AKYÜREK, Haydar; BÖCKMANN, Matthias; IMBUSCH, Benedikt T.; LIPP, Johannes; GEISLER, Sandra; LANGE, Christoph, 2020. The International Data Spaces Information Model – An Ontology for Sovereign Exchange of Digital Content. In: PAN, Jeff Z.; TAMMA, Valentina; D'AMATO, Claudia; JANOWICZ, Krzysztof; FU, Bo; POLLERES, Axel; SENEVIRATNE, Oshani; KAGAL, Lalana (eds.). *The Semantic Web – ISWC 2020* [online]. Cham: Springer International Publishing. Vol. 12507, pp. 176–192 [visited on 2025-12-02]. ISBN 978-3-030-62465-1 978-3-030-62466-8. Available from DOI: 10.1007/978-3-030-62466-8_12.

BAIDATA, [n.d.]. *BAIDATA | International Data Spaces Association (IDSA) | Semantic Standardization of Data Spaces: Effective Interoperability* [https://baidata.eu/en/semantic-standardization-of-data-spaces-effective-interoperability]. [visited on 2026-04-08].

BORGES, Jorge Luis; BARREIROS, José Colaço, 2003. *Ficções / Jorge Luis Borges ; Trad. José Colaço Barreiros*. Porto: Público Comunicação Social. Colecção Mil Folhas. ISBN 84-96075-62-1.

BRECHTEL, Marc, 2025. From Hesitation to Participation in Industrial Data Ecosystems: Analysis of Motives and Incentives in the Automotive Industry. *Electronic Markets* [online]. Vol. 35, no. 1, p. 73 [visited on 2025-12-02]. ISSN 1019-6781, ISSN 1422-8890. Available from DOI: 10.1007/s12525-025-00818-3.

CASOLARI, F.; BUTTABONI, C.; FLORIDI, L., 2023. The EU Data Act in Context: A Legal Assessment. *International Journal of Law and Information Technology*. Vol. 31, no. 4, pp. 399–412. Available from DOI: 10.1093/ijlit/eaae005.

CASTELLVI, Silvia; KUNG, Antonio; ALANGOT, Bithin; SAMSON, Eric, 2025. *Data Spaces Standardization Landscape – Europe and International* [online]. 2025-07. [visited on 2025-10-01]. Tech. rep. Zenodo. Available from DOI: 10.5281/ZENODO.15856706.

CHIRVASUTA, T.; KURTEVA, A.; HOFMAN, W.; RUKANOVA, B.; TAN, Y.-H., 2025. Aligning the FEDeRATED Upper Ontology with Battery and Electronics Ontologies to Aid Circular Economy Monitoring in Practice. In: *Lecture Notes in Networks and Systems*. Springer Science and Business Media Deutschland GmbH. Vol. 1284 LNNS, pp. 44–61. Available from DOI: 10.1007/978-3-031-85363-0_4.

CIMMINO, Andrea; POVEDA-VILLALÓN, María; GARCÍA-CASTRO, Raúl, 2024. Ontologies and Semantic Interoperability. In: ZIEGLER, Sébastien; RADÓCZ, Renáta; QUESADA RODRIGUEZ, Adrian; MATHEU GARCIA, Sara Nieves (eds.). *Springer Handbook of Internet of Things* [online]. Cham: Springer International Publishing, pp. 407–425 [visited on 2025-11-07]. ISBN 978-3-031-39649-6 978-3-031-39650-2. Available from DOI: 10.1007/978-3-031-39650-2_17.

CORCHO, O.; EKAPUTRA, F.J.; HEIBI, I.; JONQUET, C.; MICSIK, A.; PERONI, S.; STORTI, E., 2024. A Maturity Model for Catalogues of Semantic Artefacts. *Scientific Data*. Vol. 11, no. 1. Available from DOI: 10.1038/s41597-024-03185-4.

CSVW - CSV on the Web, [n.d.] [https://csvw.org/]. [visited on 2026-09-06].

CURRY, Edward; SCERRI, Simon; TUIKKA, Tuomo, 2022. Data Spaces: Design, Deployment, and Future Directions. In: CURRY, Edward; SCERRI, Simon; TUIKKA, Tuomo (eds.). *Data Spaces : Design, Deployment and Future Directions*. Cham: Springer International Publishing, pp. 1–17. ISBN 978-3-030-98636-0. Available from DOI: 10.1007/978-3-030-98636-0_1.

CWA 18291, [n.d.]. *CWA 18291:2025 En* [online]. [visited on 2026-09-10]. Available from: https://www.nen.nl/en/cwa-18291-2025-en-344399.

DATA SPACES SUPPORT CENTRE, 2026. *DSSC Blueprint (Version 3.0)*. Report. European Union Digital Europe Programme.

DAVID, R.; ALEXIEV, V.; IVANOV, P.; VAN DEN BERG, W.; WIJBENGA, J.P.; STORNEBRINK, M., 2025. Federated Vocabulary Hubs as a Building Block for Semantic Layers in Data Spaces. In: *CEUR Workshop Proceedings*. CEUR-WS. Vol. 4007.

ERIKSSON, Hans-Erik; PENKER, Magnus, 2000. *Business Modeling with UML: Business Patterns at Work*. New York Weinheim: Wiley. ISBN 978-0-471-29551-8.

FIERENS, M., 2024. Initiating Interdisciplinary Research for Future-Proof Data Protection in the Context of Data Spaces and Semantic Interoperable Data Sharing. In: *CEUR Workshop Proceedings*. CEUR-WS. Vol. 3891.

FRANKLIN, Michael; HALEVY, Alon; MAIER, David, 2005. From Databases to Dataspaces: A New Abstraction for Information Management. *ACM SIGMOD Record* [online]. Vol. 34, no. 4, pp. 27–33 [visited on 2026-02-19]. ISSN 0163-5808. Available from DOI: 10.1145/1107499.1107502.

GANZHA, Maria; PAPRZYCKI, Marcin; PAWŁOWSKI, Wiesław; SOLARZ-NIESŁUCHOWSKI, Bartłomiej; SZMEJA, Paweł; WASIELEWSKA, Katarzyna, 2021. Semantic Interoperability. In: PALAU, Carlos E.; FORTINO, Giancarlo; MONTESINOS, Miguel; EXARCHAKOS, George; GIMÉNEZ, Pablo; MARKARIAN, Garik; CASTAY, Valérie; FUART, Flavio; PAWŁOWSKI, Wiesław; MORTARA, Marina; BASSI, Alessandro; GEVERS, Frans; IBÁÑEZ-SÁNCHEZ, Gema; HUET, Ignacio (eds.). *Interoperability of Heterogeneous IoT Platforms* [online]. Cham: Springer International Publishing, pp. 133–165 [visited on 2025-10-15]. ISBN 978-3-030-82445-7 978-3-030-82446-4. Available from DOI: 10.1007/978-3-030-82446-4_5.

GREGOR, Shirley; HEVNER, Alan R, 2013. Positioning and Presenting Design Science Research for Maximum Impact. *MIS quarterly*, pp. 337–355. ISSN 0276-7783.

HALEVY, Alon; FRANKLIN, Michael; MAIER, David, 2006. Principles of Dataspace Systems. In: *Proceedings of the Twenty-Fifth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* [online]. New York, NY, USA: Association for Computing Machinery, pp. 1–9 [visited on 2024-12-27]. PODS '06. ISBN 978-1-59593-318-8. Available from DOI: 10.1145/1142351.1142352.

HARMELINK, Rogier; JOOSTEN, Reinoud; TOPAN, Engin; ADRIAANSE, Arjen; VAN HILLEGERSBERG, Jos, 2024. Data: To Share or Not to Share? A Semi-Systematic Literature Review in (Rational) Data Sharing in Inter-Organizational Systems. *Discover Data* [online]. Vol. 2, no. 1, p. 13 [visited on 2025-10-15]. ISSN 2731-6955. Available from DOI: 10.1007/s44248-024-00018-y.

HEINZ, Daniel; BENZ, Carina; FASSNACHT, Marcel; SATZGER, Gerhard, 2022. *Past, Present and Future of Data Ecosystems Research: A Systematic Literature Review*. Available from DOI: 10.5445/IR/1000148750.

HENDERSON, Deborah; EARLEY, Susan; DATA ADMINISTRATION MANAGEMENT ASSOCIATION (eds.), 2017. *DAMA-DMBOK: Data Management Body of Knowledge*. Second edition. Basking Ridge, New Jersey: Technics Publications. ISBN 978-1-63462-236-3 978-1-63462-234-9 978-1-63462-247-9.

HEVNER, Alan, 2007. A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*. Vol. 19.

HMIMOU, Achraf; JOVANOVIC, Petar; NADAL, Sergi; ROMERO, Oscar; QUERALT, Anna, 2025. Operationalizing and Automating Data Validation in Data Spaces. *Data Science and Engineering* [online] [visited on 2025-12-02]. ISSN 2364-1185, ISSN 2364-1541. Available from DOI: 10.1007/s41019-025-00317-7.

HOBERMAN, S., 2015. *Data Model Scorecard: Applying the Industry Standard on Data Model Quality*. Technics Publications. ISBN 978-1-63462-082-6.

HOSEINI, S.; BURGDORF, A.; PAULUS, A.; MEISEN, T.; QUIX, C.; POMP, A., 2025. Challenges and Opportunities of LLM-Augmented Semantic Model Creation for Dataspaces. In: *Lecture Notes in Computer Science*. Springer Science and Business Media Deutschland GmbH. Vol. 15345 LNCS, pp. 183–200. Available from DOI: 10.1007/978-3-031-78955-7_17.

IDSA Knowledge Base, [n.d.] [https://kb.internationaldataspaces.org/]. [visited on 2026-09-06].

INTERNATIONAL DATA SPACES ASSOCIATION, 2025. *Front Matter | IDS-RAM 4 | IDS Knowledge Base* [https://docs.internationaldataspaces.org/ids-knowledgebase/ids-ram-4/frontmatter/frontmatter]. [visited on 2026-03-04].

ISO/IEC 11179-1, [n.d.]. *ISO/IEC 11179-1:2023* [https://www.iso.org/standard/78914.html]. [visited on 2026-09-06].

ISO/IEC 19763-1, [n.d.]. *ISO/IEC 19763-1:2023* [https://www.iso.org/standard/84749.html]. [visited on 2026-09-06].

ISO/IEC 21838-1, [n.d.]. *ISO/IEC 21838-1:2021* [https://www.iso.org/standard/71954.html]. [visited on 2026-09-06].

JSON Schema, [n.d.] [https://json-schema.org/]. [visited on 2026-09-06].

JUSSEN, Ilka; MÖLLER, Frederik; SCHWEIHOFF, Julia; GIESS, Anna; GIUSSANI, Giulia; OTTO, Boris, 2024. Issues in Inter-Organizational Data Sharing: Findings from Practice and Research Challenges. *Data & Knowledge Engineering* [online]. Vol. 150, p. 102280 [visited on 2024-05-27]. ISSN 0169-023X. Available from DOI: 10.1016/j.datak.2024.102280.

KERNSTOCK, P; HARMS, C; HEIN, A; KRCMAR, H, 2025. Establishing and Governing Data Ecosystems at the Crossroads of Centralization and Decentralization. *ELECTRONIC MARKETS*. Vol. 35, no. 1. ISSN 1019-6781. Available from DOI: 10.1007/s12525-025-00810-x.

KIRSTEIN, Fabian; BOHLEN, Vincent, 2022. IDS as a Foundation for Open Data Ecosystems. In: OTTO, Boris; TEN HOMPEL, Michael; WROBEL, Stefan (eds.). *Designing Data Spaces* [online]. Cham: Springer International Publishing, pp. 225–240 [visited on 2025-10-23]. ISBN 978-3-030-93974-8 978-3-030-93975-5. Available from DOI: 10.1007/978-3-030-93975-5_14.

KOVACH, A; MONTALVILLO, L; LANZA, J; SOTRES, P; URBIETA, A, 2026. Understanding Data Spaces: A Systematic Mapping Study of Foundations, Technical Building Blocks, and Sectoral Adoption. *COMPUTER SCIENCE REVIEW*. Vol. 59. ISSN 1574-0137. Available from DOI: 10.1016/j.cosrev.2025.100819.

LIS, D; OTTO, B, 2021. Towards a Taxonomy of Ecosystem Data Governance. In: BUI, TX (ed.). *PROCEEDINGS OF THE 54TH ANNUAL HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES*, pp. 6067–6076. ISBN 978-0-9981331-4-0.

LIS, Dominik; GELHAAR, Joshua; OTTO, Boris, 2023. Data Strategy and Policies: The Role of Data Governance in Data Ecosystems. In: CABALLERO, Ismael; PIATTINI, Mario (eds.). *Data Governance* [online]. Cham: Springer Nature Switzerland, pp. 27–55 [visited on 2025-10-23]. ISBN 978-3-031-43772-4 978-3-031-43773-1. Available from DOI: 10.1007/978-3-031-43773-1_2.

LNENICKA, M.; NIKIFOROVA, A.; LUTEREK, M.; MILIĆ, P.; RUDMARK, D.; NEUMAIER, S.; KEVIĆ, K.; ZUIDERWIJK, A.; RODRÍGUEZ BOLÍVAR, M.P., 2024. Understanding the Development of Public Data Ecosystems: From a Conceptual Model to a Six-Generation Model of the Evolution of Public Data Ecosystems. *Telematics and Informatics*. Vol. 94. Available from DOI: 10.1016/j.tele.2024.102190.

MÖLLER, Frederik; JUSSEN, Ilka; SPRINGER, Virginia; GIESS, Anna; SCHWEIHOFF, Julia Christina; GELHAAR, Joshua; GUGGENBERGER, Tobias; OTTO, Boris, 2024. Industrial Data Ecosystems and Data Spaces. *Electronic Markets* [online]. Vol. 34, no. 1, p. 41 [visited on 2025-09-23]. ISSN 1019-6781, ISSN 1422-8890. Available from DOI: 10.1007/s12525-024-00724-0.

NAGEL, Lars; LYCKLAMA, Douwe, 2022. How to Build, Run, and Govern Data Spaces. In: OTTO, Boris; TEN HOMPEL, Michael; WROBEL, Stefan (eds.). *Designing Data Spaces : The Ecosystem Approach to Competitive Advantage* [online]. Cham: Springer International Publishing, pp. 17–28 [visited on 2025-01-03]. ISBN 978-3-030-93975-5. Available from DOI: 10.1007/978-3-030-93975-5_2.

NIZAMIS, Alexandros; IOANNIDIS, Dimosthenis; GKONIS, Panagiotis; TRAKADAS, Panagiotis, 2024. Introducing an Enhanced Metadata Broker for Manufacturing Data Spaces. In: *Proceedings of the 4th Eclipse Security, AI, Architecture and Modelling Conference on Data Space* [online]. Mainz Germany: ACM, pp. 37–40 [visited on 2025-10-16]. ISBN 979-8-4007-0984-5. Available from DOI: 10.1145/3685651.3686699.

ODRL Information Model 2.2, [n.d.] [online]. [visited on 2026-09-10]. Available from: https://www.w3.org/TR/odrl-model/.

OLIVEIRA, MIS; LIMA, GDB; LOSCIO, BF, 2019. Investigations into Data Ecosystems: A Systematic Mapping Study. *KNOWLEDGE AND INFORMATION SYSTEMS*. Vol. 61, no. 2, pp. 589–630. ISSN 0219-1377. Available from DOI: 10.1007/s10115-018-1323-6.

OTTO, Boris, 2022. The Evolution of Data Spaces. In: OTTO, Boris; TEN HOMPEL, Michael; WROBEL, Stefan (eds.). *Designing Data Spaces* [online]. Cham: Springer International Publishing, pp. 3–15 [visited on 2025-12-02]. ISBN 978-3-030-93974-8 978-3-030-93975-5. Available from DOI: 10.1007/978-3-030-93975-5_1.

OWL - Semantic Web Standards, [n.d.] [https://www.w3.org/OWL/]. [visited on 2026-09-06].

PEFFERS, Ken; TUUNANEN, Tuure; ROTHENBERGER, Marcus A.; CHATTERJEE, Samir, 2007. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems* [online]. Vol. 24, no. 3, pp. 45–77 [visited on 2025-11-01]. ISSN 0742-1222, ISSN 1557-928X. Available from DOI: 10.2753/MIS0742-1222240302.

PETTENPOHL, Heinrich; SPIEKERMANN, Markus; BOTH, Jan Ruben, 2022. International Data Spaces in a Nutshell. In: OTTO, Boris; TEN HOMPEL, Michael; WROBEL, Stefan (eds.). *Designing Data Spaces* [online]. Cham: Springer International Publishing, pp. 29–40 [visited on 2025-10-23]. ISBN 978-3-030-93974-8 978-3-030-93975-5. Available from DOI: 10.1007/978-3-030-93975-5_3.

PROV-O: The PROV Ontology, [n.d.] [online]. [visited on 2026-09-10]. Available from: https://www.w3.org/TR/prov-o/.

RDF - Semantic Web Standards, [n.d.] [https://www.w3.org/RDF/]. [visited on 2026-09-06].

Shapes Constraint Language (SHACL), [n.d.] [https://www.w3.org/TR/shacl/]. [visited on 2026-09-06].

SINGH, Ashneet Khandpur; SÁNCHEZ, Marcel Ortiz; CAPARROS, Marc Garnica; DE LOS MOZOS, Mario Reyes; CASTELLVI, Silvia; DALMOLEN, Simon; TSIAKAS, Kosmas; ITSIOS, Paschalis; MARIOLIS, Ioannis; GIAKOUMIS, Dimitrios; RODRÍGUEZ, Silvia; VELASCO, Asier Aguayo; CABELLO, Daniel; LOPEZ, Borja Perez, 2025. A Security Analysis of European Data Space Architectures. *Data Science and Engineering* [online] [visited on 2025-12-02]. ISSN 2364-1185, ISSN 2364-1541. Available from DOI: 10.1007/s41019-025-00311-z.

SKOS Simple Knowledge Organization System Reference, [n.d.] [https://www.w3.org/TR/skos-reference/]. [visited on 2026-09-06].

STÄBLER, M.; KÖSTER, F.; SCHLUETER-LANGDON, C., 2023. Towards Solving Ontological Dissonance Using Network Graphs. In: Association for Information Systems.

STEINBUSS, Sebastian, 2020. *IDSA Rule Book* [online]. 2020-12. [visited on 2026-09-06]. Tech. rep. Zenodo. Available from DOI: 10.5281/ZENODO.5658294.

STEINBUSS, Sebastian; ALMEIDA, António Henrique; WOUTER, van den Berg; STHEFAN GABRIEL, Berwanger, 2025. *Semantic Interoperability in Data Spaces* [online]. 2025-11. [visited on 2026-01-22]. Tech. rep. Zenodo. Available from DOI: 10.5281/ZENODO.17630664.

STEINBUSS, Sebastian; TURKMAYALI, Anil, 2024. *Semantic Interoperability in Data Spaces* [online]. 2024-04. [visited on 2025-11-07]. Tech. rep. Zenodo. Available from DOI: 10.5281/ZENODO.10964377.

SUSHA, I; BROEK, TV; VAN VEENSTRA, AF; LINAKER, J, 2023. An Ecosystem Perspective on Developing Data Collaboratives for Addressing Societal Issues: The Role of Conveners. *GOVERNMENT INFORMATION QUARTERLY*. Vol. 40, no. 1. ISSN 0740-624X. Available from DOI: 10.1016/j.giq.2022.101763.

TONNARELLI, M.; KUMARA, I.; DRIESSEN, S.; TAMBURRI, D.A.; VAN DEN HEUVEL, W.-J.; OOR, P., 2025. Data Catalog Tools: A Systematic Multivocal Literature Review. *Journal of Systems and Software*. Vol. 230. Available from DOI: 10.1016/j.jss.2025.112584.

VENABLE, John; PRIES-HEJE, Jan; BASKERVILLE, Richard, 2012. A Comprehensive Framework for Evaluation in Design Science Research. In: HUTCHISON, David; KANADE, Takeo; KITTLER, Josef; KLEINBERG, Jon M.; MATTERN, Friedemann; MITCHELL, John C.; NAOR, Moni; NIERSTRASZ, Oscar; PANDU RANGAN, C.; STEFFEN, Bernhard; SUDAN, Madhu; TERZOPOULOS, Demetri; TYGAR, Doug; VARDI, Moshe Y.; WEIKUM, Gerhard; PEFFERS, Ken; ROTHENBERGER, Marcus; KUECHLER, Bill (eds.). *Design Science Research in Information Systems. Advances in Theory and Practice* [online]. Berlin, Heidelberg: Springer Berlin Heidelberg. Vol. 7286, pp. 423–438 [visited on 2026-09-06]. ISBN 978-3-642-29862-2 978-3-642-29863-9. Available from DOI: 10.1007/978-3-642-29863-9_31.

VOGT, Lars; STRÖMERT, Philip; MATENTZOGLU, Nicolas; KARAM, Naouel; KONRAD, Marcel; PRINZ, Manuel; BAUM, Roman, 2025. Suggestions for Extending the FAIR Principles Based on a Linguistic Perspective on Semantic Interoperability. *Scientific Data* [online]. Vol. 12, no. 1, p. 688 [visited on 2025-12-02]. ISSN 2052-4463. Available from DOI: 10.1038/s41597-025-05011-x.

W3C XML Schema, [n.d.] [https://www.w3.org/XML/Schema]. [visited on 2026-09-06].

WHITE, Stephen A.; MIERS, Derek, 2008. *BPMN Modeling and Reference Guide: Understanding and Using BPMN ; Develop Rigorous yet Understandable Graphical Representations of Business Processes*. Lighthouse Point, Fla: Future Strategies Inc. ISBN 978-0-9777527-2-0.

YAN, Dongdong; LI, Xiaofang; YANG, Qin; LIU, Ran, 2024. Key Technologies of Government Data Heterogeneity and Interoperation: A Survey. In: *2024 IEEE 7th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*. Vol. 7, pp. 1673–1679. Available from DOI: 10.1109/IAEAC59436.2024.10503635.

Appendix A

Prompts Utilizadas

A.1 Prompt de avaliação e melhoria do artefacto

1. I've attached a diagram that I'm working on for the data model selection process of the data space creation process. This is supposed to be a tool to support developers when going through this process in real applications. I'm trying to use UML sequence diagrams to standardize the notation process. Is the diagram well structured and following the UML specification?

2. Create a full report on the changes that need to happen in HTML, poiting to specific points in the process. Be detailed and critical. Please also write the report in Portuguese from Portugal.

3. Dadas estas recomendações dá-me um diagrama revisto por ti e que sirva este processo. Dá-me o resultado num formato compatível com o draw.io para eu possa depois fazer alterações.

A.2 Prompt Análise Normas

Atua como um especialista em Normas Internacionais (ISO/IEC), Arquiteturas Semânticas de Dados e Governação de Metadados.

Analisa o conjunto fornecido de normas ISO/IEC (focando especificamente em repositórios de metadados, estruturas ontológicas e interoperabilidade semântica) para extrair especificações técnicas operacionais para a construção de um processo formal de governação de modelos de dados em espaços de dados setoriais.

Para cada norma/parte analisada, extrai o nível máximo de detalhe estrutural e normativo, apresentando o resultado rigorosamente no seguinte formato:

1. Identificação da Norma:

- *Referência Completa (Número da Norma:Ano, Título Oficial).*

- *Âmbito Central e Objetivo Funcional (2–3 frases sintetizando o problema arquitetural exato que esta norma resolve).*

2. Metamodelo e Construtos Estruturais:

- *Principais Classes, Entidades do Metamodelo e Atributos (lista as entidades arquiteturais primárias definidas, tais como Model_Element, Administered_Item, Registration_State, etc.).*
- *Relações Explícitas e Restrições (especifica como as classes se relacionam, ex.: herança, composição, asserções e cardinalidades).*

3. Diretrizes Procedimentais e Operacionais:

- *Requisitos Obrigatórios vs. Recomendados (distingue os requisitos normativos obrigatórios [«shall»] das diretrizes informativas de recomendação [«should»]).*
- *Ciclo de Vida, Fluxos de Trabalho ou Estados Administrativos (ex.: transições de estado como Candidate, Recorded, Qualified; gestão de erros; regras de versionamento).*
- *Convenções de Nomenclatura Explícitas ou Regras de Definição de Dados (se aplicável, extrai a sintaxe, regras lexicais e critérios de não-circularidade).*

4. Citações Normativas Diretas:

- *Citação Direta 1 (definição central ou princípio basilar, incluindo o número da Cláusula/Subcláusula).*
- *Citação Direta 2 (regra operacional ou técnica de implementação, incluindo o número da Cláusula/Subcláusula).*

5. Interoperabilidade Semântica e Ligação aos Espaços de Dados:

- *Contributo específico para a resolução da Heterogeneidade Ontológica/Semântica.*
- *Pontos de contacto concretos com Espaços de Dados, Metadata Brokers ou Vocabulary Hubs Federados.*

No final, gera uma Matriz Comparativa Abrangente com as seguintes colunas:
| Norma/Parte | Papel Arquitetural Primário | Classes Principais do Metamodelo | Ciclo de Vida / Controlos de Qualidade | Aplicação Concreta na Governação de Dados |

Diretrizes:

- *Mantém rigor técnico e precisão normativa estritos.*
- *Não faças resumos vagos; nomeia explicitamente as cláusulas, classes, estruturas de dados e atributos de estado definidos no texto.*

- *Dá prioridade à densidade técnica e à reprodutibilidade procedimental em detrimento de descrições conceptuais genéricas.*

Appendix B

Subprocessos

Este anexo tem o objetivo de complementar o artefacto principal 5.3 desenvolvido ao longo do capítulo 5, bem como dar continuidade ao subprocesso de exemplo 5.2 e notação utilizada na 5.4, seguindo o mesmo padrão para a construção dos diferentes subprocessos.

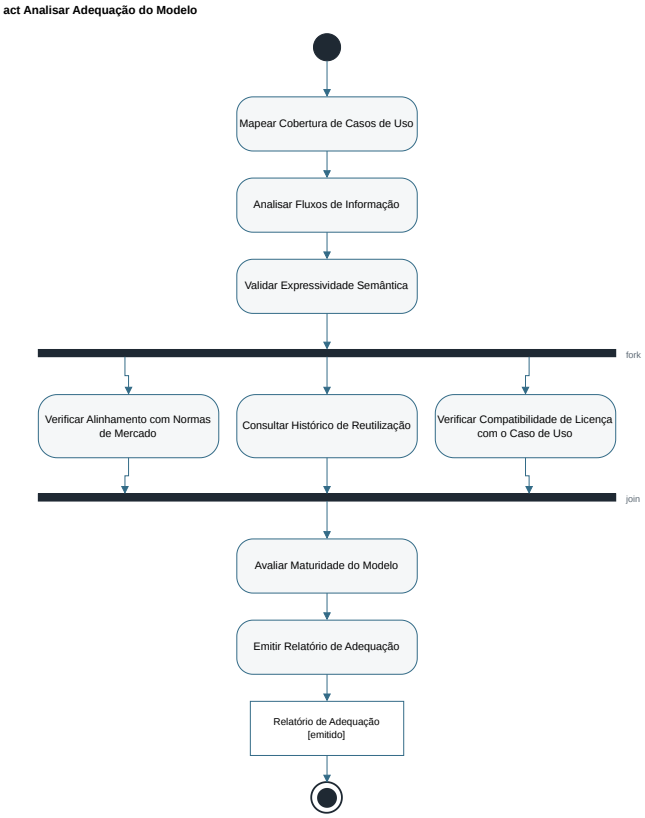


Figure B.1: Subprocesso 2: Analisar Adequação do Modelo.

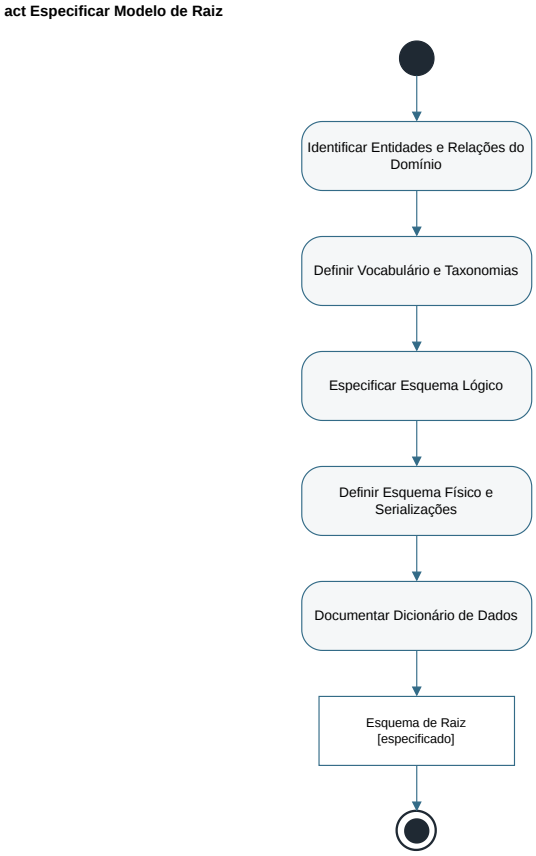


Figure B.2: Subprocesso 3: Especificar Modelo de Raiz.

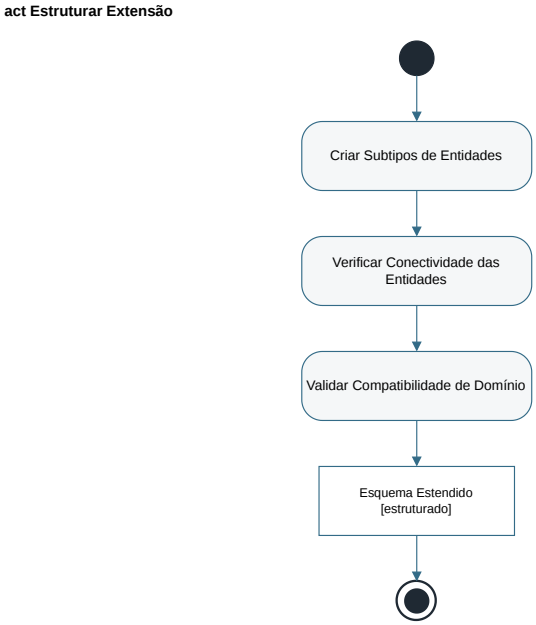


Figure B.3: Subprocesso 4: Estruturar Extensão.

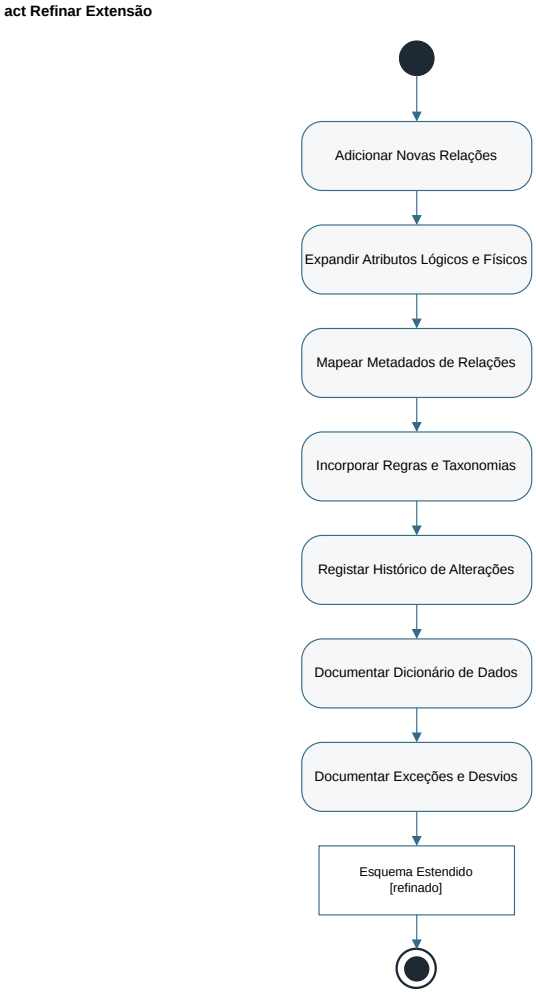


Figure B.4: Subprocesso 5: Refinar Extensão.

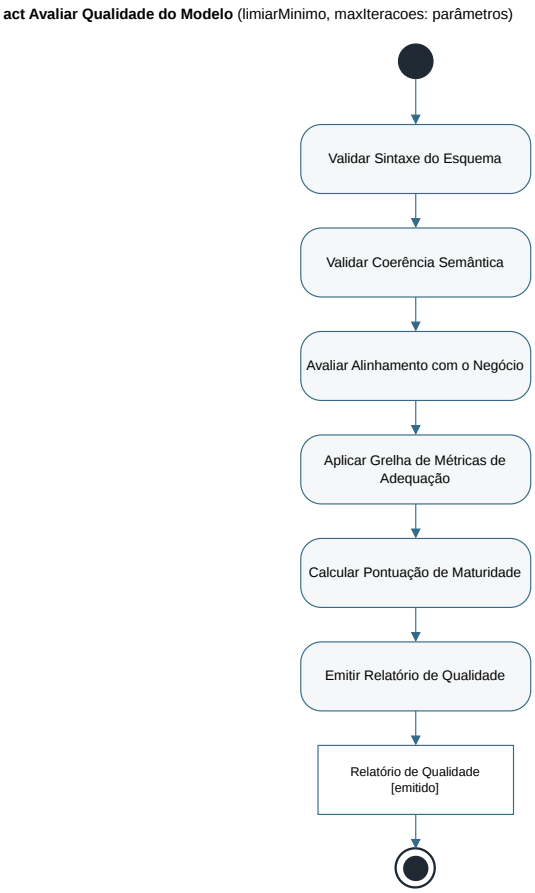


Figure B.5: Subprocesso 6: Avaliar Qualidade do Modelo.

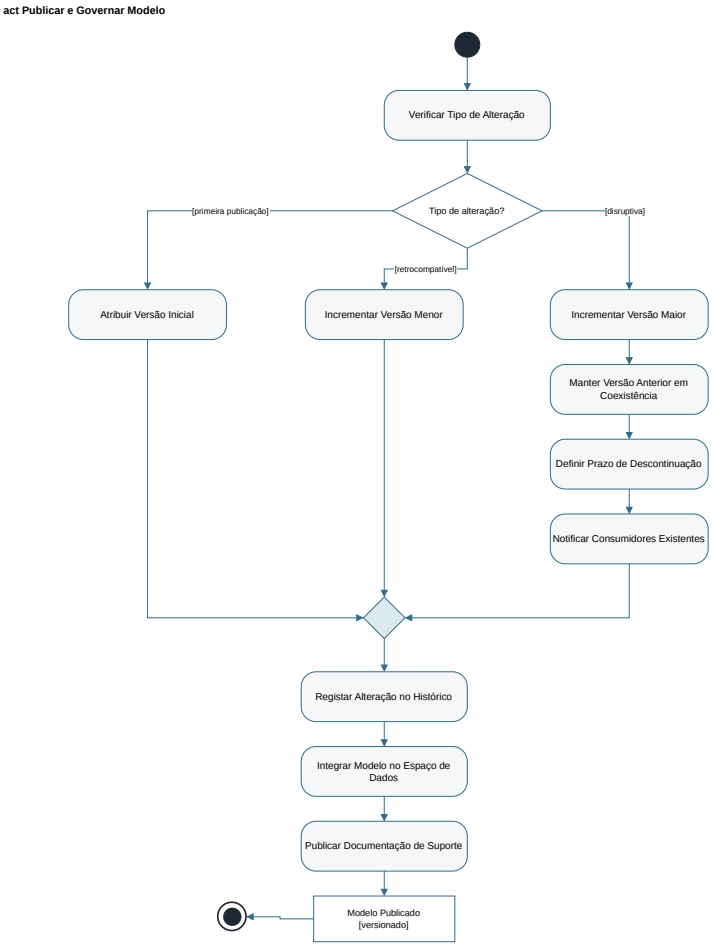


Figure B.6: Subprocesso 7: Publicar e Governar Modelo.

Appendix C

Processo Inicial Pré-Validação

Neste anexo encontra-se o processo que foi desenvolvido a priori com base às respostas ao *building block Data Models*, na análise da Blueprint 3.0 e nas boas práticas do DAMA-DMBOK, sendo este artefacto a base para o desenvolvimento do processo final.

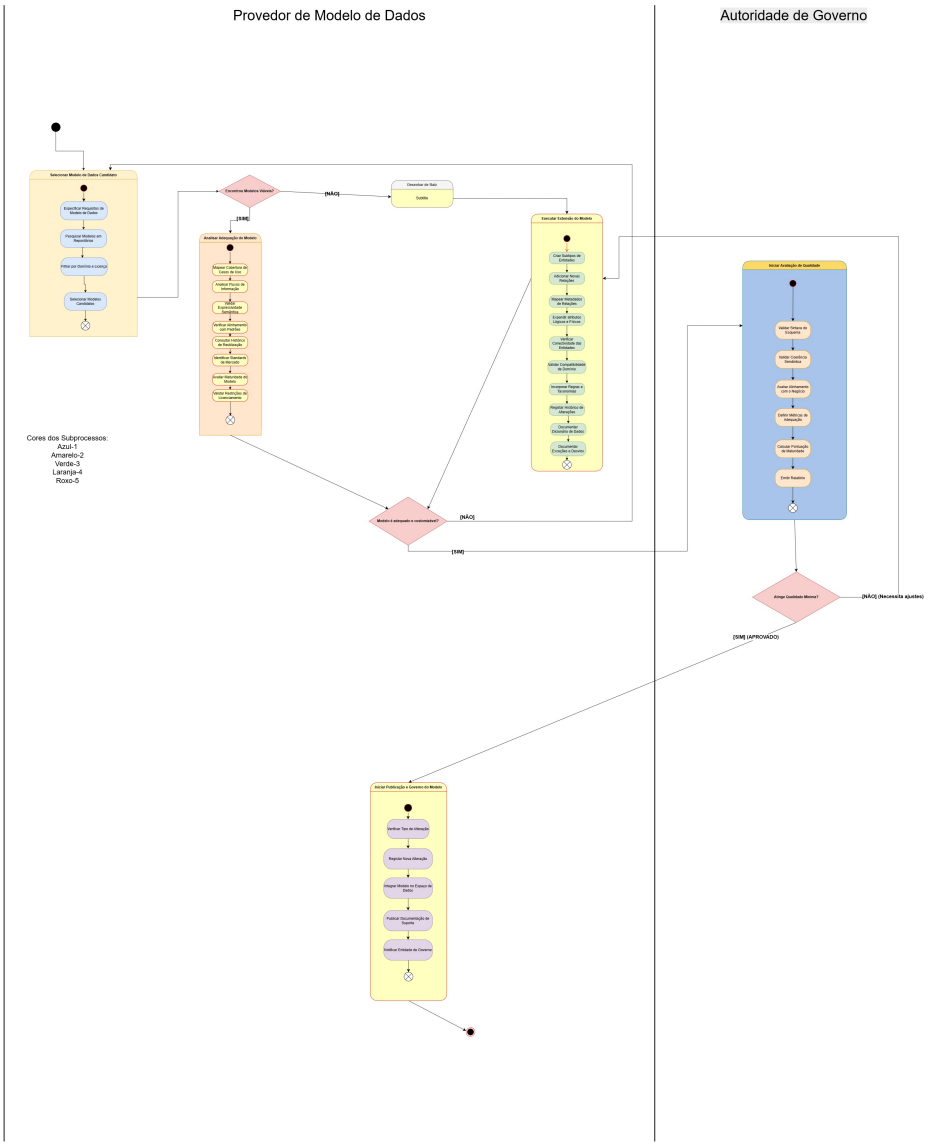


Figure C.1: Protótipo de Processo Pré-Validação

Appendix D

Fichas de Atividade

O seguinte anexo apresenta o conjunto de fichas de atividade de suporte ao processo 5.3, especificando de forma detalhada cada atividade dentro dos subprocessos mapeados, decorrentes do processo principal.

D.1

Ficha de Atividade 1: Selecionar Modelo Candidato

- **Papel:** Fornecedor de Modelo de Dados (*Submitter / Steward*).
- **Objetivo:** Pesquisar e identificar modelos de dados candidatos no catálogo semântico federado (*Vocabulary Hub*) com base em requisitos e compatibilidade de licença, mitigando redundâncias e promovendo a reutilização de esquemas.
- **Artefacto de Entrada Global (Nó de Objeto de Entrada):** Requisitos de Partilha de Dados (necessidade de estabelecimento de nova transação ou interface no Espaço de Dados).
- **Artefacto de Saída Global (Object Node):** Conjunto de Modelos Candidatos triado (identificado de forma unívoca através de URIs/IRDIs ativos no repositório de metadados).
- **Subprocessos e Atividades Internas Mapeados:**
  - **Especificar Requisitos de Modelo de Dados:** Levantamento dos requisitos lógicos e de domínio da partilha.
    Normas de Suporte: ISO/IEC 11179-1:2023 (Cláusula 7: Enquadramento conceptual) e ISO/IEC 11179-30:2023 (Identificação e caracterização conceptual preliminar).
  - **Pesquisar Modelos em Repositórios:** Busca ativa no repositório federado de metadados (*Vocabulary Hub* / *Metadata Broker*).
    Normas de Suporte: ISO/IEC 11179-35:2023 (Classes `Model` e `Modelling_Language` para a catalogação lógica de esquemas) e ISO/IEC 11179-3:2023 (Cláusula 11: *Item Mapping* para relações de equivalência semântica).

- **Filtrar por Compatibilidade de Licença:** Verificar se as licenças dos modelos candidatos admitem a reutilização e o tipo de partilha de dados pretendido.
  Normas de Suporte: ISO/IEC 11179-34:2024 (Licenciamento – classe `Reference_Document` de suporte ao filtro de compatibilidade de licença) e ISO/IEC 11179-33:2023 (Classe `Data_Set_Distribution` e seu atributo `rights` para validação de restrições).
- **Filtrar por Domínio e Cobertura:** Avaliar se o escopo dos modelos candidatos se adequa ao setor (ex: manufatura, mobilidade, etc.).
  Normas de Suporte: ISO/IEC 11179-3:2023 (Cláusula 10: *Classification_Scheme* para classificação contextual dos modelos) e ISO/IEC 11179-32:2023 (Importações de sistemas de conceitos via `imported_concept_system` para herança de domínios).
- **Selecionar Modelos Candidatos:** Seleção do modelo mais adequado para prosseguir para a análise detalhada de cobertura.
  Normas de Suporte: ISO/IEC 11179-6:2023 (Identificação unívoca de itens de metadados para garantir a rastreabilidade da seleção) e ISO/IEC TR 19763-9 (*On-demand model selection*).

D.2

Ficha de Atividade 2: Analisar Adequação do Modelo

- **Papel:** Fornecedor de Modelo de Dados (*Steward técnico do proponente*).
- **Objetivo Geral:** Avaliar o alinhamento semântico e a maturidade dos elementos do modelo candidato selecionado em relação às necessidades de negócio, determinando a herança, compatibilidade e conformidade.
- **Artefacto de Entrada Global (Nó de Objeto de Entrada):** Conjunto de Modelos Candidatos triado.
- **Artefacto de Saída Global (Object Node):** Relatório de Adequação emitido (Decisão formal: "Requer Adaptação?" – Sim/Não).
- **Subprocessos e Atividades Internas Mapeados:**
  - **Mapear Cobertura de Casos de Uso:** Mapeamento conceptual das classes e propriedades contra o caso de uso.
    Normas de Suporte: ISO/IEC 11179-35:2023 (Classes `Model_Element` e `Model_Element_Chara`
  - **Analisar Fluxos de Informação:** Detalhar a dinâmica de dados e transições descritas no modelo.
    Normas de Suporte: ISO/IEC 11179-34:2024 (Modelação de pipelines e dados computáveis – classe `Computation_Step` e fluxos de *input/output*).

- **Validar Expressividade Semântica:** Verificar a capacidade do modelo para expressar regras lógicas de forma inequívoca.
  Normas de Suporte: ISO/IEC 11179-32:2023 (Verificação de axiomas e restrições via `Assertion` e relações binárias).
- **Verificar Alinhamento com Normas de Mercado:** Validar a conformidade com esquemas semânticos padronizados do setor.
  Normas de Suporte: ISO/IEC 19763-12 (Metamodelo para registo de modelos de informação) e série ISO/IEC 21838 (*Top-level Ontologies* para validação conceptual contra ontologias superiores neutras).
- **Consultar Histórico de Reutilização:** Analisar se o modelo já foi aplicado de forma consistente noutros ecossistemas.
  Normas de Suporte: ISO/IEC 11179-6:2023 (Histórico de alterações, ciclos de vida e transições administrativas).
- **Verificar Compatibilidade de Licença com o Caso de Uso:** Assegurar que as restrições contratuais e direitos de acesso não colidem com o caso de uso real.
  Normas de Suporte: ISO/IEC 11179-33:2023 (Mapeamento de direitos e licenças em classes `Data_Set`).
- **Avaliar Maturidade do Modelo:** Realizar avaliação qualitativa do modelo candidato contra indicadores pré-estabelecidos.
  Normas de Suporte: ISO/IEC 11179-30:2023 (Avaliação de atributos de definição e de identificação) e referencial de maturidade para catálogos semânticos de Corcho et al. (2024).
- **Emitir Relatório de Adequação:** Consolidação formal dos pareceres e geração do veredito final.
  Normas de Suporte: ISO/IEC 11179-3:2023 (Geração de relatórios de mapeamento de registos comuns).

D.3

Ficha de Atividade 3: Especificar Modelo de Raiz

- **Papel:** Fornecedor de Modelo de Dados (*Submitter*).
- **Objetivo Geral:** Conceber, modelar e documentar um novo modelo de dados completo do zero sempre que o catálogo semântico não possua candidatos viáveis ou adaptáveis para responder aos requisitos de partilha de dados.
- **Artefacto de Entrada Global (Nó de Objeto de Entrada):** Requisitos de Partilha de Dados (com transição originada pelo fluxo decisório: “Nenhum candidato viável encontrado” no conector).

- **Artefacto de Saída Global (Object Node):** Esquema de Raiz especificado (*Schema*, diagramas de classe UML e metadados descritivos estruturados).
- **Subprocessos e Atividades Internas Mapeados:**
  - **Identificar Entidades e Relações do Domínio:** Isolar os conceitos basilares e as suas dependências do ponto de vista do domínio de negócio.
    Normas de Suporte: ISO/IEC 11179-35:2023 (Classes `Model` e `Model_Element`) e ISO 1087:2019 (Fundamentos de terminologia e ciência dos conceitos).
  - **Definir Vocabulário e Taxonomias:** Estruturar as relações hierárquicas e sinonímias entre conceitos.
    Normas de Suporte: ISO/IEC 11179-32:2023 (Instanciação de `Concept_System`, `Relation` e `Link` e mapeamento direto a hierarquias de conceitos SKOS).
  - **Especificar Schema Lógico:** Elaborar o modelo conceptual e lógico (ex: UML, diagramas de entidade-relação).
    Normas de Suporte: ISO/IEC 19763-12 (Metamodelo para registo de modelos de informação) e ISO/IEC 11179-31:2023 (Estruturação lógica de domínios lógicos conceituais).
  - **Definir Schema Físico e Serializações:** Mapear o modelo lógico em serializações técnicas acionáveis por computadores (XML Schema, JSON-LD, RDF/OWL).
    Normas de Suporte: ISO/IEC 11179-3:2023 (Tratamento de notação técnica através da classe `Notation`) e W3C DCAT v3 (Para padronização dos metadados de distribuição física de esquemas).
  - **Documentar Dicionário de Dados:** Descrever minuciosamente cada elemento, tipo de dado e regra semântica aplicável.
    Normas de Suporte: ISO/IEC 11179-4:2004 (Princípios e requisitos obrigatórios de formulação de definições de dados) e ISO/IEC 11179-5:2015 (Princípios de nomenclatura de metadados para garantir consistência lexical e sintática).

D.4

Ficha de Atividade 4: Estruturar Extensão

- **Papel:** Fornecedor de Modelo de Dados (*Submitter técnico*).
- **Objetivo Geral:** Projetar as novas classes, atributos e relações necessários para estender o modelo candidato existente, assegurando a herança e o acoplamento lógico relacional com o modelo base do Espaço de Dados.
- **Artefacto de Entrada Global (Nó de Objeto de Entrada):** Relatório de Adequação emitido (com decisão “Requer Adaptação = Sim”).

- **Artefacto de Saída Global (Object Node):** Esquema Estendido estruturado (diagrama UML de extensão com novos atributos e relações herdeiras desenhados).
- **Subprocessos e Atividades Internas Mapeados:**
  - **Criar Subtipos de Entidades:** Criar novos elementos de metadados herdando comportamentos de classes originais.
    Normas de Suporte: ISO/IEC 11179-35:2023 (Associações de composição e dependência de elementos: classe `Model_Element_Model_Relationship`) e ISO/IEC 11179-32:2023 (Classe `Link` para especializações em taxonomias).
  - **Verificar Conectividade das Entidades:** Validar as relações de dependência, integridade referencial e caminhos lógicos entre o modelo base e a extensão.
    Normas de Suporte: ISO/IEC 11179-32:2023 (Relações binárias e assertivas via `Binary_Relation` e `Assertion`).
  - **Validar Compatibilidade de Domínio:** Garantir que as restrições lógicas e valores permissíveis dos novos atributos são compatíveis com a semântica de base do Espaço de Dados.
    Normas de Suporte: ISO/IEC 11179-31:2023 (Restrições entre domínios conceituais e domínios de valores).

D.5

Ficha de Atividade 5: Refinar Extensão

- **Papel:** Fornecedor de Modelo de Dados (*Submitter e Steward*).
- **Objetivo Geral:** Aplicar regras de qualidade semântica, convenções linguísticas e princípios de desenho de metadados sobre a estrutura estendida, enriquecendo-a com metadados administrativos, dicionalização e exceções para submissão.
- **Artefacto de Entrada Global (Nó de Objeto de Entrada):** Esquema Estendido estruturado.
- **Artefacto de Saída Global (Object Node):** Esquema Estendido refinado (ficheiro XML/RDF de metadados pronto para submissão à Entidade de Governo).
- **Subprocessos e Atividades Internas Mapeados:**
  - **Adicionar Novas Relações:** Enriquecer o modelo com ligações semânticas adicionais.
    Normas de Suporte: ISO/IEC 11179-32:2023 (Mapeamento de relações do SKOS-CORE, como `skos:broader`, `skos:narrower`, `skos:related`).

- **Expandir Atributos Lógicos e Físicos:** Atribuir caraterísticas de tipologia de dados e cardinalidade.
  Normas de Suporte: ISO/IEC 11179-31:2023 (Atributos de herança da classe de representação `Value_Meaning`).
- **Mapear Metadados de Relações:** Descrever o significado e a proveniência dos relacionamentos criados.
  Normas de Suporte: ISO/IEC 11179-3:2023 (Classes comuns da Cláusula 8: `Designation` e `Definition` para modelação das relações).
- **Incorporar Regras e Taxonomias:** Enquadrar a extensão em taxonomias controladas ou vocabulários padrão setoriais.
  Normas de Suporte: ISO/IEC 11179-3:2023 (Cláusula 10: `Classification_Scheme_Item` e `item_classification`) e ISO/IEC TR 20943-6 (*Framework for generating ontologies*).
- **Registar Histórico de Alterações:** Registar a linhagem e mudanças realizadas ao modelo de dados.
  Normas de Suporte: W3C PROV-O (Utilização de ontologia de proveniência básica) e ISO/IEC 11179-3:2023 (Campos comuns `explanatory_comment` e `origin`).
- **Documentar Dicionário de Dados:** Confeção rigorosa das definições textuais dos novos elementos e propriedades.
  Normas de Suporte: ISO/IEC 11179-4:2004 (Formulação e validação de definições de dados, regras de clareza, concisão e não-circularidade) e ISO/IEC 11179-5:2015 (Consistência lexical e regras lexicais para designações).
- **Documentar Exceções e Desvios:** Identificar e formalizar os pontos onde a extensão se desvia do padrão original por motivos operacionais do caso de uso.
  Normas de Suporte: ISO/IEC 11179-3:2023 (Uso da classe `Constraint_Set` para documentação de desvios).

D.6

Ficha de Atividade 6: Avaliar Qualidade do Modelo

- **Papel:** Entidade de Governo do Espaço de Dados (*Steward auditor e Control Committee*).
- **Objetivo Geral:** Realizar a auditoria semântica e técnico-estrutural das propostas de modelo submetidas, aplicando testes formais e calculando métricas e score de maturidade.
- **Artefacto de Entrada Global (Nó de Objeto de Entrada):** Esquema Estendido refinado OU Esquema de Raiz especificado.
- **Artefacto de Saída Global (Object Node):** Relatório de Qualidade emitido (Parecer final de aprovação ou rejeição, contendo desvios e score numérico).

- **Subprocessos e Atividades Internas Mapeados:**
  - **Validar Sintaxe do Esquema:** Verificar a conformidade do arquivo submetido em termos de legibilidade técnica básica (schemas XML, sintaxe Turtle/RDF).
    Normas de Suporte: ISO/IEC 11179-3:2023 (Classe de representação `Notation`) e normas de validação W3C para serializações.
  - **Validar Coerência Semântica:** Auditar a ausência de ambiguidades, circularidade de termos ou desalinhamentos lógicos no grafo conceptual.
    Normas de Suporte: ISO/IEC 11179-4:2004 (Requisitos obrigatórios de formulação e rejeição de circularidades) e ISO/IEC 11179-32:2023 (Validação de heranças e relações lógicas via `Assertion`).
  - **Avaliar Alinhamento com o Negócio:** Garantir que o modelo proposto cobre as transações reais e a semântica da cadeia de suprimentos sectorial.
    Normas de Suporte: ISO/IEC 11179-35:2023 (Mapeamento de relações e dependências de negócio).
  - **Aplicar Grelha de Métricas de Adequação:** Medição sistemática dos atributos de qualidade semântica.
    Normas de Suporte: Diretrizes de qualidade de metadados do Anexo C da ISO/IEC 11179-6:2023 e referencial de qualidade terminológica da ISO 1087:2019.
  - **Calcular Pontuação de Maturidade:** Obtenção de um *score* ponderado que determina se o modelo candidato cumpre o limiar mínimo de governação.
    Normas de Suporte: ISO/IEC 11179-6:2023 (Critérios de avanço no ciclo de maturidade) e referencial de catálogos de Corcho et al. (2024).
  - **Emitir Relatório de Qualidade:** Consolidação formal de todos os pareceres, que alimenta diretamente o nó de decisão "Atinge qualidade mínima?" do processo de topo.
    Normas de Suporte: ISO/IEC 11179-3:2023 (Registo administrativo e de qualidade comuns).

D.7

Ficha de Atividade 7: Emitir Comprovativo de Conformidade

- **Papel:** Entidade de Governo do Espaço de Dados (*Registar responsável*).
- **Objetivo Geral:** Formalizar a aprovação do modelo ou extensão avaliada, emitindo o certificado digital de conformidade semântica e promovendo o ciclo de vida do metadado para o estado *Recorded* ou *Qualified*.
- **Artefacto de Entrada Global (Nó de Objeto de Entrada):** Relatório de Qualidade emitido (com parecer positivo com base no cálculo do limiar semântico).

- **Artefacto de Saída Global (Object Node):** Comprovativo de Conformidade Semântica Emitido (certificado digital assinado eletronicamente e associado unicamente à proposta).
- **Normas e Referenciais de Suporte:**
  - **ISO/IEC 11179-6:2023:** Categorias de ciclo de vida e estado administrativo de registo (transição para *Recorded* e *Qualified*).
  - **ISO/IEC 11179-3:2023:** Cláusula 9 (classe `Registration_State` e classe de enumeração `Registration_Status`).

D.8

Ficha de Atividade 8: Registar Rejeição Fundamentada

- **Papel:** Entidade de Governo do Espaço de Dados (*Registar / Steward com funções de arbitragem*).
- **Objetivo Geral:** Documentar no sistema de governação semântica as razões de não-conformidade técnica e as falhas de qualidade identificadas na proposta do modelo de dados, registando a falha de integridade, decrementando tentativas e gerindo o loop de retificação.
- **Artefacto de Entrada Global (Nó de Objeto de Entrada):** Relatório de Qualidade emitido (com parecer de rejeição, quando a pontuação de qualidade se situa abaixo do limiar de maturidade mínimo).
- **Artefacto de Saída Global (Object Node):** Notificação de Rejeição Fundamentada Registada (parecer técnico fundamentado com a indicação detalhada das falhas lógicas e o número de tentativas restantes de loop de processo).
- **Normas e Referenciais de Suporte:**
  - **ISO/IEC 11179-6:2023:** Cláusula D.1.2 (Procedimentos para registo de problemas, erros terminológicos e gestão do estado de registo).
  - **ISO/IEC 11179-3:2023:** Classe `Registration_State` e instanciação dos atributos `unresolved_issue` e `administrative_note`.
  - **ISO/IEC 11179-4:** Diretrizes de formulação e requisitos textuais para notas administrativas e de correção.

D.9

Ficha de Atividade 9: Registar Modelo no Catálogo do Espaço de Dados

- **Papel:** Entidade de Governo do Espaço de Dados (*Registar responsável pelo catálogo comum*).
- **Objetivo Geral:** Persistir lógica e fisicamente as classes, propriedades e relações do modelo semântico aprovado na base de dados centralizada ou federada de metadados do Espaço de Dados, tornando-o um elemento descoberto, rastreável e provido de histórico de linhagem.
- **Artefacto de Entrada Global (Nó de Objeto de Entrada):** Comprovativo de Conformidade Semântica Emitido.
- **Artefacto de Saída Global (Object Node):** Registo do Modelo Instanciado no Catálogo do Espaço de Dados (esquema físico gravado e associado aos URIs/PIDs ativos do catálogo).
- **Normas e Referenciais de Suporte:**
  - **ISO/IEC 11179-3:2023:** Instanciação de classes e subclasses de `Administered_Item` para persistência formal no repositório.
  - **ISO/IEC 11179-34:** Padrões de reprodutibilidade e gestão de metadados computáveis.
  - **W3C DCAT v3:** Padronização dos metadados de catálogo, distribuição física e identificadores persistentes (PIDs).
  - **W3C PROV-O:** Modelação formal de linhagem e proveniência através das propriedades `prov:wasAttributedTo`, `prov:wasGeneratedBy` e `prov:wasDerivedFrom`.

D.10

Ficha de Atividade 10: Publicar e Governar Modelo

- **Papel:** Fornecedor do Modelo de Dados (em articulação com a Entidade de Governo do Espaço de Dados).
- **Objetivo Geral:** Disponibilizar o modelo de dados em tempo de execução (*Runtime*) no *Vocabulary Hub* e no *Metadata Broker* para validação de esquemas, controlando o versionamento, depreciação, contratos e direitos de uso semânticos.
- **Artefacto de Entrada Global (Nó de Objeto de Entrada):** Registo do Modelo Instanciado no Catálogo do Espaço de Dados.
- **Artefacto de Saída Global (Object Node):** Modelo Publicado versionado (disponível para descoberta em *Runtime* por conectores de dados via *Vocabulary Hub*).
- **Subprocessos e Atividades Internas Mapeados:**

- **Verificar Tipo de Alteração:** Determinar a natureza do impacto da nova publicação (primeiro registo, menor/retrocompatível ou maior/disruptiva).
  Normas de Suporte: ISO/IEC 11179-6:2023 (Procedimentos de *change management* e evolução de esquemas da Cláusula D.4).
- **Atribuir Versão Inicial:** Atribuição do código inicial em caso de novo registo (`v1.0.0`).
  Normas de Suporte: ISO/IEC 11179-3:2023 (Identificação e controlo de versões na Cláusula 5) e propriedades de identificador do DCAT v3.
- **Incrementar Versão Menor:** Incrementar o versionamento semântico secundário em caso de alteração retrocompatível.
  Normas de Suporte: W3C DCAT v3 (Versões lógicas utilizando as propriedades estruturadas `dcat:version` e `dcat:previousVersion`).
- **Incrementar Versão Maior:** Processamento e alteração do versionamento primário em caso de rutura ou descontinuação. Inicia um fluxo paralelo de salvaguarda operacional: Manter Versão Anterior em Coexistência, Definir Prazo de Descontinuação e Notificar Consumidores Existentes.
  Normas de Suporte: ISO/IEC 11179-33:2023 (Classe de substituição reflexiva `data_set_supersession` para mapear de forma explícita modelos descontinuados e respetivos substitutos), ISO/IEC 11179-6:2023 (Processo formal de depreciação do Anexo D) e propriedades `dcat:hasCurrentVersion` do DCAT v3.
- **Registar Alteração no Histórico:** Manter o registo auditável de transições no MDR.
  Normas de Suporte: ISO/IEC 11179-3:2023 (Classe `Registration_State` e atributo `effective_date`).
- **Integrar Modelo no Espaço de Dados:** Publicar logicamente o modelo de dados de referência no *Vocabulary Hub* em *Runtime*.
  Normas de Suporte: ISO/IEC 11179-35:2023 (Publicação técnica e integração de modelos físicos no catálogo).
- **Publicar Documentação de Suporte:** Gerar uma especificação técnica legível por humanos que descreva exaustivamente as classes, propriedades e axiomas da extensão semântica.
  Normas de Suporte: ISO/IEC 11179-1:2023 (Cláusula 8: Relação entre metadados lógicos e interfaces de utilizador legíveis por humanos).

D.11

Ficha de Atividade 11: Encerrar Candidatura e Registar Lições

- **Papel:** Ambas as *Lanes* (Fornecedor de Modelo de Dados e Entidade de Governo do Espaço de Dados).

- **Objetivo Geral:** Concluir formalmente o ciclo administrativo do pedido de registo, arquivando a transação e registando lições técnicas de adequação.
- **Artefacto de Entrada Global (Nó de Objeto de Entrada):**Notificação de Publicação [modelo publicado] OU Comunicação de Rejeição Fundamentada [rejeitado] .
- **Artefacto de Saída Global (Object Node):** Registo de Candidatura Encerrado e Lições Armazenadas.
- **Normas e Referenciais de Suporte:**
  - **ISO/IEC 11179-3:2023:** Cláusula 9 (Classes `Submission_Record` e `Stewardship_Record` para preservação administrativa de responsabilidade de manutenção do catálogo).
  - **ISO/IEC 11179-6:2023:** Fecho dos procedimentos administrativos e arquivo histórico de transações.
  - **Data Act:** Artigo 33.º (Princípios de transparência, rastreabilidade e boa administração pública).