



UNIVERSIDADE ESTADUAL PAULISTA
CAMPUS DE MARÍLIA
Faculdade de Filosofia e Ciências

Programa de Pós-Graduação em Ciência da Informação

Emanuelle Torino

ARQUITETURA DE DADOS NO CONTEXTO DA CIÊNCIA DA INFORMAÇÃO

Marília
2023

Emanuelle Torino

ARQUITETURA DE DADOS NO CONTEXTO DA CIÊNCIA DA INFORMAÇÃO

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Informação como requisito para a obtenção do título de Doutora em Ciência da Informação – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (Unesp), Campus de Marília.

Área de concentração: Informação, Tecnologia e Conhecimento

Linha de pesquisa: Informação e Tecnologia

Orientadora: Silvana Aparecida Borsetti
Gregorio Vidotti

Marília
2023



[4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Esta licença permite remixe, adaptação e criação a partir do trabalho, para fins não comerciais, desde que sejam atribuídos créditos ao(s) autor(es) e que licenciem as novas criações sob termos idênticos.

Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

Revisão de André Luiz Machado

T683a

Torino, Emanuelle

Arquitetura de dados no contexto da Ciência da Informação /
Emanuelle Torino. -- Marília, 2023

331 f. : il.

Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Faculdade de Filosofia e Ciências, Marília

Orientadora: Silvana Aparecida Borsetti Gregorio Vidotti

1. Arquitetura da informação. 2. Ecologia informacional
complexa. 3. Arquitetura de dados. 4. Ecologia complexa de
dados. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da
Faculdade de Filosofia e Ciências, Marília. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

IMPACTO POTENCIAL DA PESQUISA

A presente tese, desenvolvida no âmbito da Informação e Tecnologia, possui potencial de contribuição e impacto social, educacional, tecnológico e econômico, dada a criação de um conceito no contexto da Ciência da Informação. Tal fato reafirma o impacto educacional, ao apresentar a Arquitetura de Dados como um novo campo de pesquisa para a área. No que tange ao impacto tecnológico, há aportes para a estruturação da camada de dados de sistemas de informação e ambientes informacionais digitais – considerando aspectos éticos, legais e técnicos – para que possam ser facilmente localizados, acessados, processados e (re)utilizados em diferentes contextos, incluindo a constituição de infraestruturas de dados acionáveis e que possam ser úteis para processos diários e para a tomada de decisão. Essa estruturação da camada de dados e a possibilidade de acionamento e processamento automáticos podem trazer impactos econômicos, oriundos, por exemplo, do reuso de recursos existentes e da tomada de decisão quanto a investimentos, políticas públicas, indicadores e ações que contribuam com o desenvolvimento socioeconômico. Neste contexto, os Objetivos do Desenvolvimento Sustentável (ODS) podem ser apoiados com o uso de dados de múltiplas fontes, por meio dos quais desafios globais podem ser mapeados, monitorados e solucionados.

POTENTIAL RESEARCH IMPACT

This thesis, developed within the scope of Information and Technology, has potential for contribution and social, educational, technological, and economic impact, due to the creation of a concept in the context of Information Science. This reaffirms the educational impact by presenting Data Architecture as a new research field for the area. Regarding the technological impact, there are contributions for structuring the data layer of information systems and digital informational environments - considering ethical, legal and technical aspects - so that they can be easily located, accessed, processed and (re)used in different contexts. This includes the constitution of actionable data infrastructures that can be useful for daily processes and for decision making. This structuring of the data layer and the possibility of automatic triggering and processing can bring economic impacts arising, for example, from the reuse of existing resources and decision-making regarding investments, public policies, indicators, and actions that contribute to socioeconomic development. In this context, the Sustainable Development Goals (SDGs) can be supported by using data from multiple sources, through which global challenges can be mapped, monitored, and solved.

Emanuelle Torino

ARQUITETURA DE DADOS NO CONTEXTO DA CIÊNCIA DA INFORMAÇÃO

Tese de doutorado apresentada ao Programa de Pós-Graduação em Ciência da Informação como requisito para a obtenção do título de Doutora em Ciência da Informação – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (Unesp), Campus de Marília.

Área de concentração: Informação, Tecnologia e Conhecimento

Linha de Pesquisa: Informação e Tecnologia

Data de Aprovação: 25 de novembro de 2022.

BANCA EXAMINADORA

Profa. Dra. Silvana Aparecida Borsetti Gregorio Vidotti

Programa de Pós-Graduação em Ciência da Informação (PPGCI)
Universidade Estadual Paulista (UNESP) – Presidenta e Orientadora

Prof. Dr. Henry Poncio Cruz de Oliveira

Programa de Pós-Graduação em Ciência da Informação (PPGCI) da
Universidade Federal da Paraíba (UFPB) – Examinador Externo

Prof. Dr. Luis Fernando Sayão

Programa de Pós-Graduação em Ciência da Informação (PPGCI) do
Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)/Universidade
Federal do Rio de Janeiro (UFRJ) – Examinador Externo

Prof. Dr. Caio Saraiva Coneglian

Programa de Pós-Graduação em Ciência da Informação (PPGCI)
Universidade Estadual Paulista (UNESP) – Examinador Interno
Universidade de Marília (UNIMAR)

Prof. Dr. Ricardo César Gonçalves Sant'Ana

Programa de Pós-Graduação em Ciência da Informação (PPGCI) da Universidade
Estadual Paulista (UNESP) – Examinador Interno

AGRADECIMENTOS

Hoje, 27 de julho de 2021, decidi me valer do método quadripolar para inverter a ordenação lógica comumente adotada para o desenvolvimento de uma tese e, com isso, antes de concluí-la, iniciei a redação dos agradecimentos.

Agradeço a Deus, que para mim representa uma força universal, que me permite tantas coisas, como acreditar nos sonhos, fazer planos, a ponto de torná-los realidade, um desses sonhos foi chegar ao doutorado e, se você está lendo os agradecimentos, também me foi permitido concluí-lo.

Dos meus, quero iniciar falando dos que caminharam antes de mim, acreditaram nos seus sonhos e fizeram suas conquistas. Agradeço aos meus avós Joaquim e Antonia, por sempre acreditarem em mim, pelo incentivo, preocupação, carinho e conforto. A partida do Joca me fez buscar nos estudos refúgio. Na fase de redação desta tese, a Tonha foi encontrá-lo e, em meio a um novo vazio, busquei forças para finalizá-la. Espero que vocês tenham sentido minha gratidão e amor enquanto estiveram aqui. Sou muito grata por todos os valores que me deram e tudo o que sou tem um pouco de vocês. “Só enquanto eu respirar Vou me lembrar de você[s]” (Fernando Anitelli).

Aos meus pais, Valdecir e Tereza, cabeça nas estrelas e pé no chão, por me ensinarem que posso voar, desde que saiba onde encontrar chão firme. Pai, você também foi embora durante meu doutorado. Então, espero que daí de cima você sinta orgulho de mim, como sentiu nas conquistas que pôde acompanhar daqui. Mãe, sua força inspira e por mais longe que seja o voo, sei que sempre terei em você apoio e porto seguro.

A tia Jeane que, junto aos meus pais e avós, exerce papel importante na minha formação, pelo cuidado, carinho, preocupação, dedicação, confiança e amor sem medidas.

Meus irmãos, Lígia, Leandro e Janyne. Depois de crescidos, aprendemos que a distância não separa e a união se comprova sempre que necessária! É nos momentos difíceis que entendemos quem caminha ao nosso lado, e nossa família é o retrato perfeito desta fala. Agradeço pela confiança que construímos a cada dia, por acreditarem em mim e por me permitirem compartilhar a vida com vocês e as famílias que construíram. Aqui incluo a Liliane, cunhada que faz parte dessa construção familiar.

Aos meus sobrinhos, Cauã, Duda, Heitor e Laís, que são partes de mim e, por isso, meus. Vocês representam o caminho que foi iniciado antes de nós, percorrido pelos nossos antepassados, por mim e por vocês. Tenho em vocês o brilho da vida que nasce e se renova. Vocês são meu eu do passado e, de alguma forma, devo ser um pouco do que serão no futuro. É muito bom tê-los comigo e, por vocês, construo cada um dos meus dias. Estarei sempre perto.

Agradeço às conexões que temos ao longo da vida pessoal e profissional, nesta última, àquelas que desencadearam uma aproximação com o GPNTI e com a professora Silvana, no mesmo momento em que eu me abria ao desejo de retomar os estudos.

Assim, agradeço à Silvana – em primeiro lugar como pessoa e depois como orientadora e pesquisadora –, pelo carinho com que sempre me recebeu, por me ensinar a ser uma pesquisadora, por acreditar em mim e me auxiliar na trilha da pesquisa que, embora às vezes solitária, pode ser apresentada e discutida, de forma a ter apoio em uma rede de pesquisadores com diferentes formações humanas e profissionais, e, com isso, me possibilitar ser parte de uma “família acadêmica” e compartilhar conhecimentos, não apenas com ela, mas com tantos acadêmicos que ela formou. Assim, destaco meus “irmãos acadêmicos” Sandra, Kamilo, Caio, Ana Maria, Jean, Henry, Fernando, Mariana e Gustavo. Em vocês tenho exemplos de pesquisadores.

Do ponto de vista acadêmico, agradeço ainda a todos os professores do PPGCI Unesp, sobretudo aos da Linha 1, e aos membros do GPNTI pelo compartilhamento de informações e conhecimentos, pelas pesquisas realizadas e tão importantes para a Ciência da Informação, pelos momentos no Laboratório e por toda a interação realizada ao longo dos anos. Como “a pesquisa não para e nem tira férias” (VIDOTTI, 20--?), ainda nos encontraremos em muitas oportunidades.

Agradeço ainda aos professores que participaram da banca avaliadora desta tese, pelo aceite, participação efetiva, cuidado, atenção e carinho com que conduziram o processo, demonstrando que o rigor e a ética são preponderantes no processo de pesquisa, sem deixar de lado o desenvolvimento humano do pesquisador. Meus agradecimentos aos professores e pesquisadores Caio, Henry, Ricardo e Sayão.

Aos amigos que fiz ao longo da vida, que deixaram um pouco de si em mim e me auxiliam na construção do que sou, pelos incontáveis momentos juntos, pelas

boas risadas, pelo colo nos momentos necessários, pela parceria estabelecida e por entenderem que não é apenas o tempo e a presença física constante que demonstram o quanto são importantes na minha história. Certamente, aqui é onde posso errar em citar nomes. Por isso, agradeço, com todo o meu carinho, às pessoas que compartilharam comigo parte da minha e das suas histórias. E, para tentar ser justa, embora não acredite muito que consiga sê-lo, menciono alguns dos amigos que estiveram mais próximos durante meu processo de doutorado, pelo apoio moral, alimentar e ético, por fazerem o possível para entenderem que a vida exige uma pausa para o doutorado e por me mostrarem que o doutorado precisa de pausa para a vida: Márcia, Marcos, Letícia, Elder, Fabiano, Fabiani, Azis, Dineri e Renata.

Finalmente, agradeço a você que está lendo este texto, pelo interesse em conhecer a pesquisa realizada, e espero que encontre nela algumas respostas e muitas perguntas, pois apenas assim permaneceremos fazendo ciência. Agradeço por dedicar tempo à leitura, por refletir com o texto e, caso haja interesse, por continuar as investigações na temática. Estou aberta a discussões acadêmicas e, quem sabe, parcerias para trabalhos futuros.

A todos, minha gratidão!

EPÍGRAFE

“A informação que adquire um novo significado para nós pode ser comparada ao florescimento de uma bela árvore na primavera. Quando começamos a entender uma língua estrangeira, ou a entender o significado de um poema, a conquista de fazer com que a informação anterior adquira um significado novo ou mais profundo - a conquista de obter um *insight* - pode ser um evento tão profundo e tão alegre que nos resta questionar sobre o processo que nos levou a tal estado de graça.

Assim, começamos a nos perguntar sobre o significado do significado. Em nossos pensamentos, passamos a reconhecer quão único é tal ato - um ato que pode despertar em nós uma resposta emocional tão profunda. Intuitivamente, atribuímos a resposta como sendo única para nós mesmos e para outros seres humanos.

E aí está a falácia. Pois o que estamos testemunhando é a extremidade superior de um espectro de fenômenos em que um receptor de informação é capaz de discriminar entre os itens de informação recebida e fundir tal informação com a informação armazenada na própria estrutura de conhecimento do receptor. É um fenômeno que tem uma longa história evolutiva por trás dele.”

(STONIER, 1997, p. 181, tradução nossa).

RESUMO

A Arquitetura da Informação atua no projeto, planejamento e estrutura de ambientes, visando melhorar a forma como a informação é organizada, disponibilizada e apresentada e, com isso, favorecer o acesso, a interação e a experiência de uso. Para tanto, é necessário que haja planejamento da arquitetura do ambiente, utilizando-se do arcabouço teórico-prático da Arquitetura da Informação e de seus entregáveis que favorecem o processo de projetar a estruturação, desenvolvimento e/ou a manutenção dos ambientes informacionais. A anatomia da Arquitetura da Informação é compreendida de forma *top-down*, que considera conteúdo, contexto e usuário para desenvolver o projeto adequado de um ambiente, e *bottom-up*, considerando principalmente o conteúdo e as ferramentas para alavancá-lo. Assim, atua no âmbito das soluções tecnológicas, ao passo que a Arquitetura de Dados atua no projeto da infraestrutura de dados e do Ciclo de Vida dos Dados, visando favorecer o consumo dos dados por máquinas. A presente pesquisa parte das seguintes inquietações: A Arquitetura da Informação é suficiente para estruturar a camada de dados dos ambientes informacionais digitais? A Arquitetura de Dados no contexto da Ciência da Informação pode contribuir com a Arquitetura da Informação *bottom-up*? Como a Arquitetura de Dados se relaciona com a Arquitetura da Informação? E, para respondê-las, objetiva-se investigar e compreender a Arquitetura de Dados a fim de conceituá-la no contexto da Ciência da Informação e relacioná-la com a Arquitetura da Informação. A articulação metodológica da pesquisa utiliza o delineamento quadripolar, que permite estruturar de forma sistemática o desenvolvimento de uma investigação não linear, que possibilita que o seu percurso e a abordagem ao objeto/fenômeno de pesquisa determinem os caminhos a serem seguidos, permitindo alterar, de forma cíclica os diferentes achados da investigação. Como resultados apresenta: Conceito de Arquitetura de Dados no Contexto da Ciência da Informação; Tríade da Arquitetura de Dados; Fluxo de Dados na Arquitetura de Dados; Conceito operacional de Dado; Conceito operacional de Ecologia Complexa de Dados; Reposicionamento da Arquitetura da Informação Pervasiva para a abordagem Informacional; Reposicionamento da Ecologia Informacional Complexa para a Arquitetura da Informação.

Palavras-chave: Arquitetura da Informação. Ecologia Informacional Complexa. Arquitetura de Dados. Ecologia Complexa de Dados. Ciclo de Vida dos Dados. Metadados. Estrutura Sintática de Dados. Estrutura Semântica de Dados. Princípios FAIR. Princípios CARE.

ABSTRACT

Information Architecture works in projecting, planning and structuring environments, aiming to improve the way information is organized, made available and presented, and thus favouring access, interaction and user experience. To this end, it is necessary to plan the architecture of the environment, using Information Architecture's theoretical-practical framework and its deliverables that favour the process of projecting the structuring, development, and/or maintenance of informational environments. Information Architecture's anatomy is understood in a top-down manner, which considers content, context and user in order to develop the appropriate design of an environment, and bottom-up, considering mainly content and tools to leverage it. Thus, it acts in the realm of technological solutions, while Data Architecture acts in the design of the data infrastructure and Data Life Cycle, aiming to favour data consumption by machines. This research is based on the following questions: Is Information Architecture sufficient to structure the data layer of digital informational environments? Can Data Architecture, in the context of Information Science, contribute to bottom-up Information Architecture? How does Data Architecture relate to Information Architecture? And, to answer them, we aim to investigate and understand Data Architecture in order to conceptualize it in the context of Information Science and relate it to Information Architecture. The research methodological articulation uses the quadripolar design, which allows the development of a non-linear investigation to be structured in a systematic way, allowing the research path and approach towards the object/phenomenon to determine the course of action to be followed, allowing cyclical alteration of the different research findings. As results, it presents: Concept of Data Architecture in the Context of Information Science; Triad of Data Architecture; Data Flow in Data Architecture; Operational Concept of Data; Operational Concept of Complex Data Ecology; Repositioning Pervasive Information Architecture towards Informational approach; Repositioning Complex Information Ecology towards Information Architecture.

Keywords: Information Architecture. Complex Informational Ecology. Data Architecture. Complex Data Ecology. Data Life Cycle. Metadata. Syntactic Data Structure. Semantic Data Structure. FAIR Principles. CARE Principles.

LISTA DE FIGURAS

Figura 1 – Delineamento quadripolar da pesquisa	27
Figura 2 – Abordagens da Arquitetura da Informação.....	46
Figura 3 – Os três círculos da Arquitetura da Informação	55
Figura 4 – Sistema de Busca	87
Figura 5 – Diagrama conceitual de como os funcionários veem a Ecologia da Informação da empresa	126
Figura 6 – Iceberg da Arquitetura da Informação.....	132
Figura 7 – Mapa de conceitos de informação	163
Figura 8 – Taxonomia DRM de alto nível.....	180
Figura 9 – Áreas de padronização FEA DRM	181
Figura 10 – <i>DRM Abstract Model</i>	182
Figura 11 – Modelo para construção de conceito.....	188
Figura 12 – O triângulo conceitual	189
Figura 13 – Construção do Conceito de Arquitetura de Dados no contexto da Ciência da Informação	192
Figura 14 – Principais atividades do Ciclo de Vida dos Dados.....	194
Figura 15 – Um modelo do ciclo de aquisição enriquecido.....	196
Figura 16 – Modelo do ciclo de uso enriquecido	197
Figura 17 – O ciclo combinado de aquisição e uso	197
Figura 18 – Oito fases do modelo de Ciclo de Vida dos Dados.....	198
Figura 19 – Ciclo de Vida dos Dados: 1 e 2	200
Figura 20 – Ciclo de Vida dos Dados para Ciência da Informação – (CVD–CI)	201
Figura 21 – Ciclo de Vida dos Dados para Ciência da Informação – (CVD–CI)	203
Figura 22 – Modelo fundamental e informal de um Elemento de Dados	216
Figura 23 – Metadados	217
Figura 24 – <i>High-level Data Description metamodel</i>	219
Figura 25 – Princípios CARE para Governança de Dados Indígenas	246

Figura 26 – Desafios enfrentados ao fornecer e consumir dados na web	268
Figura 27 – Publicação de dados na Web.....	269
Figura 28 – Princípios FAIR.....	274
Figura 29 – Elementos de objeto de dados.....	278
Figura 30 – Tríade Vitruviana.....	283
Figura 31 – Tríade da Arquitetura de Dados	284
Figura 32 – Fluxo de Dados na Arquitetura de Dados.....	286

LISTA DE QUADROS

Quadro 1 – Levantamento Bibliográfico de Literatura	32
Quadro 2 – Comparativo dos Elementos da Arquitetura da Informação Pervasiva e da Arquitetura da Informação	114
Quadro 3 – Arquitetura da Informação Clássica e Arquitetura da Informação Moderna .	131
Quadro 4 – Conceitos e características dos dados	152
Quadro 5 – Conceitos e características das informações	153
Quadro 6 – Quadro comparativo entre as características de dado e informação	154
Quadro 7 – Dados e Informação	156
Quadro 8 – Quatro aspectos da Informação	160
Quadro 9 – Princípios e Subprincípios CARE	248
Quadro 10 – Relacionamento entre os Atores envolvidos no Tratamento dos Dados	253
Quadro 11 – Benefícios das Boas Práticas de Dados na Web.....	269
Quadro 12 – Facetas e Subfacetas dos Princípios FAIR	276
Quadro 13 – Relações entre Arquitetura da Informação e Arquitetura de Dados	308

LISTA DE ABREVIATURAS E SIGLAS

AD	Arquitetura de Dados
AI	Arquitetura da Informação
AIP	Arquitetura da Informação Pervasiva
ASEO	<i>Academic Search Engine Optimization</i>
BDTD	Biblioteca Digital Brasileira de Dissertações e Teses
Brapci	Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação
CARE	<i>Collective Benefit, Authority to Control, Responsibility, Ethics</i>
CI	Ciência da Informação
CEP	Comitê de Ética em Pesquisa
CNS	Conselho Nacional de Saúde
COI	<i>Communities of Interest</i>
CONEP	Comissão Nacional de Ética em Pesquisa
CRM	<i>Consolidated Reference Model</i>
CVD	Ciclo de Vida dos Dados
DC	<i>Dublin Core</i>
DIKW	<i>Data-Information Knowledge-Wisdom</i>
DoDA	<i>Department of Defense Architecture Framework</i>
DRM	<i>Data Reference Model</i>
EA	<i>Enterprise Architecture</i>
FAIR	<i>Findable, Accessible, Interoperable, Reusable</i>
FEA	<i>Federal Enterprise Architecture</i>
FEAF	<i>Federal Enterprise Architecture Framework</i>
GDI	<i>General Definition of Information</i>
GPNTI	Grupo de Pesquisa Novas Tecnologias em Informação
IA	Inteligência Artificial
IBM	<i>International Business Machines Corporation</i>
IDEF	<i>Integration Definition for Function Modeling</i>
INDA	Infraestrutura Nacional de Dados Abertos
LGPD	Lei Geral de Proteção de Dados Pessoais
LOD	<i>Linked Data</i> ou <i>Linked Open Data</i>
MARC	<i>Machine Readable Cataloging</i>
MDR	<i>Metadata Registries</i>
METS	<i>Metadata Encoding and Transmission Standard</i>
MODS	<i>Metadata Object Description Schema</i>
NIEM	<i>National Information Exchange Model</i>

Oasisbr	Portal Brasileiro de Publicações Científicas em Acesso Aberto
ORCID iD	<i>Open Researcher and Contributor ID</i>
PARC	<i>Xerox Palo Alto Research Center</i>
PDDL	<i>Open Data Commons Public Domain Dedication and License</i>
PID	<i>Persistent Identifier</i>
PPGCI	Programa de Pós-Graduação em Ciência da Informação
PREMIS	<i>PREservation Metadata: Implementation Strategies</i>
RDF	<i>Resource Description Framework</i>
RIUT	Repositório Institucional da Universidade Tecnológica Federal do Paraná
UNESP	Universidade Estadual Paulista Júlio de Mesquita Filho
TIC	Tecnologias de Informação e Comunicação
TOGA	<i>The Open Group Architecture Framework</i>
SEO	<i>Search Engine Optimization</i>
W3C	<i>World Wide Web Consortium</i>
XML	<i>Extensible Markup Language</i>

SUMÁRIO

1 INTRODUÇÃO	16
2 DELINEAMENTO METODOLÓGICO.....	25
2.1 POLO EPISTEMOLÓGICO	27
2.2 POLO TEÓRICO	29
2.3 POLO TÉCNICO	31
2.4 POLO MORFOLÓGICO	34
2.4.1 Produção Científica relacionada à Tese.....	36
3 ARQUITETURA DA INFORMAÇÃO	40
3.1 ARQUITETURA DA INFORMAÇÃO: ABORDAGEM ARQUITETURAL.....	56
3.2 ARQUITETURA DA INFORMAÇÃO: ABORDAGEM SISTÊMICA	62
3.2.1 Sistema de Organização.....	63
3.2.2 Sistema de Rotulagem.....	70
3.2.3 Sistema de Navegação.....	75
3.2.4 Sistema de Busca	82
3.2.5 Tesouros, Vocabulários Controlados e Metadados	93
3.3 ARQUITETURA DA INFORMAÇÃO: ABORDAGEM INFORMACIONAL.....	98
3.4 ARQUITETURA DA INFORMAÇÃO: ABORDAGEM PERVASIVA	103
3.4.1 Arquitetura da Informação: Reposicionamento da Abordagem Pervasiva para a Informacional	110
3.5 ECOLOGIA INFORMACIONAL COMPLEXA	124
3.6 ARQUITETURA DA INFORMAÇÃO: REFLEXÕES.....	130
4 ARQUITETURA DE DADOS	135
4.1 DADO E INFORMAÇÃO: TESSITURAS TEÓRICAS	141
4.2 ARQUITETURA DE DADOS: ANTECEDENTES.....	166
4.3 ARQUITETURA DE DADOS: APORTE PARA A CONSTRUÇÃO DO CONCEITO .	186
4.4 CONCEITO DE ARQUITETURA DE DADOS NO CONTEXTO DA CIÊNCIA DA INFORMAÇÃO	190
4.4.1 Elementos da Arquitetura de Dados no contexto da Ciência da Informação	193

4.4.1.1 <i>Ciclo de Vida dos Dados</i>	193
4.4.1.3 <i>Metadados</i>	208
4.4.1.4 <i>Estruturas Sintática e Semântica dos Dados</i>	227
4.4.1.5 <i>Princípios Éticos, Legais e Técnicos aplicados aos Dados</i>	237
4.4.1.5.1 <i>Princípios éticos aplicados aos dados</i>	239
4.4.1.5.2 <i>Princípios legais aplicados aos dados</i>	251
4.4.1.5.3 <i>Princípios técnicos aplicados aos dados</i>	266
4.4.2 <i>Arquitetura de Dados no contexto da Ciência da Informação: Tríade e Fluxo</i> ..	283
4.4.3 <i>Entregáveis da Arquitetura de Dados</i>	290
4.4.4 <i>Arquiteto de Dados</i>	291
4.5 ECOLOGIA COMPLEXA DE DADOS	294
4.6 ARQUITETURA DE DADOS: REFLEXÕES	299
5 ARQUITETURA DA INFORMAÇÃO E ARQUITETURA DE DADOS: RELAÇÕES POSSÍVEIS	304
6 CONCLUSÕES	310
REFERÊNCIAS	318

1 INTRODUÇÃO

As Tecnologias de Informação e Comunicação (TIC) trouxeram consigo avanços inegáveis para diferentes setores e atividades, que se apoiaram no advento da informática, da internet e da *World Wide Web* para desenvolver e disponibilizar sistemas de informação e ambientes informacionais digitais.

Os sistemas de informação são responsáveis pelo armazenamento e gestão de dados de diferentes domínios, com potencial de uso para finalidades diversas, mas que estão sempre vinculados a conteúdos, contextos e usuários.

Esses avanços suscitaram em pesquisadores como Wurman (1996) e Rosenfeld e Morville (1998) o interesse em aproximar desses ambientes a arquitetura, o design, a ergonomia, a usabilidade, a interação humano-computador e outros aspectos relacionados, por meio da Arquitetura da Informação (AI).

A Arquitetura da Informação atua no projeto, planejamento e estrutura dos ambientes, visando melhorar a forma como a informação é organizada, disponibilizada e apresentada e, assim, favorecer o acesso, a interação e a experiência de uso. Para tanto, é necessário que haja um planejamento da arquitetura do ambiente, utilizando-se do arcabouço teórico-prático da Arquitetura da Informação e de seus entregáveis que favorecem o processo de projetar a estruturação, o desenvolvimento e/ou a manutenção dos ambientes informacionais.

Destaca-se que a Arquitetura da Informação não se ocupa da estética, mas do projeto de organização do espaço, visando permitir que o ambiente possa ser habitado e o usuário possa localizar a informação que necessita, atentando-se à usabilidade, acessibilidade e experiência de uso.

Para Rosenfeld, Morville e Arango (2015) a crescente disponibilidade de informações acessíveis por diversos aparatos tecnológicos possui inúmeros benefícios; contudo, pode se tornar um desafio para que as informações possam ser localizadas e compreendidas pelos usuários.

Al nos permite pensar sobre esses problemas por meio de duas perspectivas importantes: que os produtos e serviços de informação são percebidos pelas pessoas como *lugares feitos de informação*, e que esses ambientes de informação podem ser *organizados para*

*facilitar a localização e compreensão.*¹ (ROSENFELD, MORVILLE; ARANGO, 2015, n. p., tradução nossa).

Para Oliveira (2014, p. 19) “A Arquitetura da Informação tem sido usada como um vetor para o melhoramento do acesso e do uso da informação em ambientes analógicos ou digitais desde a década de 1980.” O autor apresenta diferentes abordagens disciplinares para a AI, desde sua concepção, com base arquitetural, apoiada pela arquitetura e design; passando pela teoria dos sistemas, na abordagem sistêmica; pela abordagem informacional que se relaciona com a Ciência da Informação; até a abordagem pervasiva, com aspectos de ubiquidade e design de serviços.

Com base nos estudos de Everden e Everden (2003), Oliveira (2014, p. 95) explicita como chave da abordagem sistêmica “[...] a AI como mecanismo de estruturação de ‘interfaces e camadas’ (partes) para integrar ‘o todo’ de forma holística.”

Elucida-se, assim, que a AI atua nas diferentes camadas dos ambientes informacionais analógicos e/ou digitais e nos sistemas de informação, considerando as quatro abordagens anteriormente mencionadas. Em todas as abordagens, há o que Rosenfeld e Morville (1998), Morville e Rosenfeld (2006) e Rosenfeld, Morville e Arango (2015) chamam de princípios básicos de AI, constituídos por sua anatomia e componentes, também chamados de sistemas.

A anatomia da AI é compreendida de forma *top-down* e *bottom-up*. A AI *top-down* considera conteúdo, contexto e usuário para desenvolver o projeto adequado de um ambiente, ou seja, atua no planejamento e na estruturação da camada de interface, visando o uso da informação. Enquanto a AI *bottom-up* considera principalmente o conteúdo e as ferramentas para alavancá-lo, de forma que atua no âmbito das soluções tecnológicas (HAGEDORN, 2000).

É necessário entender que o avanço das tecnologias e dos algoritmos de processamento utilizados pelos sistemas de Recuperação da Informação conduzem o usuário à realização de buscas por meio de ferramentas de busca ou sistemas de descoberta e entrega, que agregam conteúdos de uma ampla gama de ambientes informacionais. Com isso, é a Arquitetura da Informação *bottom-up* que auxilia na

¹ “IA allows us to think about problems through two important perspectives: that information products and services are perceived by people as places made of information, and that these information environments can be organized for optimum findability and understandability.”

localização da informação de interesse do usuário por meio da ferramenta de busca que, após acionada pelo usuário, vai direcioná-lo para o ambiente informacional de origem do conteúdo recuperado.

Dessa forma, a Arquitetura da Informação *top-down* fica em segundo plano, podendo ou não ser utilizada pelo usuário após o contato com a informação, momento em que toma a decisão de manter-se no ambiente e utilizar a navegação e/ou o sistema de busca, ou ainda retornar à ferramenta de busca para acessar outros ambientes informacionais a partir do resultado de busca primário. Nesse contexto, é possível ainda que o usuário tenha acesso, por meio de ferramentas de busca ou sistemas de descoberta e entrega, a catálogos de ambientes informacionais analógicos, cujo acesso à informação é possível apenas de forma física.

Com isso, a relevância da Arquitetura da Informação *bottom-up* é reforçada, no que tange à necessidade de manter a informação estruturada, para suportar a busca e a navegação, de forma que os conteúdos possam ser apresentados na camada de interface do ambiente informacional.

Porém, Resmini e Rosati (2011, p. 26, tradução nossa) entendem que não faz parte da abordagem sistêmica unir ao projeto de AI questões de “[...] conexão de dados, largura de banda, custos, tipologia de servidor, e limites de armazenamento que, normalmente não estão nas preocupações da AI, pois tendem a ser bastante específicos e de natureza tecnológica.²”

Para Oliveira (2014, p. 96), “[...] a abordagem sistêmica influencia a AI, sobretudo em aplicações no campo dos Sistemas de informação, porém incorpora preocupações para além das necessidades de informação dos sujeitos e das organizações, incorpora preocupações estritamente técnicas.”

Diante do exposto, é possível compreender que a Arquitetura da Informação não tem como objetivo o processo de armazenamento e tratamento dos dados enquanto atributo codificado³, atividade imprescindível para o adequado funcionamento do ambiente informacional e/ou sistema de informação, considerando os sistemas de Rosenfeld e Morville (1998), Morville e Rosenfeld (2006) e Rosenfeld,

² “[...] questions of data connections, bandwidth, costs, server topology, and storage limits that are not normally part of the mindset of IAs and that tend to be rather specific and technological in nature.”

³ Considera-se que os metadados são constituídos por atributo codificado e campo de valor. Assim, atributos são “Aspectos da informação sobre um objeto de conteúdo. Os atributos podem ser campos, tags e meta-tags.” (HAGEDORN, 2000, p. 2, tradução nossa). E o campo de valor consistem em “Dados descritivos sobre o objeto de conteúdo. Os valores dos atributos podem ser controlados (de acordo com um vocabulário) ou não controlados. Os valores dos atributos podem ser metainformação.” (HAGEDORN, 2000, p. 2, tradução nossa).

Morville e Arango (2015). Embora atue no tratamento dos dados enquanto campo de valor³, considerando os tesouros, vocabulários controlados e metadados presentes em Morville e Rosenfeld (2006).

Morville e Rosenfeld (2006) apontam que há componentes invisíveis na Arquitetura da Informação, nomeadamente: tesouros, vocabulários controlados e metadados – que, nesta tese são compreendidos como sistema de representação da informação –; algoritmos, que atuam sobretudo no sistema de recuperação do ambiente informacional, mas sem relegar a importância que os algoritmos das ferramentas de busca exercem no processo de recuperação da informação; e “melhores apostas”, que consistem na forma de agrupar e ranquear os resultados de busca.

Assim, ainda que considerando a presença de aspectos de tratamento dos dados na *AI bottom-up*, é perceptível que os projetos de *AI* não atuam diretamente na camada de dados dos sistemas de informação e/ou ambientes informacionais digitais, embora a tenham como insumo para a camada de interface.

Considerando essa lacuna, recorre-se a Capurro (2003), para quem a Ciência da Informação tem duas raízes, a Biblioteconomia clássica e a Computação Digital. O que possibilita afirmar que, muito embora a área tenha como objeto a informação, sua característica indissociável dos dados, utilizados em diferentes processos – dentre os quais se destacam a coleta, o armazenamento, a representação, a recuperação, o acesso, o (re)uso e a preservação –, é indubitavelmente possível que a Ciência da Informação seja área fronteira ao estudo dos dados.

Justifica-se, ainda, o estudo na Ciência da Informação, uma vez que a própria *AI*, embora seja uma área interdisciplinar, possui em sua estrutura muitos aspectos presentes na Ciência da Informação, sobretudo os relacionados à organização e representação da informação e às TIC.

Desta forma, considerando que a Arquitetura da Informação atua principalmente na interface dos ambientes informacionais e sistemas de informação, visando, por meio dela, nortear o planejamento, a organização e o uso da informação em determinados contextos, é necessário pensar, a partir dos conceitos de *AI*, como é possível atuar em outras camadas dos sistemas de informação e ambientes informacionais digitais, notadamente na camada de dados, com aportes da Ciência da Informação.

Merece destaque, ainda, a abordagem pervasiva da AI, um campo de estudo recente, que se apresenta a partir da necessidade de aproximar a AI dos avanços tecnológicos que possibilitam que os sujeitos façam interações com a informação disponível em múltiplos canais e ambientes informacionais. Para tanto, considera o hibridismo de espaços de vivência humana e os múltiplos dispositivos que permitem acessar a informação, seja ela analógica ou digital.

Além disso, Oliveira (2014, p. 134) inclui na conceituação da Arquitetura da Informação Pervasiva a presença de Ecologias Informacionais Complexas, referindo-se “[...] ao conjunto de espaços, ambientes, canais, mídias, tecnologias e sujeitos com seus comportamentos, todos interligados e conectados de maneira holística pela informação.” Assevera o autor que a Ecologia Informacional Complexa consiste em um todo constituído pelas partes e organizado de forma sistêmica.

Contudo, tais Ecologias Informacionais Complexas podem se constituir a partir do (re)uso de dados disponíveis na camada de dados dos sistemas, vislumbrando-se, ainda, que os dados sejam provenientes de diferentes sistemas de informação, disponíveis em ambientes internos ou externos, o que reforça a relevância do presente estudo.

Nesse sentido, é necessário que haja comunicação entre os diferentes sistemas de informação e ambientes informacionais digitais, o que requer que os dados por eles armazenados sejam compreensíveis por humanos e máquinas, cujo processo depende da adequada estruturação para que sejam localizáveis, acessíveis, processáveis e interoperáveis.

No que tange à camada de dados de um sistema de informação ou de um ambiente informacional digital, é necessário que, de forma análoga ao que se faz na Arquitetura da Informação, haja uma arquitetura, denominada Arquitetura de Dados (AD). A Arquitetura de Dados no contexto da Ciência da Informação, conceituada nesta tese, estabelece a estrutura e o fluxo de dados, considerando as formas de localização, coleta/consumo, representação, armazenamento, organização, processamento, acesso e (re)uso.

Assim, o cenário atual de crescente disponibilidade de dados e informações, utilizados em diferentes contextos, sistemas de informação e ambientes informacionais que podem se constituir em Ecologias Informacionais Complexas, apresenta algumas inquietações que culminaram nas questões de pesquisa desta tese:

- a) A Arquitetura da Informação é suficiente para estruturar a camada de dados dos ambientes informacionais digitais?
- b) A Arquitetura de Dados no contexto da Ciência da Informação pode contribuir com a estruturação da camada de dados dos ambientes informacionais para que sejam consumidos na Arquitetura da Informação *bottom-up*?
- c) Como a Arquitetura de Dados se relaciona com a Arquitetura da Informação?

Diante das questões de pesquisa, tem-se como hipótese que a Arquitetura da Informação é insuficiente para a adequada estruturação da camada dos dados dos ambientes informacionais digitais, o que requer a aproximação com a Arquitetura de Dados.

Para responder a esses questionamentos, o objetivo geral desta tese consiste em investigar e compreender a Arquitetura de Dados, a fim de conceituá-la no contexto da Ciência da Informação e relacioná-la com a Arquitetura da Informação.

E, com o intuito de atingi-lo, são estabelecidos como objetivos específicos:

- a) discutir Arquitetura da Informação e Ecologia Informacional Complexa no contexto da Ciência da Informação;
- b) investigar e compreender a Arquitetura de Dados e a Ecologia Complexa de Dados;
- c) estabelecer os elementos conceituais essenciais à Arquitetura de Dados no contexto da Ciência da Informação;
- d) investigar as relações entre Arquitetura da Informação e Arquitetura de Dados no contexto da Ciência da Informação.

O presente estudo no âmbito da linha de pesquisa Informação e Tecnologia do Programa de Pós-Graduação em Ciência da Informação (PPGCI) da Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), vinculado ao Grupo de Pesquisa Novas Tecnologias em Informação (GPNTI), situa o desenvolvimento do conceito de Arquitetura de Dados no contexto da Ciência da Informação no Brasil e possibilita iniciar discussões contemporâneas acerca da temática na área.

De igual maneira, a atuação profissional da pesquisadora, na qual o tratamento e a integração de dados têm possibilitado avanços práticos no fazer enquanto profissional da informação (cujo desenvolvimento se faz de forma empírica), pode trazer à análise científica problemáticas cotidianas e tratá-las sob o prisma da

investigação científica, potencializando achados que podem propiciar avanços à área de Ciência da Informação, inclusive reposicionando outras disciplinas fronteiriças.

Além disso, a investigação possui viés pragmático, uma vez que a Arquitetura de Dados e a Ecologia Complexa de Dados podem ser materializáveis, com potencial aplicação em diferentes ambientes e contextos.

Neste sentido, há a possibilidade de aproximar a Ciência da Informação de áreas correlatas ao contribuir com elementos de estruturação da camada de dados dos sistemas de informação e/ou ambientes informacionais digitais, com benefícios para diferentes instituições nas quais a aplicação prática dos resultados desta pesquisa podem contribuir para a melhoria dos sistemas, por meio da Arquitetura de Dados e da constituição de Ecologias Complexas de Dados.

Como exemplo, no contexto das instituições de ensino superior, isso se materializa a partir da estruturação de dados para uso e reuso, visando criar ecologias de pesquisa, compreendendo a conexão de dados internos e externos, como Currículo Lattes, identificadores persistentes de pessoas – como o *Open Researcher and Contributor ID* (ORCID iD) –, diretórios de grupos de pesquisa, resultados de pesquisas publicados em fontes externas – a exemplo de periódicos e indexadores como *Scopus*, *Web of Science* e *SciELO* –, patentes e softwares, plataformas como a Sucupira e, a vinculação destes com dados internos, como os de docentes e estudantes.

Isso corrobora com a afirmação de Santos e Vidotti (2009, p. 4):

A Ciência da Informação deveria ter ou criar mais espaços de investigação que permitam a compreensão das Tecnologias de Informação e Comunicação (TIC) para a potencialização de competências informacionais, para a criação de arquiteturas informacionais e computacionais mais inclusivas, para a conceituação de usos da informação em ambientes informacionais digitais, para a aprendizagem de metalinguagens e para a representação da informação.

E com Sant’Ana (2016, p. 138):

A CI pode, portanto, complementar seu papel na proposta de novos caminhos para que os usuários, de forma mais democrática e aberta possam acessar e utilizar dados, reduzindo a assimetria informacional que pode surgir entre os que os detêm e os que dele precisam.

Essas afirmações reforçam a necessidade e a relevância do estudo da Arquitetura de Dados no contexto da Ciência da Informação (CI), a fim de discuti-la, conceituá-la e relacioná-la à Arquitetura da Informação.

No contexto da Arquitetura da Informação, há diferentes estudos relacionados à camada da interface dos ambientes informacionais. Porém a estruturação da camada de dados não está claramente elucidada, o que reforça a necessidade da proposição e aproximação com a Arquitetura de Dados.

A pesquisa se justifica, ainda, na afirmação de Borko (1968, p. 2): a “Ciência da Informação como uma disciplina tem como meta fornecer um corpus teórico sobre informação que propiciará a melhoria de várias instituições e procedimentos dedicados à acumulação e transmissão de conhecimento”. Postula o autor que a área de Ciência da Informação possui amplo campo de pesquisa, sistematizada em nove categorias, sendo que a presente pesquisa se enquadra em três delas, ao abordar: (1) padrões de comunicação, (3) análise da linguagem e linguística computacional, (6) arquitetura de sistemas.

A presente tese está estruturada em seis seções:

- a) Introdução: apresenta a pesquisa e posiciona os problemas de pesquisa, a hipótese, o objetivos geral e os específicos, a justificativa e a motivação;
 - b) Delineamento Metodológico: posiciona a estrutura da pesquisa por meio do delineamento quadripolar, a partir do qual apresenta os elementos que compõem os polos epistemológico, teórico, técnico e morfológico;
 - c) Arquitetura da Informação: apresenta e discute as abordagens arquitetural, sistêmica, informacional e pervasiva, bem como a Ecologia Informacional Complexa no âmbito da Ciência da Informação. Propõe o reposicionamento da abordagem Pervasiva da Arquitetura da Informação para a abordagem Informacional, bem como o reposicionamento da Ecologia Informacional Complexa, da Arquitetura da Informação Pervasiva para a Arquitetura da Informação;
 - d) Arquitetura de Dados: apresenta e discute Arquitetura de Dados em diferentes domínios e conceitua Arquitetura de Dados no contexto da Ciência da Informação, com base na Teoria do Conceito, além de discutir Ecologia Complexa de Dados;
 - e) Arquitetura da Informação e Arquitetura de Dados: relações possíveis: elabora o relacionamento entre Arquitetura da Informação e Arquitetura de Dados, com base nos estudos realizados ao longo da tese;
-

- f) **Conclusões:** apresenta os resultados da pesquisa, bem como os achados da investigação, elaborando-os de modo a responder às perguntas de pesquisa, à hipótese, e ao objetivos geral e aos específicos.
-

2 DELINEAMENTO METODOLÓGICO

A articulação metodológica da pesquisa utiliza método quadripolar, cuja adoção pela Ciência da Informação foi sugerida por Silva e Ribeiro em 2002 “para alicerçar a investigação global [...] como alternativa à tradicional investigação linear/sequencial” (VECHIATO, 2013, p. 24).

O método quadripolar foi concebido por De Bruyne, Herman e De Schoutheete (1977), cuja obra esclarece que a metodologia não deve ser reduzida a uma abordagem ou ao uso de técnicas de coleta. Ao contrário, consiste em uma visão global do processo de pesquisa com uma dinâmica interna própria. Desta forma, a pesquisa não se constrói em momentos distintos, mas sim na articulação da abordagem, dos procedimentos e das técnicas utilizadas ao longo da investigação.

Na obra supracitada, o prefaciador Jean Ladrière afirma que “[...] a única maneira de construir uma verdadeira ciência dos fatos sociais é adotar o caminho da análise sistemática [...]” (DE BRUYNE; HERMAN; DE SCHOUTHEETE, 1977, p. 11), e que a ciência se desenvolve de forma dupla, em conteúdos e critérios que se utilizam de métodos próprios que possibilitam estabelecer o eixo de evolução de forma progressiva, oriundo da realização da pesquisa e da abordagem aos objetos/fenômenos estudados. Ladrière afirma ainda que “Foi a natureza dos objetos que de algum modo impôs o tipo de progressão e, portanto, os métodos” (p. 14).

Para De Bruyne, Herman e De Schoutheete (1977, p. 27) o objetivo da metodologia científica é “[...] esclarecer a unidade subjacente a uma multiplicidade de procedimentos científicos particulares, ela ajuda a desimpedir os caminhos da prática concreta da pesquisa dos obstáculos que esta encontra.” Assim, os autores entendem que a progressão de cada pesquisa é variável e que uma reflexão global possibilita distinguir vias metodológicas que apresentam procedimentos colocados à disposição do pesquisador, sugerindo uma metodologia geral cujo procedimento é dado por metodologias específicas.

A obra de De Bruyne, Herman e De Schoutheete (1977) estabelece o método quadripolar. Contudo, destacamos que o método quadripolar não se restringe ou finaliza em um método de pesquisa composto por etapas sequenciais e predefinidas. Ao contrário, a reflexão acerca da dinâmica da pesquisa constitui o método quadripolar como um delineamento de pesquisa.

O delineamento quadripolar foi concebido como um espaço formado por quatro polos, que estabelecem as diretrizes que orientam o processo de pesquisa científica sem reduzi-lo a um conjunto de procedimentos e/ou etapas sequenciais e lineares.

Assim, o delineamento quadripolar permite estruturar de forma sistemática o desenvolvimento de uma investigação não linear, que possibilita que o seu percurso e a abordagem ao objeto/fenômeno de pesquisa determinem os caminhos a serem seguidos, permitindo alterar, de forma cíclica, os diferentes achados da investigação. Desta forma, a abordagem ao objeto/fenômeno pode requerer determinado(s) método(s), e ser progressivamente elaborada por ele(s), de forma dinâmica e contínua, sem que se esgotem em si mesmos.

Para Oliveira (2014, p. 27),

O método quadripolar transcende as fronteiras de definição de um objeto ou um fenômeno científico, sua plasticidade permite que ele se ajuste às características dos objetos ou fenômenos de pesquisa, permitindo que o método seja aplicado em diversos objetos ou fenômenos, recorrendo a diversos aparatos teóricos, múltiplas epistemologias, várias técnicas e produza diferentes morfologias na pesquisa.

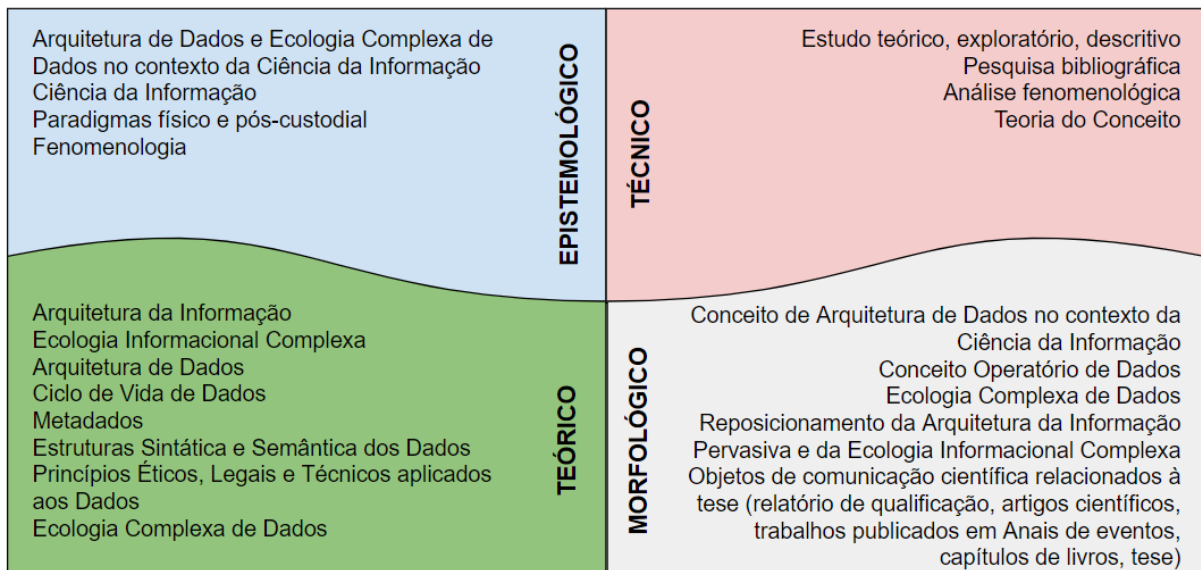
O método quadripolar, entendido nesta tese como delineamento quadripolar, foi concebido como a “[...] articulação de diferentes instâncias, de diferentes pólos que determinam um espaço no qual a pesquisa se apresenta como apanhada num campo de forças, submetida a determinados fluxos, a determinadas exigências internas.” (DE BRUYNE; HERMAN; DE SCHOUTHEETE, 1977, p. 34). Os polos epistemológico, teórico, técnico e morfológico, segundo os autores, não consistem em momentos separados da pesquisa, são aspectos particulares da sua constituição que se relacionam e condicionam pela interconexão.

O polo epistemológico consiste na reflexão da pesquisa sobre si mesma, de forma que alicerça a vigilância crítica ao objeto/fenômeno, explicita a problemática de pesquisa, assume o paradigma que orienta a investigação e determina a abordagem ao objeto, fatores que asseguram o distanciamento da pesquisa científica do senso comum. O polo teórico norteia a elaboração das hipóteses e a conceitualização que sustenta o estudo e permite interpretar fatos que auxiliam na resolução da problemática trazida pela pesquisa, por meio das contribuições teóricas do quadro de referência utilizado. O polo técnico explicita os métodos e as técnicas adotadas para a realização da pesquisa. E o polo morfológico operacionaliza o polo teórico,

apresenta o processo percorrido ao largo da investigação e estrutura a formalização dos resultados, materializados nos objetos de produção técnico-científica relacionados à pesquisa, a exemplo de publicações científicas e modelos conceituais (DE BRUYNE; HERMAN; DE SCHOUTHEETE, 1977).

O delineamento quadripolar da presente pesquisa está representado na Figura 1, que explicita seu processo.

Figura 1 – Delineamento quadripolar da pesquisa



Fonte: Autoria própria.

A Figura 1 representa os caminhos trilhados na condução da pesquisa realizada para a elaboração desta tese e reforça o entendimento do método quadripolar como um delineamento de pesquisa que permite, de forma integrada, conduzir e sustentar as decisões do pesquisador ao longo da investigação.

2.1 POLO EPISTEMOLÓGICO

O polo epistemológico, enquanto espaço norteador e estruturante da investigação, atua no modo de produção de conhecimento e nos procedimentos de validação da pesquisa. Por isso, é tido como motor interno e obrigatório para De Bruyne, Herman e De Schoutheete (1977). Para os autores, a metodologia “[...] deve apreender a ciência como um processo (não sequencial, mas vivo) e não como um produto.” (p. 43). Assim, afirmam que o olhar epistemológico para a investigação oferece ao pesquisador fundamentos imprescindíveis ao rigor, à exatidão, à precisão

do procedimento e possibilita progredir na abordagem e conhecimento do objeto/fenômeno investigado.

Neste polo, ocorre a constante vigilância ao objeto/fenômeno científico, definição da problemática de investigação e dos parâmetros discursivos, paradigmas e critérios de cientificidade (SILVA, 2006).

A presente tese tem como hipótese que a Arquitetura da Informação é insuficiente para a adequada estruturação da camada dos dados dos ambientes informacionais digitais, o que requer a aproximação com a Arquitetura de Dados.

Neste sentido, tem como fenômenos de pesquisa a Arquitetura de Dados e a Ecologia Complexa de Dados no contexto da Ciência da Informação, visando apresentar contribuições para a estruturação da camada de dados dos ambientes informacionais digitais e dos sistemas de informação, bem como para a Arquitetura da Informação *bottom-up*. Para tanto, a Arquitetura de Dados é conceituada no contexto da Ciência da Informação.

Tem-se como abordagem a fenomenologia, tida por Husserl (2008, p. 20) como “[...] doutrina universal das essências, em que se integra a ciência da essência do conhecimento.”, que deve atuar livre de pressupostos para tratar os “[...] fenômenos, enfocando exclusivamente a eles, deixando de lado quaisquer questões sobre suas origens causais e sua natureza fora do próprio ato de consciência.” (MOREIRA, 2002, p. 62-63).

Martins e Theóphilo (2016, p. 42) destacam que o estudo do fenômeno é “[...] entendido como aquilo que se mostra ou se revela por si mesmo.” Para os autores, todo fenômeno possui essências, que consistem nas características essenciais do fenômeno, que não se reduz a nenhuma das suas características.

A essência não é a coisa ou a qualidade. Ela é somente o ser da coisa ou da qualidade, isto é, um puro possível para cuja definição a existência não entra em conta, muito embora se dê através dela. (MARTINS; THEÓPHILO, 2016, p. 43).

Para os autores, na fenomenologia, a busca da essência do fenômeno ocorre na forma como ele se apresenta ao pesquisador.

Ou seja, é a exploração do fenômeno tal como é dado à consciência, livre de qualquer crença e de qualquer juízo. Isso deriva da busca do verdadeiro conhecimento, que, para alcançar a essência, deve ser feita através de método que não seja influenciado por conhecimentos prévios e nem da forma como o fenômeno se apresenta. Portanto prescinde da existência física e real do fenômeno (objeto). Na busca pela essência do fenômeno, parte-se da redução das características

subjetivas a invariante do objeto de pesquisa. Ou seja, o fenômeno é reduzido à sua essência através da desnudação de sua forma, persistindo, portanto, somente sua essência. (MARTINS; THEÓPHILO, 2016, p. 44).

Nesta pesquisa, a fenomenologia possibilita estabelecer relações entre os fenômenos e suas essências, e contribui para a compreensão e conceituação da Arquitetura de Dados no contexto da Ciência da Informação, o que se faz por meio da Teoria do Conceito (DAHLBERG, 1978a, 1978b).

A pesquisa se insere no paradigma físico (CAPURRO, 2003), presente na Ciência da Informação, principalmente nos estudos de recuperação da informação e, de forma mais ampla, em uma sociedade informatizada, para o qual um objeto físico/mensagem é transmitido de um emissor a um receptor. Tem forte influência do desenvolvimento tecnológico, por isso, ocupa-se de desenvolver e melhorar métodos que otimizem a gestão de dados.

A pesquisa também se insere no paradigma pós-custodial, informacional e científico, no qual há ênfase ao acesso à informação, e as Tecnologias de Informação e Comunicação (TIC) se apresentam como um dos focos de estudo da Ciência da Informação (MALHEIRO; RIBEIRO, 2011).

Um dos traços do paradigma pós-custodial é a “prioridade máxima dada ao acesso à informação, por todos em condições bem definidas e transparentes, pois só o acesso público justifica e legitima a custódia e a preservação” (MALHEIRO; RIBEIRO, 2011, p. 59). Tal paradigma é também pautado no dinamismo e na valorização da informação, considerando as influências introduzidas pelas TIC no seu processamento.

O paradigma pós-custodial se apresenta, nesta tese, como alicerce para a estrutura e o tratamento da camada de dados de ambientes informacionais digitais e sistemas de informação, considerando a relevância para o armazenamento, acesso e (re)uso em benefício do processamento humano e maquínico.

O alicerce nestes dois paradigmas possibilita tratar das abordagens computacionais e humanas relacionadas aos fenômenos estudados.

2.2 POLO TEÓRICO

No polo teórico, apresenta-se “[...] a racionalidade do sujeito que conhece e aborda o objeto, bem como a postulação de leis, a formulação de hipóteses, teorias e

conceitos operatórios e conseqüente confirmação ou infirmação do <<contexto teórico>> elaborado” (SILVA, 2006, p. 154).

Considerando a proposta de pesquisa, os problemas e os objetivos traçados, o polo teórico é constituído a partir de enfoques conceituais provenientes da Ciência da Informação e de áreas e campos fronteiriços aos fenômenos estudados, notadamente Ciência da Computação, Gestão Empresarial e Sistemas de Informação.

Isso possibilita tecer relações entre as teorias já existentes para alicerçar a conceituação de Arquitetura de Dados no contexto da Ciência da Informação, bem como identificar o relacionamento entre a Arquitetura da Informação e a Arquitetura de Dados, campos norteadores do polo teórico.

O polo teórico consiste em um quadro de referência acerca de:

- a) Arquitetura da Informação;
- b) Ecologia Informacional Complexa;
- c) Arquitetura de Dados;
- d) Dado e Informação;
- e) Ciclo de Vida dos Dados;
- f) Metadados;
- g) Estruturas Sintática e Semântica dos Dados;
- h) Princípios Éticos, Legais e Técnicos aplicados aos Dados;
- i) Ecologia Complexa de Dados.

O polo teórico que sustenta a tese é apresentado nas seções 3 e 4.

Para a elaboração da seção 3, dedicada à Arquitetura da Informação, utilizam-se como base os autores precursores da AI, cujas obras são revisitadas, o que possibilita a discussão incorporando aspectos contemporâneos ao campo, bem como a compreensão da lacuna existente na AI *bottom-up* no que tange à estruturação da camada de dados dos ambientes informacionais.

Ao passo que a seção 4, dedicada à Arquitetura de Dados, baseia-se em um levantamento bibliográfico exploratório, que auxilia na compreensão da Arquitetura de Dados e, com isso, possibilita estabelecer seus elementos conceituais essenciais no contexto da Ciência da Informação, bem como identificar possíveis relacionamentos entre a Arquitetura da Informação e a Arquitetura de Dados.

2.3 POLO TÉCNICO

No polo técnico, ocorre o contato “[...] com a realidade objectivada, aferindo-se a capacidade de validação do dispositivo metodológico, [...] sempre tendo em vista a confirmação ou refutação das leis postuladas, das teorias elaboradas e dos conceitos operatórios formulados.” (SILVA, 2006, p. 154).

No polo técnico desta tese, foi realizado estudo teórico, exploratório, descritivo, baseado em pesquisa bibliográfica e documentos técnicos, com abordagem fenomenológica. Com base em Martins e Theóphilo (2016, p. 45), cumpre destacar que:

[...] não há “o” ou “um” método fenomenológico, mas uma atitude do homem de abertura para compreender aquilo que se mostra, livre de conceitos predefinidos. A atitude fenomenológica é a atitude de retornar um caminho que conduza a enxergar o existir assim como ele se mostra. Isto é, trata-se de reorientar o olhar, a fim de desvendar o fenômeno além da aparência [...].

No que tange à construção do polo teórico, realiza-se levantamento bibliográfico da literatura científica – nos idiomas português, inglês e espanhol; nas fontes de informação nacionais e internacionais: Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (Brapci), Portal Brasileiro de Publicações Científicas em Acesso Aberto (Oasisbr) e *Web of Science*, sendo a última acessada por meio da plataforma Periódicos Capes⁴.

As temáticas do levantamento bibliográfico estão relacionadas aos objetivos propostos e ao quadro de referência apresentado no polo teórico. Considerando o percurso realizado no estudo, optou-se pela sistematização do levantamento da literatura realizado para embasar a seção 4, cuja discussão é proeminente nesta tese.

Neste levantamento, utilizam-se as seguintes expressões de busca: “Arquitetura de Dados”, “*Data Architecture*”, “Ecologia Complexa de Dados”, “Ecologia de Dados”, “*Complex Data Ecology*”, “*Data Ecology*”, cujos resultados estão sistematizados no Quadro 1.

As buscas apresentadas no Quadro 1 foram atualizadas em setembro de 2022, a fim de considerar as publicações recentes indexadas nas bases de dados.

⁴ Vale mencionar que, nesta pesquisa, foi realizada a autenticação pelo domínio UNESP. Considerando que o acesso aos conteúdos via Periódicos Capes não é universal, é possível que o a busca realizada com outra autenticação possa gerar resultados diferentes.

Quadro 1 – Levantamento Bibliográfico de Literatura

Termo de busca	Base de Dados	Quantidade de registros recuperados	Quantidade de registros recuperados após o filtro	Quantidade de registros selecionados para análise
Arquitetura de Dados	Brapci	1	1	0
	Oasisbr	33	33	2
	Scopus	0	0	0
	Web of Science	0	0	0
Data Architecture	Brapci	1	1	0
	Oasisbr	37	37	3
	Web of Science	675	195	32
Ecologia Complexa de Dados	Brapci	0	0	0
	Oasisbr	0	0	0
	Web of Science	0	0	0
Ecologia de Dados	Brapci	0	0	0
	Oasisbr	0	0	0
	Web of Science	0	0	0
Complex Data Ecology	Brapci	0	0	0
	Oasisbr	0	0	0
	Web of Science	0	0	0
Data Ecology	Brapci	0	0	0
	Oasisbr	0	0	0
	Web of Science	0	0	0
Total		747	267	37

Fonte: Autoria própria.

Nas buscas realizadas na base de dados *Web of Science* utilizando o termo “*Data Architecture*”, considerando a quantidade inicial de 675 registros recuperados, foram aplicados filtros a fim de reduzir a quantidade de itens recuperados, limitando-os não apenas pelos idiomas espanhol, inglês e português, mas também pela tipologia – artigos, artigos de revisão, capítulos de livros e livros e, pela área do conhecimento: *Computer Science* (122), *Engineering* (88), *Information Science Library Science* (21), *Science Technology Other Topics* (13), reduzindo os resultados de 675 para 195 registros⁵.

A partir dos resultados de busca, verifica-se a preponderância das discussões acerca da Arquitetura de Dados nas áreas de Ciência da Computação e Engenharia, com poucas ocorrências na área de Ciência da Informação.

A partir da análise dos registros oriundos do levantamento bibliográfico de literatura acerca da Arquitetura de Dados, observa-se que algumas obras recuperadas abordam o termo com poucas ocorrências no texto, enquanto outras, com maior incidência de ocorrências do termo ao longo do texto, utilizam a Arquitetura de Dados no contexto do desenvolvimento de sistemas de forma macro ou na abordagem do Big Data, contudo, sem se ater a discussões relacionadas à que realizamos nesta tese. Infere-se que esta abordagem esteja relacionada à predominância de textos provenientes da Ciência da Computação.

Por outro lado, há obras que abordam a Arquitetura de Dados com alguma proximidade à que se realiza neste estudo, ainda que ao contextualizar pesquisas de domínios específicos. Nestes casos, há a menção à temática da AD, contudo, sem apresentar uma conceituação ou modelo conceitual.

De igual maneira, para a construção do conceito de Arquitetura de Dados no contexto da Ciência da Informação, enquanto elemento do polo morfológico desta tese, utiliza-se a Teoria do Conceito de Dahlberg (1978a, 1978b) como elemento do polo técnico.

Assim, o levantamento bibliográfico de literatura acerca de Arquitetura de Dados e de Ecologia Complexa de Dados e as morfologias obtidas reforçam a relevância da proposta desta tese, sobretudo para a Ciência da Informação.

⁵ Expressão de busca utilizada: Resultados para “*data architecture*” (Tópico) and Artigo or Artigo de revisão (Tipos de documento) and English or Spanish (Idiomas) and *Computer Science* or *Engineering* or *Information Science Library Science* or *Science Technology Other Topics* (Áreas de pesquisa).

2.4 POLO MORFOLÓGICO

No polo morfológico, “[...] formalizam-se os resultados da investigação levada a cabo, através da representação do objecto em estudo e da exposição de todo o processo de pesquisa e análise que permitiu a construção científica em torno dele.” (SILVA, 2006, p. 155).

No polo morfológico, como resultado principal, em atendimento ao objetivo geral da pesquisa, tem-se o conceito de Arquitetura de Dados no contexto da Ciência da Informação.

Além desse, outros resultados foram alcançados:

- a) a elaboração do conceito operacional de Dados;
- b) os elementos da Arquitetura de Dados no contexto da Ciência da Informação;
- c) os entregáveis da Arquitetura de Dados;
- d) a definição de Arquiteto de Dados;
- e) a Tríade da Arquitetura de Dados;
- f) o Fluxo de Dados na Arquitetura de Dados;
- g) a Ecologia Complexa de Dados.

No que tange à Arquitetura da Informação, já consolidada na área de Ciência da Informação, a discussão apresentada nesta tese contribui com os seguintes resultados:

- a) o Reposicionamento da abordagem Pervasiva da Arquitetura da Informação para a abordagem Informacional;
- b) decorrente do reposicionamento da abordagem Pervasiva da Arquitetura da Informação para a abordagem Informacional, tem-se o Reposicionamento da Ecologia Informacional Complexa para a Arquitetura da Informação;
- c) a compreensão de que no projeto de Arquitetura da Informação, tendo a informação a característica pervasiva, é necessário um projeto de Arquitetura da Informação macro, aplicável aos diferentes ambientes, canais, mídias e tecnologias que compõem a Ecologia Informacional Complexa.

Considerando a característica do fazer científico e do processo de pesquisa, que perpassa, mas não se restringe a um único objeto finalístico, os resultados obtidos

são materializados na forma tradicional de apresentação de resultados científicos, privilegiando a comunicação científica ampliada relacionada à pesquisa, cujos objetos são: relatório de qualificação, artigos científicos, trabalhos publicados em Anais de eventos, capítulos de livros e a própria tese.

Todos os resultados da pesquisa são disponibilizados em repositório digital, de forma a conectar a produção científica com outros objetos desenvolvidos e favorecer a comunicação científica ampliada. Neste sentido, sempre que possível utiliza-se o Repositório Institucional da Universidade Tecnológica Federal do Paraná (RIUT)⁶, instituição de vínculo profissional da pesquisadora. Adicionalmente, são utilizados outros repositórios digitais de instituições de vínculo de coautores dos trabalhos, especialmente o Repositório Institucional da Unesp, dado o vínculo da pesquisadora com o Programa de Pós-Graduação em Ciência da Informação no qual a tese se desenvolveu e o vínculo da orientadora da pesquisa de doutorado.

Silva e Ribeiro (2002, p. 90) afirmam que “[...] consumada, neste quarto pólo, a exposição de todo o processo que levou à construção do objecto científico, pode reiniciar-se o ciclo dinâmico de investigação quadripolar, a qual propende visivelmente para a acumulação em espiral do conhecimento construído.”

Desta forma, a investigação, a partir do delineamento quadripolar, possibilita a vigilância e a alimentação constantes e alternadas dos quatro polos, possibilitando alterações de percursos e gerando dinamicidade, conforme o próprio processo de construção de conhecimentos, o que se faz possível pela característica não linear do delineamento quadripolar.

Apreende-se, então, que o processo investigativo e os fenômenos abordados neste estudo não se esgotam com a realização desta pesquisa, muito embora seus objetivos sejam respondidos no decorrer do processo de elaboração da tese.

Durante a realização da pesquisa, alguns resultados a ela relacionados foram comunicados por meio de objetos de produção científica, especificamente artigos, trabalhos publicados em eventos e capítulos de livros, que são relacionados na seção 2.4.1.

⁶ Universidade Tecnológica Federal do Paraná.

2.4.1 Produção Científica relacionada à Tese

O polo morfológico desta tese explicita que os resultados de pesquisa relacionados a esta tese são materializados em objetos de produção científica. Durante o processo investigativo, alguns resultados foram publicados e são, agora, elencados nesta seção, a fim de explicitar as temáticas investigadas, bem como seus relacionamentos.

NOBREGA, H. G. M.; TORINO, E.; VIDOTTI, S. A. B. G.; HALICKI, J. P. C.; OLIVEIRA, T. M. de. Integração de dados de ECTI entre diferentes sistemas de informação: proposta de solução. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 22. 2022, Porto Alegre. **Anais** [...] Porto Alegre: Universidade Federal do Rio Grande do Sul, 2022. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/viewFile/1241/800>.

TORINO, E.; BRITO, J. F.; TREVISAN, G. L.; VIDOTTI, S. A. B. G. Infraestrutura e serviços de gestão de dados de pesquisa: uma avaliação no âmbito da Universidade Estadual Paulista (UNESP). **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 27, n.1, p. 1-28, mar. 2022. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/85188>.

TORINO, E.; BRITO, J. F.; RODAS, C. M.; VIDOTTI, S. A. B. G. A relação entre arquitetura da informação e experiência do usuário sob a ótica dos pesquisadores da ciência da informação brasileira. **Biblos**, Rio Grande, v. 36, n. 1, p. 219-237, jan./jun. 2022. Disponível em: <https://periodicos.furg.br/biblos/article/view/13769/986>.

TORINO, E.; CONEGLIAN, C. S.; VIDOTTI, S. A. B. G. Coleta e reúso de dados de resultados de pesquisa para a constituição de CRIS institucional. *In*: SANTOS, Gildenir Carolino; SHINTAKU, Milton. (org.). **Ecossistemas e inovações tecnológicas**: da construção às boas práticas. Campinas: UNICAMP/BCCL; Brasília, DF: IBICT, 2022. Cap. 10, p. 231-252. Disponível em: <https://econtents.bc.unicamp.br/omp/index.php/ebooks/catalog/view/157/190/677>. Acesso em: 17 ago. 2022.

CONEGLIAN, C. S.; TORINO, E.; VIDOTTI, S. A. B. G. Inteligência Artificial e Ciência de Dados em CRIS institucional: modelo conceitual. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 21. 2021, Rio de Janeiro. **Anais** [...] Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2021. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/337>.

TORINO, E.; CONEGLIAN, C. S.; VIDOTTI, S. A. B. G. Estruturas de representação para reúso de dados no contexto da ecologia de pesquisa: CRIS institucional. **Informação & Informação**, Londrina, v. 25, n. 3, p. 1-27, jul./set. 2020. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/41946>.

TORINO, E.; PINTO, C. A. B. S.; VIDOTTI, S. A. B. G. Dos produtos e serviços à ciência de serviços: um novo olhar para a ciência da informação. *In*: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO, 7. 2017, Londrina. **Anais [...]**. Londrina: Universidade Estadual de Londrina, 2017. Disponível em:

<http://www.uel.br/eventos/cinf/index.php/secin2017/secin2107/paper/viewFile/460/284>.

TORINO, E.; ROA-MARTÍNEZ, S. M.; VIDOTTI, S. A. B. G. Dados de pesquisa: disponibilização ou publicação?. *In*: SHINTAKU, M.; SALES, L. F.; COSTA, M. Costa. (org.). **Tópicos sobre dados abertos para editores científicos**. Botucatu: ABEC, 2020. p. 183-201. Disponível em:

<https://www.abecbrasil.org.br/arquivos/Topicos dados abertos editores científicos.pdf>.

TORINO, E.; SAAVEDRA FILHO, N. C. Integração e reuso de dados para o povoamento semiautomático de dissertações e teses no Repositório Institucional da UTFPR. **Informação & Informação**, Londrina, v. 26, n. 2, p. 306-332, abr./jun. 2021. Disponível em:

<http://www.uel.br/revistas/uel/index.php/informacao/article/view/40037/pdf>.

TORINO, E.; TREVISAN, G. L.; CONEGLIAN, C. S.; BOTEGA, L. C.; SEGUNDO, J. E. S.; VIDOTTI, S. A. B. G. Enriquecimento semântico para a disponibilização de dados abertos: teoria e prática. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 25, p. 1-19, 2020. Disponível em:

<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2020.e67991>.

TORINO, E.; TREVISAN, G. L.; VIDOTTI, S. A. B. G. Dados abertos CAPES: um olhar à luz dos desafios para publicação de dados na web. **Ciência da Informação**, Brasília, v. 48, n. 3, p. 38-46, set./dez. 2019. Disponível em:

<http://revista.ibict.br/ciinf/article/view/4866/4428>.

TORINO, E.; TREVISAN, G. L.; VIDOTTI, S. A. B. G. Otimização da produção acadêmico-científica para mecanismos de busca acadêmicos (ASEO). **Biblos**, Rio Grande, v. 33, n. 2 p. 4-19, 2019. Disponível em:

<https://periodicos.furg.br/biblos/article/view/9734>.

TORINO, E.; VIDOTTI, S. A. B. G. Boas práticas para dados na web: análise do portal Dados Abertos Capes. **Informação & Sociedade**: Estudos, João Pessoa, v. 31, n. 1, p. 1-25, jan./mar. 2021. Disponível em:

<https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/50790/33168>.

TORINO, E.; VIDOTTI, S. A. B. G. Representação da informação e interoperabilidade entre ambientes científicos: um enfoque ao papel do autor. *In*: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO, 8. 2019, Londrina. **Anais [...]** Londrina: Universidade Estadual de Londrina, 2019. Disponível em:

<http://www.uel.br/eventos/cinf/index.php/secin2019/secin2019/paper/viewFile/602/410>.

TORINO, E.; VIDOTTI, S. A. B. G.; ALVES, R. C. V.; SANTOS, P. L. V. A. C. Enlace de objetos digitais no contexto da publicação científica ampliada. **Tendências da**

Pesquisa Brasileira em Ciência da Informação, v. 12, n. 2, p. 1-9, 2019.
Disponível em: <https://revistas.ancib.org/index.php/tpbci/article/view/494>.

TORINO, E.; VIDOTTI, S. A. B. G.; VECHIATO, F. L. Contribuições do atributo Metadados para a Encontrabilidade da Informação. **Em Questão**, v. 26, n. 2, p. 437-457, 2020. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/93072>.

TORINO, E.; VIDOTTI, S. A. B. G.; SANT'ANA, R. C. G. Ciclo de vida de dados no processo de publicação e acesso à produção científica. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 20. 2019, Florianópolis. **Anais** [...] Florianópolis: Universidade Federal de Santa Catarina, 2019. Disponível em: <https://conferencias.ufsc.br/index.php/enancib/2019/paper/viewFile/576/612>.

TORINO, E.; VIDOTTI, S. A. B. G.; TREVISAN, G. L. Otimização da produção acadêmico-científica para mecanismos de busca acadêmicos. *In*: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO, 8. 2019, Londrina. **Anais** [...] Londrina: Universidade Estadual de Londrina, 2019. Disponível em: <http://www.uel.br/eventos/cinf/index.php/secin2019/secin2019/paper/view/553/381>.

VIDOTTI, S. A. B. G.; TORINO, E.; CONEGLIAN, C. S. #SejaJUSTOeCUIDADOSO: princípios FAIR e CARE na gestão de dados de pesquisa. *In*: SALES, L. F.; VEIGA, V. S. de O.; HENNING, P.; SAYÃO, L. F. (org.). **Princípios FAIR aplicados à gestão de dados de pesquisa**. 1ed. Rio de Janeiro: IBICT, 2021. p. 209-221. Disponível em: <https://ridi.ibict.br/handle/123456789/1182>.

Destaca-se que, tendo em vista a relevância dos movimentos de Ciência Aberta e Acesso Aberto, e considerando seus benefícios para a ciência e a sociedade, bem como as questões relacionadas aos direitos autorais e aos direitos de acesso e uso que protegem o autor e ao mesmo tempo o usuário da informação de problemas oriundos do uso indevido de materiais protegidos integralmente por *copyright*, as publicações anteriormente relacionadas foram realizadas em periódicos, eventos e livros de acesso aberto. O mesmo se faz com a presente tese, que adota uma licença *Creative Commons*.

Com isso, a visibilidade das publicações é ampliada, por meio do acesso público, gratuito e com restrições mínimas necessárias, expostas por meio das licenças *Creative Commons* atribuídas, além da possibilidade de interoperabilidade e coleta dos metadados representacionais por diferentes bases de dados e mecanismos de buscas gerais e acadêmicos, que inegavelmente ampliam o alcance das referidas publicações e democratizam o acesso aos resultados de uma pesquisa realizada em uma universidade pública.

As estruturações conceituais realizadas nesta tese são compreendidas como morfologias e, portanto, são elementos do polo morfológico distribuídas na tese. Também se constitui elemento morfológico a própria tese que ora se apresenta.

3 ARQUITETURA DA INFORMAÇÃO

A Arquitetura da Informação é uma disciplina que tem sido estudada na Ciência da Informação desde sua gênese. Diferentes pesquisadores se dedicam a compreendê-la. Há indícios de que a Arquitetura da Informação tenha se originado na *International Business Machines Corporation* (IBM) e *Xerox Palo Alto Research Center* (PARC) (RONDA LEÓN, 2008; RESMINI; ROSATI, 2011; OLIVEIRA, 2014; BRANDT, 2020), onde o termo foi empregado no contexto das Tecnologias da Informação, com enfoque na aplicação prática.

Ronda León (2008) realizou uma análise histórica do conceito da Arquitetura da Informação, bem como sua evolução e relacionamento com outras disciplinas. Em seu estudo, o autor relaciona o surgimento do termo Arquitetura da Informação às atividades práticas profissionais desenvolvidas, sobretudo nos campos do design, da interação humano-computador e dos sistemas de informação, sendo posteriormente empregada, já conceitualmente, no projeto da informação em ambientes informacionais.

Nesse sentido, Ronda León (2008, tradução nossa) afirma que “A primeira abordagem da AI, na década de 70, estava voltada para o problema básico da desorganização da informação que nos rodeia, cuja solução consistia em proporcionar ordem a essa informação no ambiente computacional nascente.”⁷ Afirma o autor que, na década de 70, considerando o desenvolvimento da computação, as empresas passaram a adotá-la para o desenvolvimento de sistemas de informação e a criação de bases de dados, visando “[...] a gestão de dados resultantes dos processos internos.”⁸ (RONDA LEÓN, 2008, tradução nossa). Vale destacar que a tecnologia existente à época não permitia que os referidos sistemas de informação se relacionassem.

No contexto da IBM, o termo utilizado foi Arquitetura, ao se referirem às características, inovações e arquitetura de um sistema em lançamento, o IBM System/360. Neste contexto, Amdahl, Blaauw e Brooks Junior publicaram, em 1964, um artigo para apresentar os objetivos, a concepção, os fundamentos, as

⁷ “El primer enfoque de la AI, en la década del 70, se focalizaba en el problema básico de la desorganización de la información que nos rodea, cuya solución consistía en proporcionar un orden a dicha información en el naciente entorno computacional.”

⁸ “[...] la gestión de los datos resultantes de los procesos internos.”

características e os problemas de engenharia encontrados no desenvolvimento do IBM System/360, no qual enfatizam: “O termo arquitetura é usado aqui para descrever os atributos de um sistema como visto pelo programador, ou seja, a estrutura conceitual e comportamento funcional, diferente da organização do fluxo de dados e controles, o design lógico e a implementação física.”⁹ (AMDAHL; BLAAUW; BROOKS JUNIOR, 1964, tradução nossa).

De igual maneira, a Xerox Corp., na década de 1970, após a aquisição de uma empresa de informática, instalou em Palo Alto o *Xerox Palo Alto Research Center* (PARC), um centro de pesquisa dedicado à informação, ciência e tecnologia. O objetivo da companhia era que “[...] a Xerox Corp. seja ‘a arquiteta da informação’ para o escritório comercial. Assim, adotamos para o PARC o objetivo de pesquisa de conceber e testar um sistema que poderia se tornar esta arquitetura da informação.”¹⁰ (PAKE, 1985, p. 54, tradução nossa). Segundo o autor, que foi o primeiro diretor do PARC, a criação visava tirar proveito das tecnologias digitais que estavam em desenvolvimento na época, vislumbrando a fusão com as tecnologias de imagem analógicas, já conhecidas pela Xerox.

Assim, considera-se importante ponderar que os termos “arquitetura” utilizado, pela IBM (AMDAHL; BLAAUW; BROOKS JUNIOR, 1964), e “arquitetura da informação”, utilizado pela Xerox (PAKE, 1985), estão diretamente relacionados à Ciência da Computação, pois tratam da arquitetura de sistemas de informação, do processo de modernização das tecnologias utilizadas para o fluxo de informação, no sentido da transmissão de pacotes utilizando *Ethernet*, do desenvolvimento de *hardware*, *software* e dispositivos para computadores. Contudo, IBM e Xerox não abordam a informação em si, no que tange à sua estrutura, organização, representação, acesso e uso, e, com isso, distanciam-se da Arquitetura da Informação como discutida na Ciência da Informação.

Corroboram essa afirmação Resmini e Rosati (2011, p. 20, tradução nossa):

Não é discutível que estamos falando de arquiteturas de computador aqui, discos e caixas e fios e hubs, mas a maneira pela qual o termo arquitetura é abstraído e conceituado em conexão com estrutura e

⁹ “The term architecture is used here to describe the attributes of a system as seen by the programmer, i.e., the conceptual structure and functional behavior, as distinct from the organization of the data flow and controls, the logical design, and the physical implementation.”

¹⁰ “[...] the Xerox Corp. ‘the architect of information’ for the business office. Thus we adopted for PARC the research goal of devising and testing a system that could become this architecture of information.”

comportamento e não apenas *layouts* físicos lançou as bases para a extensão subsequente de seu uso para outras áreas da computação.¹¹

Em seu estudo, Ronda León (2008) afirma que, àquela época, a Arquitetura da Informação carecia de amplo referencial teórico, em decorrência da característica prática da disciplina, cuja teoria estava relacionada às experiências práticas dos profissionais que a desenvolviam.

Ainda em busca dos precursores da Arquitetura da Informação no domínio da Ciência da Informação, de acordo com Rosenfeld, “Richard Saul Wurman cunhou o termo cerca de trinta anos atrás, e outros desde então (incluindo nós) vieram com várias definições do termo, algumas bastante semelhantes, outras não.”¹² (ROSENFELD; MORVILLE, 1998, p. 163, tradução nossa). Os autores destacam, ainda, que o emprego da AI é realizado há séculos no processo de organização e navegação dos diferentes artefatos, por exemplo, livros, websites, mapas, bibliotecas e cidades.

Assim, considera-se que, no campo teórico, os precursores da AI, Wurman (1996), Rosenfeld e Morville (1998) e Morville e Rosenfeld (2006) trataram do seu conceito inicial, que tem sido utilizado científica e profissionalmente em diferentes contextos. Neste sentido, Wurman (1996) se concentra no *layout* e na apresentação da informação em ambientes bidimensionais, ao passo que Rosenfeld e Morville (1998) e Morville e Rosenfeld (2006) tratam da estrutura e organização da informação em websites, entendendo-os como ambientes multidimensionais.

Posteriormente, outros estudos teóricos trouxeram incorporações aos conceitos estabelecidos por Wurman, Rosenfeld e Morville, em autores como Vidotti, Cusin e Corradi (2008), Resmini e Rosati (2011), Camargo e Vidotti (2011) Siqueira (2012), Vechiato (2013), Oliveira (2014), Lacerda (2015), Rosenfeld, Morville e Arango (2015), Brandt (2020).

Na Ciência da Informação, em geral, os estudos de AI se baseiam na definição de Morville e Rosenfeld (2006, p. 4, tradução nossa), para os quais a Arquitetura da Informação consiste em:

¹¹ “It is not disputable that we are talking computer architectures here, disks and boxes and wires and hubs, but the way in which the term architecture is abstracted and conceptualized in connection with structure and behavior and not just physical layouts laid the basis for the subsequent extension of its use to other areas of computing.”

¹² “Richard Saul Wurman coined the term about thirty years ago, and others since then (including us) have come up with varying definitions of the term, some quite similar, some not.”

1. O projeto estrutural de ambientes informacionais compartilhados.
2. A combinação de sistemas de organização, rotulagem, busca e navegação em websites e intranets.
3. A arte e ciência de moldar produtos e experiências de informação para suportar a usabilidade e a localização.
4. Uma disciplina emergente e uma comunidade de prática focada em trazer princípios de design e arquitetura para o cenário digital.¹³

Cumpre destacar que a primeira edição da obra de Rosenfeld e Morville (1998) se concentra em websites (em suas várias manifestações, incluindo intranets e portais corporativos), o que ocorre também nas edições posteriores 2002 e 2006. Contudo, a quarta edição, publicada por Rosenfeld, Morville e Arango (2015), é marcada pela mudança de cenário da época, um cenário no qual os canais, mídias, dispositivos e tecnologias evoluíram, mantendo válidas as mesmas estruturas apresentadas nas obras publicadas por Rosenfeld e Morville anteriormente. Assim, passam a tratar de vários canais de interação, considerando os dispositivos disponíveis à época. Com isso, é possível verificar que a dissociação da informação de um único ambiente de armazenamento torna os preceitos da AI aplicáveis a diferentes tecnologias, mídias, canais e ambientes informacionais.

Assim, a definição apresentada por Morville e Rosenfeld (2006) foi atualizada em Rosenfeld, Morville e Arango (2015, p. 24, tradução nossa)

1. O projeto estrutural de ambientes informacionais compartilhados.
2. A combinação de sistemas de organização, rotulagem, busca e navegação em ecossistemas digitais, físicos e cross-channel.
3. A arte e a ciência de moldar produtos e experiências de informação para suportar a usabilidade, a localização e a compreensão.
4. Uma disciplina emergente e uma comunidade de prática focada em trazer princípios de design e arquitetura para o cenário digital.¹⁴

Em entrevista concedida a Resmini, Rosenfeld afirma que “[...] podemos aplicar a arquitetura da informação em todos os lugares, não apenas na Web.”¹⁵ (ROSENFELD, 2021, p. 79, tradução nossa), e continua “[...] a arquitetura da

¹³ “1. The structural design of shared information environments.

2. The combination of organization, labeling, search, and navigation systems within web sites and intranets.

3. The art and science of shaping information products and experiences to support usability and findability.

4. An emerging discipline and community of practice focused on bringing principles of design and architecture to the digital landscape.”

¹⁴ “1. The structural design of shared information environments

2. The synthesis of organization, labeling, search, and navigation systems within digital, physical, and cross-channel ecosystems

3. The art and science of shaping information products and experiences to support usability, findability, and understanding

4. An emerging discipline and community of practice focused on bringing principles of design and architecture to the digital landscape”

¹⁵ “[...] we can apply information architecture everywhere, not just to the Web.”

informação está em toda parte.”¹⁶ (ROSENFELD, 2021, p. 83, tradução nossa).¹⁷ Ao que Resmini, ao se referir à Arquitetura da Informação inicialmente discutida no domínio da web, acrescenta “[...] a informação digital tornou-se uma parte penetrante e integral do tecido da realidade de uma forma que não era sequer imaginável naquela época. A próxima afirmação óbvia seria que o papel da arquitetura da informação é ainda mais crucial hoje do que há dez ou vinte anos.”¹⁸ (ROSENFELD, 2021, p. 80, tradução nossa).

Os estudos de AI se alicerçam ainda no conceito de Hagedorn (2000, p. 5, tradução nossa): “Arquitetura da Informação: arte e ciência da organização da informação para ajudar efetivamente as pessoas a satisfazer suas necessidades de informação. Envolve investigação, análise, projeto e implementação.”¹⁹

Camargo e Vidotti (2011) destacam a característica multidisciplinar da Arquitetura da Informação, que envolve aspectos do design, da ergonomia, da usabilidade, da interação humano-computador, da Ciência da Computação, da Ciência da Informação e de outras áreas, à medida que trabalha com a organização e a forma de apresentação da informação. Contudo, para as autoras, a AI também aborda a representação e os metadados, camada imperceptível ao usuário. A Arquitetura da Informação, além de auxiliar na estruturação dos ambientes digitais, ao promover acesso à informação armazenada, auxilia os processos de gestão, sobretudo da informação e do conhecimento.

Lacerda (2015, p. 110) defende “[...] uma visão da AI na qual a ‘informação’ é sua matéria-prima, materializada em ‘objetos’ delimitados por ‘espaços’ intencionalmente concebidos de modo a promover ‘experiências de usuários’.”

Enquanto Rosenfeld, Morville e Arango (2015, p. 58-59, tradução nossa) afirmam que a arquitetura objetiva projetar ambientes que atendam e comuniquem suas funções sociais de forma eficaz, assim como a arquitetura da informação, que

¹⁶ “[...] *information architecture is everywhere.*”

¹⁷ O percurso realizado nesta tese nos levou a revisitar os autores que são precursores da Arquitetura da Informação. Com isso, as quatro edições da obra “*Information Architecture*”, conhecida como o livro do Urso Polar, foram analisadas. Nas primeiras edições (1998, 2002, 2006), os autores se referem a websites, intranets e extranets, ao passo que na edição de 2015 referem-se a ecossistemas digitais, físicos e *cross-channel*, de forma que, ao longo da tese, foram utilizadas citações às edições da obra que mencionam websites, intranet e extranet, mas cujo contexto se aplica às tecnologias e dispositivos atuais.

¹⁸ “[...] *digital information has become a pervasive, integral part of the fabric of reality in a way that was not even imaginable back then. The obvious next statement would be that the role of information architecture is even more crucial today than it was ten or twenty years ago.*”

¹⁹ “*The art and science of organizing information to help people effectively fulfill their information needs. Information architecture involves investigation, analysis, design and implementation.*”

“[...] define composições de elementos semânticos, como rótulos de navegação, títulos de seção e palavras-chave, e produz o projeto de princípios, objetivos e diretrizes que capturam o sentimento pretendido do lugar [...]”²⁰, ou seja, seu contexto.

Para Camargo e Vidotti (2011, p. 24):

A arquitetura da informação é uma área do conhecimento que oferece uma base teórica para tratar aspectos informacionais, estruturais, navegacionais, funcionais e visuais de ambientes informacionais digitais por meio de um conjunto de procedimentos metodológicos a fim de auxiliar no desenvolvimento e no aumento da usabilidade de tais ambientes e de seus conteúdos.

Assim, verifica-se que a AI se apresenta como um campo de atividade interdisciplinar, que objetiva otimizar questões relacionadas à usabilidade e à encontrabilidade da informação²¹ em ambientes digitais. Dessa forma, trata de aspectos relacionados à organização, representação, armazenamento, preservação, acesso e uso da informação, atentando-se ao comportamento informacional.

Oliveira (2014) revisita a literatura científica para embasar e representar, conceitualmente, o que considera como quatro abordagens disciplinares relacionadas à Arquitetura da Informação (Figura 2), bem como as suas fundações interdisciplinares, a saber:

- a) arquitetural, “um campo que desenvolve padrões e orientações arquiteturais para tornar os espaços de informação organizados e acessíveis” (p. 91), cujas áreas interdisciplinares são arquitetura e design, por isso o diálogo entre função e beleza;
- b) sistêmica, fundamentada na teoria geral dos sistemas e na atuação no campo dos sistemas de informação, para os quais os ambientes de informação são um conjunto de sistemas que constituem a parte de um todo complexo. O autor destaca a aproximação com a Ciência da Informação realizada por Vidotti, Cusin e Corradi (2008) que, a partir da obra de Morville e Rosenfeld (2006), promovem uma ligação com a abordagem informacional;

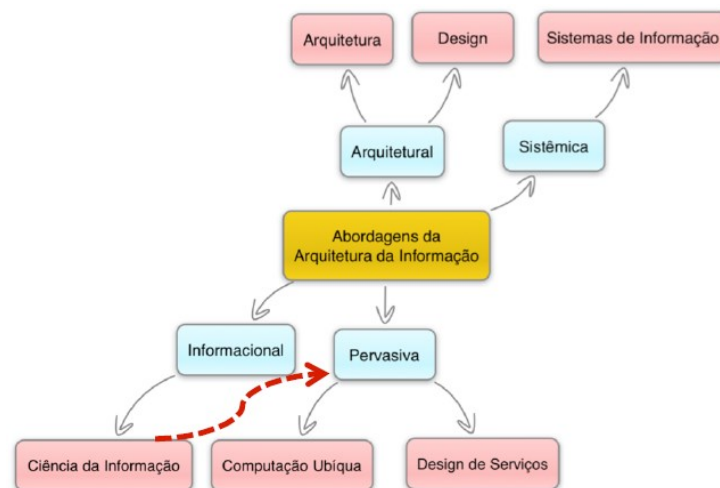
²⁰ “[...] defines compositions of semantic elements such as navigation labels, section headings, and keywords, and produces the design principles, goals, and guidelines that capture the intended feeling of the place [...]”

²¹ “A encontrabilidade da informação sustenta-se fundamentalmente na interseção entre as funcionalidades de um ambiente informacional e as características dos sujeitos informacionais.” (VECHIATO; VIDOTTI, 2014, p. 164).

- c) informacional, possui relação direta com a Biblioteconomia e a Ciência da Informação e com aspectos do tratamento, organização, armazenamento, recuperação e preservação da informação;
- d) e pervasiva, uma abordagem emergente, cujo campo conceitual ainda está em construção e que se relaciona com a computação ubíqua e com o design de serviços, baseada “na complexidade, no holismo e na integração de espaços, ambientes, tecnologias, canais e comportamentos dos sujeitos” (p. 132).

O autor destaca ainda que tais abordagens podem se mesclar, conectar ou sobrepor, e que nenhuma delas se extingue, embora possa, no percurso histórico da disciplina, se enfraquecer.

Figura 2 – Abordagens da Arquitetura da Informação



Fonte: Oliveira (2014, p. 83).

A abordagem arquitetural está relacionada às discussões de Wurman, que nasceram da perspectiva de melhorar a compreensão da informação para a interação em ambientes urbanos. A discussão se concentra no planejamento da informação de forma que ela possa ser estável, funcional e adequadamente apresentada, em benefício da compreensão humana.

As abordagens sistêmica e informacional, embora utilizem bases diversas, se conectam em alguns aspectos, ao considerar a necessidade de estruturar e organizar a informação em determinado ambiente informacional e/ou sistema de informação. Para Oliveira (2014), a abordagem sistêmica da AI, embasada na Teoria dos Sistemas, atua de forma holística na estruturação das partes que constituem o todo,

e sua atuação está mais atrelada às soluções tecnológicas do que ao uso da informação. Enquanto a abordagem informacional, estreitamente ligada à Ciência da Informação, atua em aspectos mais relacionados à organização, representação e uso da informação.

Ao passo que a abordagem pervasiva, discutida por Oliveira (2014), faz uma aproximação das teorias já consolidadas na AI aos conceitos de ubiquidade e pervasividade, marcas do avanço tecnológico na atualidade, em decorrência do protagonismo dos sujeitos na interação e uso das TIC. Assim, a Arquitetura da Informação Pervasiva, para Oliveira (2014), consiste em uma nova abordagem da AI, ao considerar “[...] os processos de hibridização dos espaços humanos onde os sujeitos vivem, trabalham e divertem-se aos ambientes de informação digital.” (p. 109), somado ao advento das TIC que possibilitam que a informação disponível possa ser acessada por diferentes dispositivos, que podem exigir arquiteturas diferentes, considerando a necessidade de responsividade.

Por outro lado, Resmini e Rosati (2011) utilizam a cronologia de Ronda León (2008) e apresentam três abordagens para a Arquitetura da Informação, que se diferenciam pela forma como trabalham a informação:

- a) abordagem do design da informação: “Arquitetura da informação como entendimento”²², alicerçada na obra de Wurman, que considera “[...] design e arquitetura são a base para uma ciência e arte de criar ‘instruções para espaços organizados’”²³ (p. 23);
- b) abordagem dos sistemas de informação: “Arquitetura da informação como gestão da informação dentro de uma visão de negócios.”²⁴ Na qual a informação é considerada um recurso. Os autores reconhecem que houve, nesta abordagem, uma alteração a partir da obra de Rosenfeld e Morville, com uma aproximação do design de experiência do usuário (p. 25);
- c) abordagem da ciência da informação: “Arquitetura da informação para a web”²⁵, cujo tratamento empírico inicial partiu de Rosenfeld e Morville, como formar de estruturar e organizar websites. (p. 27).

²² *“Information architecture as understanding”*

²³ *“design and architecture are the basis for a science and art of creating ‘instruction(s) for organized space”*

²⁴ *“Information architecture as information management within a business vision”*

²⁵ *“Information architecture for the Web”*

Este apanhado histórico inicial foi realizado com o intuito de identificar, no domínio da Ciência da Informação, os autores tidos como precursores da Arquitetura da Informação, cujas obras compõem o polo teórico deste estudo.

Diante do exposto e considerando os objetivos específicos desta tese (a e d), especialmente no que concerne a discutir a Arquitetura da Informação e Ecologia Informacional Complexa no contexto da Ciência da Informação e investigar as relações entre Arquitetura da Informação e Arquitetura de Dados no contexto da Ciência da Informação, tem-se a necessidade de buscar nas bases conceituais da Arquitetura da Informação (AI), no contexto da Ciência da Informação, elementos que contribuam para a estruturação do polo teórico do estudo.

Desta forma, a abordagem aos fenômenos estudados aponta como percurso a análise exploratória das obras supramencionadas, visando embasar o primeiro capítulo do polo teórico desta tese, notadamente a Arquitetura da Informação (AI), de forma a contribuir para compreender as lacunas existentes, bem como os relacionamentos da AI com da Arquitetura de Dados (AD), cuja conceituação na Ciência da Informação é objetivo desta tese.

A análise exploratória tem início em Wurman (1996), que introduziu sua obra afirmando que, naquela época, estávamos vivenciando um tsunami de dados produzidos e compartilhados por diferentes mídias e dispositivos, em velocidade e quantidade aceleradas. “Trata-se de um maremoto de dados não relacionados e crescentes, formados por bits e bytes, vinda em uma cacofonia de espuma desorganizada, descontrolada e incoerente. Está cheia de destroços e detritos.”²⁶ (WURMAN, 1996, p. 15, tradução nossa).

A analogia feita por Wurman (1996) utilizando o maremoto, um fenômeno natural resultante de um abalo sísmico que causa o movimento acelerado das ondas do mar, e o fenômeno crescente de bits e bytes, que navegam os ambientes informacionais, com relacionamentos cacófatos é ainda retrato comum nos ambientes informacionais digitais.

Rosenfeld e Morville (1998, p. 20, tradução nossa), afirmam que “A maneira como organizamos, rotulamos e relacionamos as informações influencia a maneira como as pessoas as compreendem.”²⁷

²⁶ “This is a tidal wave of unrelated, growing data formed in bits and bytes, coming in an unorganized, uncontrolled, incoherent cacophony of foam. It’s filled with flotsam and jetsam.”

²⁷ “The way we organize, label, and relate information influences the way people comprehend that information.”

Vale destacar que Morville e Rosenfeld (2006), ao apresentarem a definição de Arquitetura da Informação, reforçam a sua complexidade, e afirmam que a representação, quer seja da informação, de um conteúdo ou de um rótulo, não é capaz de expressar de forma completa e precisa a entidade representada. Assim sendo, a representação pode ser compreendida de forma diversa pelos usuários.

Nesse sentido, Wurman, ao se designar Arquiteto da Informação, afirma:

Quero dizer arquiteto na criação de princípios sistêmicos, estruturais e ordenados para fazer algo funcionar - a criação cuidadosa de artefato, ideia ou política que informa porque é claro. Eu uso a palavra informação no seu sentido mais verdadeiro. A maior parte da palavra informação contém a palavra informar, então eu chamo as coisas de informação apenas se elas me informarem, não se forem apenas coleções de dados, de coisas.²⁸ (WURMAN, 1996, p. 16, tradução nossa).

Wurman (1991) defende que não há uma explosão de informações e sim de dados, que precisam se tornar informação compreensível. “A compreensão da diferença entre os dados brutos e aqueles que podem ajudar na compreensão e aumentar o conhecimento, entre a informação como coisa e a informação como significado, tornará vocês um processador de informação mais competente.” (WURMAN, 1991, p. 45), e continua afirmando que é necessário reconhecer o que é compreensível e que, em alguns casos, a própria informação prejudica sua compreensão.

O autor critica, ainda, o cuidado excessivo com o projeto e as formas de apresentação, que podem tornar a informação menos compreensível, e a prevalência do belo em detrimento do funcional, ainda que isso não se faça de forma consciente. A Arquitetura da Informação seria a alteração deste cenário, no desenvolvimento de trabalhos em diferentes suportes, dentre os quais destaca a interface eletrônica.

Rosenfeld e Morville (1998) destacam a relevância de atentar-se a quem será o usuário e quais as razões para uso do ambiente, visando estabelecer o que será apresentado durante o uso, para que o ambiente seja adequadamente estruturado e utilizável. Neste sentido, afirmam que o trabalho do arquiteto da informação é estruturar o ambiente para que as pessoas que o utilizem tenham uma experiência positiva, que as faça voltar a utilizá-lo. Para tanto, não deve haver preocupação

²⁸ “I mean architect as in the creating of systemic, structural, and orderly principles to make something work-the thoughtful making of either artifact, or idea, or policy that informs because it is clear. I use the word information in its truest sense. Most of the word information contains the word inform, so I call things information only if they inform me, not if they are just collections of data, of stuff.”

excessiva com a estética ou com o funcionamento do ambiente, mas sim com a forma de unir essas duas características em uma estrutura adequada para o ambiente informacional.

Cada ambiente informacional requer uma Arquitetura da Informação que corresponda aos objetivos e usuários que pretende atingir. Isso exige que o arquiteto da informação seja um usuário crítico de ambientes informacionais, para que molde sua atuação, somando a isso a atenção aos objetivos do ambiente informacional e do seu mantenedor, além das necessidades dos múltiplos usuários potenciais, o que possibilita o desenvolvimento de projetos centrados no usuário, utilizando as tecnologias adequadas (ROSENFELD; MORVILLE, 1998; MORVILLE; ROSENFELD, 2006). Segundo os autores, os projetos centrados no usuário podem exigir contato com pessoas internas e externas ao projeto, a fim de coletar suas perspectivas, considerando aspectos positivos e negativos, e utilizá-las no planejamento e desenvolvimento do ambiente informacional.

Para o planejamento e desenvolvimento de ambientes informacionais de forma adequada, idealmente devem ser envolvidos especialistas de diferentes áreas, dentre os quais está o arquiteto da informação (WURMAN, 1996; ROSENFELD; MORVILLE, 1998; MORVILLE; ROSENFELD, 2006). Rosenfeld e Morville (1998, p. 17, tradução nossa) por sua vez, alertam que “[...] todos (incluindo nós) são influenciados por sua perspectiva disciplinar. Se possível, tente garantir que outras disciplinas sejam representadas em sua equipe de desenvolvimento do website para garantir uma arquitetura equilibrada.²⁹”. A formação da equipe responsável pelo projeto de AI do ambiente informacional pode ser decisiva no seu sucesso.

Para Wurman (1996, tradução nossa),

Arquiteto da Informação: 1) o indivíduo que organiza os padrões inerentes aos dados, deixando claro o complexo; 2) uma pessoa que cria a estrutura ou mapa de informação que permite que outros encontrem seus caminhos pessoais para o conhecimento; 3) a ocupação profissional emergente do século 21, atendendo às necessidades da época, com foco na clareza, na compreensão humana e na ciência da organização da informação.³⁰

²⁹ “[...] everyone (including us) is biased by their disciplinary perspective. If possible, try to ensure that other disciplines are represented on your web site development team to guarantee a balanced architecture.”

³⁰ “Information Architect: 1) the individual who organizes the patterns inherent in data, making the complex clear; 2) a person who creates the structure or map of information which allows others to find their personal paths to knowledge; 3) the emerging 21st century professional occupation addressing the needs of the age focused upon clarity, human understanding and the science of the organization of information.”

Enquanto para Rosenfeld e Morville (1998, p. 13, tradução nossa), o arquiteto da informação:

- Esclarece a missão e a visão do site, equilibrando as necessidades de sua organização patrocinadora e as necessidades de seu público.
- Determina qual conteúdo e funcionalidade o site conterá.
- Especifica como os usuários encontrarão as informações no site, definindo sua organização, navegação, rotulagem e sistemas de busca.
- Mapeia como o site irá acomodar as mudanças e o crescimento ao longo do tempo.³¹

No que tange à necessidade de profissionais que atuem com Arquitetura da Informação, Morville e Rosenfeld (2006, p. 17, tradução nossa) afirmam que, independentemente da designação que se dê à função, é imprescindível que sejam “[...] profissionais com habilidades e experiência especializadas, que sabem como criar sistemas de informação úteis e utilizáveis em ambientes massivamente complexos.³²”, além de destacar a complexidade da própria disciplina de Arquitetura da Informação. Nesse sentido, reforçam que, à época, não havia cursos universitários que formassem arquitetos da informação, mas áreas com foco na informação podem ser fonte de bons profissionais para desempenhar as atividades, dentre as quais destacam: Design Gráfico e Design de Informação, Ciência da Informação e Biblioteconomia, Jornalismo, Engenharia de Usabilidade, Marketing, Ciência da Computação, Arquitetura, Gestão de Produtos.

A chamada Arquitetura da Informação clássica (RESMINI; ROSATI, 2011), cujos precursores são Rosenfeld e Morville (1998), possui o que designam como “anatomia da Arquitetura da Informação”: sistemas de organização, navegação, rotulagem e busca, que alicerçam e possibilitam a evolução estruturada dos ambientes informacionais projetados. Posteriormente, Morville e Rosenfeld (2006) incorporam à “anatomia da Arquitetura da Informação” Tesouros, Vocabulário Controlado e Metadados, sem rotulá-los como um sistema mas, notadamente, entendendo a relevância do tratamento da informação. Por isso, nesta tese,

³¹ “Clarifies the mission and vision for the site, balancing the needs of its sponsoring organization and the needs of its audiences.

- Determines what content and functionality the site will contain.
- Specifies how users will find information in the site by defining its organization, navigation, labeling, and searching systems.
- Maps out how the site will accommodate change and growth over time.”

³² “[...] professionals with specialized skills and experience, who know how to create useful, usable information systems within massively complex environments.”

designamo-los como sistema de representação da informação, de acordo com Oliveira e Vidotti (2012).

Assim, é possível visualizar em Rosenfeld e Morville (1998) e Morville e Rosenfeld (2006) que a estruturação da Arquitetura da Informação ocorre de forma *top-down* e *bottom-up*. Vale destacar que os autores não se referem a isso como abordagens da AI, e sim como parte da sua anatomia. Cumpre-se esclarecer que a própria Arquitetura da Informação é *top-down* e *bottom-up*, a primeira consiste na decomposição da AI, para tornar compreensíveis seus sistemas e subsistemas, de forma hierárquica, partindo do mais genérico para o mais específico; enquanto a segunda trata da estruturação e descrição do nível mais baixo da AI, que possibilita compor o todo a partir da menor unidade. No contexto desta tese, compreende-se que a Arquitetura de Dados se conecta à Arquitetura da Informação *bottom-up* e pode contribuir com o desenvolvimento e melhoria da Arquitetura da Informação *top-down*.

Rosenfeld e Morville (1998) e Morville e Rosenfeld (2006) afirmam que os usuários desejam localizar as informações de que necessitam e que suas necessidades são diversas. Nesse sentido, há aqueles usuários que definem com exatidão o que precisam e sabem onde e como localizar, enquanto há outros que não sabem o que estão procurando e acabam localizando algo do seu interesse durante a exploração do ambiente. Essas ações são aprofundadas na obra de Vechiato (2013), quando trata da busca de informações e descoberta acidental ou serendipidade (VECHIATO; FARIAS, 2020) no contexto da Encontrabilidade da Informação.

Ao tratar de ambientes informacionais, é importante esclarecer que, nesta tese, adota-se esta terminologia para designar qualquer ambiente que armazene e disponibilize informações, desde um ambiente físico, até os digitais, como websites, intranets, extranets e sistemas de informação, incluindo aqueles que, pelas características de disponibilização de informação analógica e digital, são considerados híbridos.

Assim, é possível esclarecer que o advento das tecnologias reposiciona os papéis desenvolvidos pelos atores que circundam e interagem com os ambientes informacionais. Neste sentido, pode-se compreender a partir da Ciência de Serviços³³

³³ O crescimento do setor de serviços despertou em pesquisadores o interesse pela investigação dessa nova lógica e, com isso, propuseram a criação de uma nova área designada como Ciência de Serviços, que considera conhecimento e as competências para favorecer relações e criar valor, não apenas a um serviço, mas a um sistema de serviços. Nesse sentido, o sistema de serviço é entendido como uma rede de pessoas, tecnologias, informações compartilhadas e outros recursos que interagem entre si e se conectam a outros sistemas de serviço para a cocriação de valor (MAGLIO *et al.*, 2009).

que o usuário não possui mais um papel estático e passivo de mero consumidor de produto, serviço ou informação. Ao contrário, à medida que interage com esses produtos, serviços, informações e ambientes, ele também age como fornecedor, em um processo designado cocriação de valor³⁴.

Isso ocorre nas interações que possibilitem a contribuição com conteúdos e feedbacks, e até mesmo ao realizar a representação da informação em ambientes de autoarquivamento, a exemplo do que ocorre no cenário acadêmico-científico por meio das submissões de artigos, trabalhos em eventos e livros, bem como no autoarquivamento em repositórios digitais. Essa alteração da interação do sujeito com o ambiente se aproxima das teorias da Ciência de Serviços, especificamente na cocriação de valor.

Além disso, no uso de ambientes digitais, esse processo, embora possa passar despercebido, ocorre quando o usuário fornece informações para o ambiente e, com elas, pode auxiliar no desenvolvimento e reposicionamento dos ambientes e informações como, por exemplo, rastreamento de consultas e de navegação, *cookies*, pesquisas *online*, cadastros em sistemas de informação, feedbacks e testes de usabilidade que, dentre outros fatores, alimentam bases de dados com informações pessoais, de preferências e de localização. O fornecimento de informações por parte do usuário também molda experiências futuras em pesquisas em motores de busca e sistemas de recomendação de produtos, até mesmo auxilia em bases de treinamento de agentes computacionais por meio de Inteligência Artificial.³⁵

Desta forma, nesta tese, considerando a necessidade de manter a tríade da Arquitetura da Informação apresentada por Rosenfeld e Morville (2006) – conteúdo, contexto e usuário –, utilizaremos o termo usuário para designar os atores humanos que interagem com os sistemas de informação e ambientes informacionais digitais, quer seja para a interação, produção e/ou uso da informação. Enquanto, no contexto da Arquitetura de Dados, o usuário principal são aplicações e agentes computacionais, os quais designaremos por máquinas, cuja interação pode ser de fornecimento ou consumo de dados, considerando ainda que ambos, fornecimento e

³⁴ Considerando o sistema de serviços, a interação entre fornecedor e consumidor possibilita a cocriação de valor para o serviço, o que gera benefícios para os serviços, seus fornecedores e consumidores (MAGLIO *et al.*, 2006).

³⁵ Destaca-se que esse tipo de interação pode trazer inúmeras possibilidades de análises, interpretações e predições. Contudo, devem ser realizadas apenas quando os dados coletados tenham propósito de uso. Além disso, é importante que os sujeitos envolvidos nas coletas estejam cientes do processo e as autorizem.

consumo, possam ocorrer concomitantemente. Assim, não trataremos apenas de usuários, mas também de consumidores e fornecedores.

Considerando que “[...] a arquitetura da informação trata de compreender e transmitir a visão geral de um website.”³⁶ (ROSENFELD; MORVILLE, 1998, p. 15, tradução nossa)³⁷, deve ficar claro que os objetivos, o público e os conteúdos do ambiente informacional devem ser discutidos e estabelecidos antes que o ambiente seja desenvolvido. Com isso, a AI auxilia na definição da forma e ordem de apresentação dos conteúdos para que não haja sobrecarga das páginas, sobretudo da principal, e, ainda, do consumidor ao localizar o que tem interesse.

Fortes arquiteturas de informação são especialmente importantes para grandes websites: para desbloquear o poder encontrado nesses enormes volumes de conteúdo, esses sites precisam de sistemas de navegação e esquemas organizacionais que apresentem as informações que as pessoas precisam saber e ocultem as que, de outra forma, entrariam no caminho.³⁸ (ROSENFELD; MORVILLE, 1998, p. 12, tradução nossa).

Verifica-se em Vidotti, Cusin e Corradi (2008) que os sistemas da Arquitetura da Informação apresentados por Rosenfeld e Morville (1998) são estruturados por elementos da Ciência da Informação e acrescidos de usabilidade e acessibilidade, preponderantes para acesso e uso.

A Arquitetura da Informação preocupa-se com a estruturação e o planejamento de interfaces digitais, constituindo-se na estruturação do ambiente com relação à organização, representação e apresentação dos conteúdo informacionais, no *design* do *website*, e nos elementos de acessibilidade e usabilidade destes espaços. (VIDOTTI; CUSIN; CORRADI, 2008, p. 176).

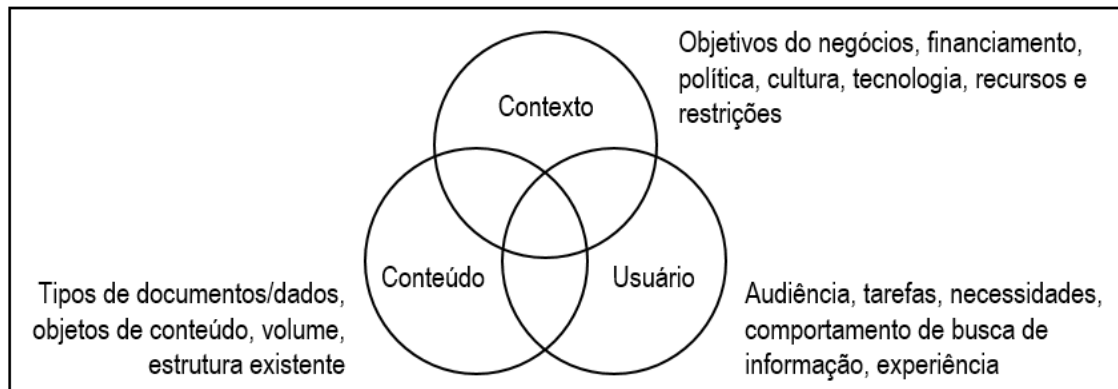
³⁶ “[...] *information architecture is about understanding and conveying the big picture of a web site.*”

³⁷ A primeira edição da obra de Rosenfeld e Morville (1998) se concentravam em sites (em suas várias manifestações, incluindo intranets e portais corporativos). Contudo, a quarta edição, publicada por Rosenfeld, Morville e Arango (2015) é marcada pela mudança de cenário da época, no qual os canais, mídias, dispositivos e tecnologias evoluíram, mantendo válidas as mesmas estruturas apresentadas nas obras publicadas por Rosenfeld e Morville em 1998, 2002 e 2006. “A experiência de interação com a informação de muitas pessoas ocorre cada vez mais por meio de aplicativos de smartphones e outros canais que não envolvem um navegador web tradicional. Além disso, à medida que os componentes e sensores do sistema ficam cada vez menores e mais baratos, o acesso bidirecional à informação está se tornando uma parte fundamental de objetos do cotidiano, como termostatos e maçanetas, que não são percebidos como dispositivos de computação tradicionais. Embora muitas dessas experiências não exijam os mesmos tipos de estruturas semânticas que os sites tradicionais, elas ainda são componentes-chave nos ecossistemas de informação e, portanto, sujeitas a muitos dos mesmos princípios de design apresentados nas edições anteriores do livro. Ao considerar o assunto de nossos projetos em abstrato – como ambientes de informação em vez de sites – podemos ver que os princípios de design que informam essas estruturas semânticas têm ampla aplicabilidade além do design para a Web.” (ROSENFELD; MORVILLE; ARANGO; 2015, p. xii, tradução nossa).

³⁸ “*Strong information architectures are especially important for large web sites: to unlock the power found in those huge volumes of content, these sites need navigation systems and organizational schemes that feature the information that people need to know and hide the stuff that would otherwise get in the way.*”

As discussões apresentadas elucidam que a Arquitetura da Informação atua especificamente com três elementos, considerados como a tríade contexto, conteúdo e usuário (Figura 3), preponderantes para o planejamento e desenvolvimento de ambientes informacionais.

Figura 3 – Os três círculos da Arquitetura da Informação



Fonte: Morville e Rosenfeld (2006, p. 25, tradução nossa).

Deve-se considerar que todo ambiente informacional está atrelado a um contexto, que pode ser o vínculo a uma pessoa, empresa, instituição ou organização, no qual estão imbricados missão, visão, valores e outras características que individualizam o ambiente (MORVILLE; ROSENFELD, 2006). Com isso, é importante que a Arquitetura da Informação atue de forma a expressar esse contexto, materializando-o no ambiente projetado, com o uso de linguagem e demais elementos de forma concernente ao contexto que será trabalhado.

O conteúdo é definido por Morville e Rosenfeld (2006) de forma ampla, abrangendo informações, documentos, serviços, metadados e demais materiais que serão disponibilizados no ambiente informacional e, quando o ambiente informacional é digital, torna-se necessário utilizar textos, imagens estáticas e dinâmicas e outros elementos para transmitir significado e realizar transações. Para tanto, alertam que há algumas facetas relacionadas ao conteúdo: propriedade, formato, estrutura, metadados, volume e dinamismo. Assim, a característica dos conteúdos e a combinação dessas facetas auxiliam na individualização do ambiente.

Quanto ao usuário, suas preferências, necessidades informacionais e comportamento influenciam na estruturação do ambiente informacional. Por isso, para Morville e Rosenfeld (2006), é imprescindível compreender quem e como usa o ambiente e o que quer dele.

Ressalta-se que a obra de Rosenfeld, Morville e Arango (2015) mantém a tríade da AI inalterada. Contudo, aponta uma alteração substancial de cenário, sobretudo pela evolução dos canais, mídias, dispositivos e tecnologias, cujos usos e interações devem ser considerados para favorecer a recuperação da informação e a descoberta nos ambientes informacionais.

É essa combinação de conteúdo, contexto e usuário que torna o projeto de Arquitetura da Informação e de Ecologia da Informação de um ambiente informacional únicos.

Diante do exposto, e considerando os objetivos desta tese e as abordagens apontadas por Oliveira (2014), as obras dos precursores da Arquitetura da Informação são revisitadas e apresentadas nas seções 3.1 a 3.4.

3.1 ARQUITETURA DA INFORMAÇÃO: ABORDAGEM ARQUITETURAL

A gênese da Arquitetura da Informação, como compreendida na Ciência da Informação, se relaciona com a Arquitetura e o Design e tem em Wurman seu precursor.

O aporte das áreas de Arquitetura e Design para a Arquitetura da Informação é apresentado por Oliveira (2014, p. 86) “No paradigma arquitetural, o campo do design se desdobra para influenciar os estudos e práticas da Arquitetura da Informação por meio do Design de Informação, do Design de Interação, do Design Gráfico, entre outros.”

Enquanto precursor, Wurman, um arquiteto, trouxe da Arquitetura e do Design Urbano elementos estruturantes para a Arquitetura da Informação, definindo-a como “ciência e a arte de criar instruções para espaços organizados” (WURMAN, 1996). Para Oliveira (2014), “A Arquitetura é um campo devotado à racionalização dos espaços em função do uso que a sociedade ou os sujeitos lhe atribuirão.”

Afirma Ronda León (2008, tradução nossa) que a atuação profissional de Wurman enfocou “[...] o design da informação em espaços urbanos, com ênfase nos processos de organização da informação, como passos preliminares para tornar a informação visivelmente compreensível para os usuários.”³⁹ Nesse sentido, Wurman declara como sua paixão “fazer a informação compreensível”.

³⁹ “[...] el diseño de información en los entornos urbanos, haciendo hincapié en los procesos de organización de la información, como pasos previos para hacerla información visiblemente comprensible para los usuarios.”

Resmini e Rosati (2011) atrelam a Wurman o que designam como abordagem do design de informação, cujas estruturas estão fortemente ligadas à arquitetura. Por isso, consideram que “[...] espera-se que os arquitetos da informação desenhem linhas e obtenham algum tipo de ordem no espaço de dados para melhor compreensão e diversão, sendo sua principal tarefa tornar essas informações mais simples, mais diretas e mais compreensíveis.”⁴⁰ (RESMINI; ROSATI, 2011, p. 23, tradução nossa).

Para tratar a Arquitetura da Informação enquanto abordagem arquitetural, recorre-se a Wurman, cuja publicação do texto intitulado *Beyond Graphics: The Architecture of Information*, em parceria com Katz, é reconhecida como o marco inicial da Arquitetura da Informação como compreendida na Ciência da Informação.

O referido texto foi publicado em outubro de 1975 no *AIA Journal*, também conhecido como *ARCHITECT: The Journal of The American Institute of Architects*, um periódico vinculado ao *American Institute of Architects*, cujo número teve como temática *Environmental Graphics* (Gráficos Ambientais), à qual, o tema e o título do texto de Wurman e Katz estão relacionados.

Nesta edição do AIA JOURNAL você viu uma ampla gama de gráficos ambientais usados para direcionar, informar, entreter, encantar. No entanto, apesar de toda a sua diversidade, os gráficos ambientais são apenas uma parte das técnicas, mídias e programas ilimitados que podem tornar a cidade compreensível para todos nós que vivemos nela e a usamos. O objetivo de todas as formas de exibir e comunicar informações é um ambiente informativo e atraente, interagindo com cidadãos informados e autoinformados.⁴¹ (WURMAN; KATZ, 1975, p. 40, tradução nossa).

Tal afirmação esclarece o teor do texto de Wurman e Katz (1975), que abordam a forma como a informação deve ser disponibilizada para que seja perceptível, funcional e útil para que as pessoas possam se localizar nas cidades.

Nesse sentido, para Wurman e Katz (1975, p. 40, tradução nossa),

A arquitetura da informação aplica-se não apenas a gráficos de grande escala aplicados ou integrados a estruturas no ambiente feito pelo homem, mas também aos meios de exibição de informações que tornam esse ambiente compreensível, interessante e navegável. Ela

⁴⁰ “[...] information architects were expected to draw lines and derive some kind of order in data space for better understanding and enjoyment, their primary task being to make this information simpler, more direct, and more comprehensible.

⁴¹ “In this issue of the AIA JOURNAL you have seen a wide range of environmental graphics used to direct, to inform, to entertain, to delight. Yet, for all their diversity, environmental graphics are but a part of the limitless techniques, media and programs that can make the city understandable to all of us who live in it and use it. The goal of all the ways of displaying and communicating information: is an informative and attractive environment interacting with informed, self-informing citizens.”

decorre do reconhecimento das várias maneiras pelas quais o ambiente pode se tornar auto-revelador.⁴²

Os autores empregam o termo Arquitetura da Informação ao discutirem a disposição e legibilidade da sinalização, a comunicação visual, o uso de mapas e diagramas, as experiências sensoriais, os ambientes e espaços como centros de informação, a poluição visual, temáticas relacionadas à arquitetura urbana, relacionando-a à informação como forma de tornar o ambiente urbano mais compreensível. Os autores entendem que “A extensão lógica de tais ideias, tanto em termos de concepção quanto de implementação, é o ambiente como arte da informação.”⁴³ (WURMAN; KATZ, 1975, p. 60, tradução nossa).

Pode-se depreender do texto perspectivas arquitetônicas aplicáveis aos ambientes informacionais, sejam eles analógicos, digitais ou híbridos, incluindo a necessidade de integração entre os diferentes elementos que compõem a Arquitetura da Informação na perspectiva da Ciência da Informação. Compreende-se, por meio da afirmação de Wurman e Katz (1975, p. 60, tradução nossa), que é possível aproximar a AI conforme compreendida por eles dos pressupostos teóricos da Ciência da Informação, nos paradigmas físico, cognitivo, social (CAPURRO, 2003) e pós-custodial (MALHEIRO; RIBEIRO, 2011).

É claro que grande parte de nossa preocupação com os meios de exibir informações está enraizada em nossa consciência de que muitas informações básicas sobre nossos ambientes criados pelo homem não existem quando e onde precisamos delas. Enquanto o ambiente inteiramente autorrevelador é uma ideia tão utópica quanto o edifício inteiramente autorrevelador ou o painel de instrumentos sem instruções, como arquitetos e designers temos uma oportunidade e responsabilidade especiais de tornar o ambiente mais compreensível e autorrevelador, por meio de um projeto mais funcional e autorrevelador.⁴⁴

No que tange ao paradigma físico, para o qual um objeto físico deve ser transmitido do emissor ao receptor por meio de um canal, as TIC e a Recuperação da

⁴² “The architecture of information applies not only to large-scale graphics applied to or integrated with structures in the manmade environment, but also the means of information display which make that environment understandable, interesting and navigable. It follows from the recognition of the various ways in which the environment can become self-revealing.”

⁴³ “The logical extension of such ideas, in terms of both conception and implementation, is the environment as information art.”

⁴⁴ “Of course, much of our concern with the means of displaying information is rooted in our awareness that so much basic information about our man-made environment does not exist when and where we need it. While the entirely self-revealing environment is as utopian an idea as the entirely self-revealing building or instruction less instrument panel, as architects and designers we have a special opportunity and responsibility to make environment more comprehensible and self-revealing, through more functional and self-revealing design.”

Informação são focos de pesquisa. É possível vislumbrar os diferentes canais, mídias e tecnologias aplicáveis aos ambientes informacionais analógicos, digitais e híbridos, de forma estruturada visando exibir, transmitir, comunicar informações de forma adequada, considerando aspectos de navegabilidade, acessibilidade e usabilidade, visando favorecer emissor e receptor e, com isso, tornar as informações localizáveis e acessíveis.

No paradigma cognitivo, o usuário, enquanto sujeito cognoscente, no contato com a informação, utiliza seus conhecimentos e modelos mentais para a criação de conhecimento. Wurman e Katz (1975) abordam questões como legibilidade, posicionamento, escala, orientação, navegação, percepção, experiência, sensações e aprendizado com a interação. Nesse sentido, é possível vislumbrar características da estrutura dos ambientes informacionais analógicos, digitais e híbridos, no que tange à organização e representação da informação, rotulagem, navegação e busca, visando favorecer a Encontrabilidade da Informação e a Experiência do Usuário⁴⁵, sobretudo no cenário atual de grande disponibilidade de informação, em que o sujeito consome e fornece informações de forma autônoma nos diferentes ambientes, o que torna necessário que os processos sejam compreensíveis e intuitivos.

No paradigma social, os aspectos e interesses sociais são considerados no processo de transmissão da informação. Diante disso, é imprescindível que a facilidade de acesso, compreensão e uso da informação, independente do seu suporte, seja uma questão preponderante em qualquer ambiente informacional, para que os sujeitos e suas demandas sejam atendidos de forma rápida e eficiente. Considera-se ainda o papel preponderante do usuário na Arquitetura da Informação.

E, no paradigma pós-custodial, a ênfase é dada ao acesso à informação, e as TIC são apresentadas como um dos focos de estudo da Ciência da Informação. Nesse sentido, é possível vislumbrar o papel dos profissionais envolvidos na Arquitetura da Informação para o tratamento adequado do conteúdo, considerando contexto e

⁴⁵ “[...] percepções e respostas das pessoas, resultantes do uso e/ou uso antecipado de um produto, sistema ou serviço. Nota 1: A experiência do usuário inclui todas as emoções, crenças, preferências, percepções, respostas físicas e psicológicas, comportamentos e realizações do usuário que ocorrem antes, durante e após o uso. Nota 2: A experiência do Usuário é uma consequência da imagem da marca, da apresentação, da funcionalidade, do desempenho do sistema, do comportamento interativo e das capacidades assistivas do sistema interativo, dos estados interno e físico do usuário, resultantes de experiências anteriores, atitudes, habilidades e personalidade e do contexto de uso. Nota 3: A usabilidade, quando interpretada a partir da perspectiva dos objetivos pessoais do usuário, pode incluir o tipo de aspectos perceptivos e emocionais tipicamente associados à experiência do usuário. Os critérios de usabilidade podem ser usados para avaliar aspectos da experiência do usuário.” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2011).

usuário, para que a informação seja acessível e compreensível pelo usuário. Além disso, o ambiente informacional deve ser intuitivo e, com isso, facilitar o acesso e o uso pelos sujeitos.

Em uma visão macro, entende-se que os ambientes urbanos podem ser considerados como ambientes informacionais, uma vez que é possível acessar e disponibilizar informações nesse contexto, e essas informações fluem e se espalham utilizando diferentes tecnologias, mídias e canais com os quais os usuários interagem para diferentes necessidades cotidianas.

Durante uma palestra, Wurman (2010, tradução nossa) discorre sobre algumas de suas experiências e interações e, aos poucos, revela sua compreensão acerca da Arquitetura da Informação, que parece simples, mas cuja aparência esconde complexidade. O autor afirma: “Queremos tornar algo compreensível para outro ser humano, para nós mesmos. [...] Algo que não entendo, quero tornar compreensível.”⁴⁶

É importante mencionar ainda a primeira Lei de Wurman “Você só entende algo relativo a algo que você entende.”⁴⁷ (WURMAN, 2010, tradução nossa). De lá apreende-se que a compreensão pode ocorrer por comparação, metáforas, analogias e pelo acionamento do conhecimento prévio, o que reforça a presença da tríade – conteúdo, contexto e usuário – para o planejamento adequado da Arquitetura da Informação de ambientes informacionais e sistemas de informação. Neste contexto, os sistemas propostos por Morville e Rosenfeld (2006) possuem papel preponderante na estruturação da informação e do ambiente informacional/sistema de informação, para que o usuário tenha facilidade na compreensão dessas estruturas e, com isso, torne o uso parte da sua estrutura cognitiva.

Embora Wurman seja reconhecido pelo uso do termo Arquitetura da Informação em 1975, durante a palestra ele afirma que já em 1962 trabalhava com AI, referindo-se à elaboração de um livro contendo a planta de 50 cidades em escala. Nesse contexto, atrela a AI à forma sistêmica de compreender e disponibilizar informação. E complementa:

A propósito, não havia computadores naquela época. Não havia wireframes. Não havia páginas da web. Não havia nenhuma dessas coisas. Isso não faz arquitetura da informação. O que faz um arquiteto

⁴⁶ “We want to make something understandable to another human being, to ourselves. [...] Something I don't understand, I want to make it understandable.”

⁴⁷ “You only understand something relative to something you understand.”

da informação é uma atitude. Um desejo, uma paixão para se comunicar sistemicamente com regras e sistemas e transferir informações para outro ser humano.⁴⁸ (WURMAN, 2010, tradução nossa).

Assim, é necessário destacar que a AI, na perspectiva de Wurman, não está inicialmente atrelada aos websites, embora sua popularização tenha ocorrido com a obra de Rosenfeld e Morville. Ao contrário, na perspectiva de Wurman, a AI tem como elementos centrais o sujeito e a informação, sendo que, para a compreensão, a informação necessita de sistematização, organização e apresentação.

E, sendo Wurman oriundo da área de Arquitetura, o embasamento da AI pode ter considerado os princípios estabelecidos na Tríade Vitruviana – *firmitas utilitas, venustas* – tidas respectivamente, como “solidez, firmeza, consistência, robustez”; “utilidade, uso, funcionalidade, proveito, vantagem”; “beleza, elegância, estética” (VITRUVIUS POLLIO, 2007, p. 82) –, que atrela a construção à Arquitetura quando ela é bem estruturada, possui e atende a uma função, e é bela.

Além disso, aspectos de organização e tratamento da informação são abordados em Wurman (2010), referindo-se a estruturas analógicas e digitais. Sua abordagem da organização da informação a coloca em destaque, uma vez que a relaciona com a forma como se pensa, e afirma que a forma como a informação é organizada pode criar uma nova informação, uma nova forma de acesso. Ao passo que, no que concerne ao tratamento da informação, o autor exemplifica com o uso de variantes de nomes no processo busca, cujo impacto está na recuperação da informação e na divergência de registros, dividindo os conteúdos recuperados nas formas variantes. Com isso, é notório o papel fundamental da representação da informação no campo de valor e de instrumentos de padronização desta representação, visando unificar elementos e favorecer a busca. Nos ambientes informacionais digitais isso se faz, utilizando metadados, mantendo consistência na estrutura e na representação.

Todo esse aparato trazido por Wurman permite reflexões, não apenas estruturais, mas também informacionais, no que se refere ao fluxo, organização,

⁴⁸ “There were no computers then, by the way. There was no wireframes. There was no web pages. There was none of those things. That doesn’t make information architecture. What makes an information architect is an attitude. A desire, a passion to communicate systemically with rules and systems, and transfer information to another human being.”

representação, acesso e uso, com foco primordial no sujeito, para o qual o ambiente e a informação devem ser compreensíveis.

A este contexto, nesta tese, considera-se que o consumo de informação por humanos e máquinas pode ser beneficiado pela estruturação da camada de dados dos ambientes informacionais digitais, o que ocorre por meio da Arquitetura de Dados. Esse processo torna os dados mais compreensíveis e processáveis por estruturas computacionais que são utilizadas em benefício humano.

3.2 ARQUITETURA DA INFORMAÇÃO: ABORDAGEM SISTÊMICA

Na abordagem chamada Arquitetura da Informação Sistêmica, há influência da Teoria Geral dos Sistemas, para a qual os ambientes informacionais são constituídos por um conjunto estruturado de sistemas ou partes interdependentes que compõem o todo.

Resmini e Rosati (2011) designam essa abordagem como sistemas de informação, por estar atrelada ao desenvolvimento de sistemas de informação e informática de negócios, e pouco conectada à experiência do usuário. Contudo, os autores afirmam que a obra de Rosenfeld e Morville (1998) alterou o foco da Arquitetura da Informação daquela época, até então fortemente ligada à tecnologia da informação, para uma visão voltada ao usuário.

Vale destacar que Rosenfeld e Morville, oriundos da área de Ciência da Informação, atuaram no início da web com o desenvolvimento de websites, e utilizavam o termo “arquitetura da informação” como uma metáfora para reforçar a importância da estrutura e da organização nesse processo. Isso pode explicar os sistemas que compõem a anatomia da Arquitetura da Informação na perspectiva de Rosenfeld e Morville (1998).

Na Arquitetura da Informação Sistêmica, os precursores (ROSENFELD; MORVILLE, 1998) estabelecem quatro sistemas que alicerçam a AI de um ambiente informacional: sistema de organização, sistema de navegação, sistema de rotulagem e sistema de busca. Segundo os autores, os três primeiros sistemas são responsáveis pela interação do usuário com o ambiente informacional, enquanto o último estrutura o funcionamento dos mecanismos de busca. Posteriormente, em Morville e Rosenfeld (2006), aos sistemas são incorporados tesouros, vocabulário controlado e metadados,

agrupados na anatomia da Arquitetura da Informação, contudo sem designação de sistema, como ocorre com os demais.

Para Rosenfeld e Morville (1998), um dos desafios da AI é tentar controlar os conteúdos e as funcionalidades do ambiente informacional. Por isso, reforçam que as decisões e requisitos podem necessitar de muitas iterações.

De acordo com Resmini e Rosati (2011), Wurman tinha a Arquitetura da Informação como o design das páginas de um website, enquanto Rosenfeld e Morville a percebiam como a estrutura, os links e as conexões entre as páginas por meio do conteúdo.

Dessa forma, inicialmente foram usados os sistemas de organização, rotulagem, navegação e busca e, posteriormente, os metadados, vocabulários controlados e tesouros, como formas de estruturar o conteúdo em um website. Formas, essas, que se aplicam a outros ambientes informacionais, sejam eles analógicos, digitais ou híbridos.

Para Rosenfeld, Morville e Arango (2015, p. 269)

Um ambiente de informação interativo [...] é uma coleção de sistemas interconectados com dependências complexas. Um único link em uma página pode fazer parte simultaneamente da estrutura, organização, rotulagem, navegação e sistema de busca do site. É útil estudar esses sistemas de forma independente, mas também é crucial considerar como eles interagem. O reducionismo não nos dirá toda a verdade.⁴⁹

Desta forma, os referidos sistemas são abordados em seções separadas, contudo, mantendo suas interconexões.

3.2.1 Sistema de Organização

“Um dicionário, ou uma enciclopédia, ou muitas das coleções de dados em nosso mundo, baseiam-se na capacidade de encontrar algo. A capacidade de encontrar algo anda de mãos dadas com a forma como é organizado.”⁵⁰ (WURMAN, 1996, p. 16, tradução nossa). Enfatiza o autor que há cinco diferentes formas de organizar a informação: por localização, por alfabeto, por tempo, por categoria, por

⁴⁹ “An interactive information environment—like a website—is a collection of interconnected systems with complex dependencies. A single link on a page can simultaneously be part of the site’s structure, organization, labeling, navigation, and searching systems. It’s useful to study these systems independently, but it’s also crucial to consider how they interact. Reductionism will not tell us the whole truth.”

⁵⁰ “A dictionary, or an encyclopedia, or many of the collections of data in our world, are based on being able to find something. The ability to find something goes hand-in-hand with how well it’s organized.”

hierarquia, e que a escolha é contextual. Contudo, esclarece que “A escolha primária de como você organiza algo é feita ao decidir como você deseja que seja encontrado.”⁵¹ (WURMAN, 1996, p. 17, tradução nossa). Além disso, para Rosenfeld e Morville (1998), deve ser definida a forma de apresentação da informação aos usuários.

Como arquitetos de informação, organizamos informações para que as pessoas possam encontrar as respostas certas para suas perguntas. Nós nos esforçamos para oferecer suporte à navegação casual e à pesquisa direcionada. Nosso objetivo é aplicar sistemas de organização e rotulagem que façam sentido para os usuários.⁵² (ROSENFELD; MORVILLE, 1998, p. 20, tradução nossa).

Desta forma, é importante atentar-se ao objetivo da organização da informação, que consiste em torná-la localizável, o que requer estruturação da informação, para que ela seja amplamente compreensível e, com isso, utilizada. Essa é uma das preocupações da AI, equilibrar interesses de fornecedores e consumidores de informação, utilizando-se das tecnologias existentes.

Em geral, na Arquitetura da Informação, a preocupação ocorre com os humanos, caracterizados como usuários, parte da tríade que alicerça a atuação da AI: conteúdo, contexto, usuário. Há de se considerar que o usuário é entendido essencialmente como uma figura humana, cujo conteúdo e contexto podem ser moldados considerando sua estrutura cognitiva e, por isso, caracterizado por Vechiato (2013), no contexto dos estudos de Encontrabilidade da Informação, como sujeito cognoscente. Contudo, com o avanço das técnicas e tecnologias, o usuário pode também passar a ser caracterizado por uma figura maquínica, a exemplo de aplicativos e aplicações computacionais, que podem ser utilizados, inclusive, com técnicas de Inteligência Artificial.

Neste contexto, o consumo da informação por máquinas depende de uma estruturação formal dos conteúdos e contextos para que sejam compreensíveis e processados com a semântica adequada, que possibilitará operações sintáticas. Tal fato pode trazer avanços significativos para o processo de estruturação de ambientes conectados, não apenas na navegação, mas também nas estruturas de dados.

⁵¹ “The primary choice of which way you organize something is made by deciding how you want it to be found.”

⁵² “As information architects, we organize information so that people can find the right answers to their questions. We strive to support casual browsing and directed searching. Our aim is to apply organization and labeling systems that make sense to users.”

Rosenfeld e Morville (1998) destacam que os ambientes web possibilitam que diferentes formas de organização sejam aplicadas aos conteúdos, o que os torna mais flexíveis e menos presos a estruturas formais de apresentação.

Quando se trata de ambientes compostos por consumo de dados de múltiplas fontes, essa flexibilidade pode ser otimizada a partir de dados adequadamente estruturados em seus ambientes de origem. Da mesma forma, a flexibilidade pode ocorrer com uso de Inteligência Artificial aplicada a ambientes moldados por perfis de usuários. Isso pode se constituir em uma tendência para a programação e para o desenvolvimento de ambientes informacionais digitais, mais flexíveis e capazes de oferecer experiências personalizadas aos usuários. Para tanto, torna-se imprescindível que os conteúdos sejam estruturados a partir de vocabulários padronizados que permitam semântica formal⁵³. Tal estrutura pode ser realizada com bancos de dados que

[...] permitem que você reutilize o mesmo conteúdo em várias formas e formatos para diferentes públicos. [...] No entanto, o modelo de banco de dados tem limitações. Os registros devem seguir regras rígidas. Em um determinado tipo de registro, cada registro deve ter os mesmos campos e, dentro de cada campo, as regras de formatação devem ser aplicadas de forma consistente nos registros.⁵⁴ (ROSENFELD; MORVILLE, 1998, p. 36, tradução nossa).

Neste contexto, a Arquitetura de Dados pode contribuir com a modelagem dos processos de conversão de registros entre diferentes sistemas para que os dados, ainda que possuam uma estrutura padronizada na base de dados de origem, possam ser convertidos para a estrutura e atender aos padrões da base de dados de destino. Destaca-se, neste sentido, que é importante manter nos metadados a proveniência dos dados e elementos da conversão, para que o processo seja compreensível a qualquer tempo.

A organização da informação é tratada no campo da Biblioteconomia e Ciência da Informação desde a sua gênese, com o objetivo de tornar os documentos compreensíveis e facilmente acessíveis. Neste sentido, os preceitos eram aplicados a ambientes controlados, como catálogos de bibliotecas. Com o advento da web e a

⁵³ “A semântica formal auxilia a estabelecer especificações precisas que transmitam significados detalhados e, por um lado, utilizar um vocabulário complexo (ontologia), pode servir como base para tarefas tais como o raciocínio automatizado.” (LÓSCIO; BURLE; CALEGARI, 2017).

⁵⁴ “[...] enable you to repurpose the same content in multiple forms and formats for different audiences. [...] However, the database model has limitations. The records must follow rigid rules. Within a particular record type, each record must have the same fields, and within each field, the formatting rules must be applied consistently across records.”

facilidade de disponibilização de conteúdos, a tarefa de organização da informação tornou-se corriqueira a qualquer indivíduo que atue no desenvolvimento de ambientes informacionais. Contudo, os responsáveis pela atividade passaram a ser indivíduos especialistas em outras áreas que, pelo fato de não serem profissionais da informação, podem desconhecer os processos de organização da informação, prejudicando a localização da informação e o uso de tais ambientes.

Estamos caminhando em direção a um mundo onde um grande número de pessoas publica e organiza suas próprias informações. À medida que fazemos isso, os desafios inerentes à organização dessas informações tornam-se mais reconhecidos e mais importantes.⁵⁵ (ROSENFELD; MORVILLE, 1998, p. 20, tradução nossa).

No processo de organização da informação, deve-se entender a relevância dos termos empregados, visto que os conteúdos são localizados pelos usuários a partir dos rótulos utilizados para descrevê-los. A complexidade deste processo consiste na subjetividade do emprego dos rótulos pelos fornecedores e da interpretação pelos consumidores que, quando humanos, aplicam interpretações próprias dos seus conhecimentos prévios e, quando máquinas, utilizam-se da programação utilizada. Diante disso, justifica-se a necessidade do uso de vocabulários padronizados e estruturação adequada dos dados para favorecer a qualidade do fornecimento e consumo de dados.

Para Rosenfeld e Morville (1998), a organização e a rotulagem não devem levar em conta os modelos mentais dos envolvidos no projeto e desenvolvimento do ambiente informacional. Ao contrário, devem ser realizadas considerando os usuários do ambiente, o que amplia o escopo de análise.

Nesse sentido e considerando a relevância da organização da informação, Rosenfeld e Morville (1998, p. 58, tradução nossa) destacam que “Os sistemas de organização são compostos de esquemas de organização e estruturas de organização. Um esquema de organização define as características compartilhadas dos itens de conteúdo e influencia o agrupamento lógico desses itens. Uma estrutura de organização define os tipos de relacionamento entre itens e grupos de conteúdo.”⁵⁶

⁵⁵ “We’re moving towards a world where tremendous numbers of people publish and organize their own information. As we do so, the challenges inherent in organizing that information become more recognized and more important.”

⁵⁶ “Organization systems are composed of organization schemes and organization structures. An organization scheme defines the shared characteristics of content items and influences the logical grouping of those items. An organization structure defines the types of relationships between content items and groups.”

Nesse sentido, os autores afirmam que o sistema de organização está intimamente relacionado aos outros sistemas da AI, como sistemas de rotulagem, navegação e busca, por meio da indexação. Por isso, enfatizam a necessidade do planejamento adequado do sistema de organização, que será a forma de navegação primária do ambiente e influenciará todos os demais sistemas.

Morville e Rosenfeld (2006, p. 66, tradução nossa) discutem sobre diferentes esquemas de organização da informação e afirmam que “O poder de um esquema de organização puro deriva de sua capacidade de sugerir um modelo mental simples que os usuários possam entender rapidamente.”⁵⁷. Infere-se que este modelo mental possa contribuir com a compreensão do usuário no uso do ambiente.

Conforme já sinalizamos nesse texto, Wurman (1991) afirma que “Você só aprende o que tem relação com o que compreende”. Tal afirmação destaca a importância do conhecimento prévio, das analogias e de conexões na estruturação da informação e do ambiente informacional, além da manutenção de estruturas conhecidas pelo usuário, ainda que diante de alguma alteração no ambiente informacional.

Por outro lado, Wurman (1991) também considera que “O que você não sabe é tão importante quanto o que sabe”, afirmação que abre a possibilidade de conduzir o usuário na compreensão durante o uso de um ambiente informacional ou de um sistema de informação, o que exige que sejam amigáveis e intuitivos.

Destaca-se, ainda, a forte relação entre a organização da informação e a navegação, sobretudo na manutenção da orientação do usuário no uso de um ambiente informacional ou em uma Ecologia Informacional Complexa. É possível afirmar que o sistema de organização determinará a forma de navegação, a estrutura das trilhas de navegação (*wayfinding*) e a possibilidade de crescimento ordenado do ambiente informacional e/ou da Ecologia Informacional Complexa como projetado.

Ao tratar o sistema de organização na AI, Rosenfeld e Morville (1998), no que tange à estrutura, abordam a organização da informação por meio de bancos de dados, exemplificando a aplicação com banco de dados relacional, e acrescentam que pode ser necessário o uso de vocabulários controlados, visando favorecer o entendimento dos consumidores no uso do ambiente. Contudo, referem-se a vocabulário controlado como termos utilizados e notas que os definem, cuja

⁵⁷ “The power of a pure organization scheme derives from its ability to suggest a simple mental model that users can quickly understand.”

característica é o emprego no campo de valor, enquanto para a estruturação dos dados, os vocabulários controlados, apesar de serem oportunos para a representação do valor do dado, são imprescindíveis no campo de atributo.

Na obra de Morville e Rosenfeld (2006), os bancos de dados são novamente abordados, definidos como “[...] uma coleção de dados organizados para facilidade e velocidade de busca e recuperação.”⁵⁸. Os autores utilizam como exemplo os bancos de dados relacionais. Porém, nesta obra, embora reforcem que o foco é o acesso à informação e não a recuperação dos dados, os autores reconhecem que há uma ligação entre a Arquitetura da Informação e os bancos de dados, o que ocorre por meio de metadados, o que possibilita aplicar estrutura e relacionamentos a ambientes heterogêneos e não estruturados.

Nesse sentido, em ambientes informacionais digitais, há relação direta com os formatos de dados e metadados empregados no processo de organização e representação da informação. Cabe mencionar que metadados são utilizados para representar entidades⁵⁹ do mundo real. Para tanto, são estruturados por meio de atributos, que consistem em uma *tag* ou *metatag* utilizadas como elemento específico sobre cada característica de uma entidade, que disponibiliza um campo de valor para representar a característica específica da entidade representada.

No que tange ao uso de metadados, tanto os atributos quanto os valores podem utilizar vocabulários controlados, que consistem em uma linguagem artificial constituída por termos, de forma relacional, e cuja função é otimizar o fornecimento e o consumo de dados em um sistema de informação.

Assim, no caso dos atributos, os vocabulários controlados podem ser esquemas de metadados, sobretudo aqueles padronizados internacionalmente, visando possibilitar o intercâmbio de dados entre máquinas – por exemplo, *Machine Readable Cataloging* (MARC), *Dublin Core* (DC), *Metadata Object Description Schema* (MODS), *Metadata Encoding and Transmission Standard* (METS) e *PREservation Metadata: Implementation Strategies* (PREMIS) – ou ainda criados e gerenciados por uma organização – o que requer análise e definição de requisitos de

⁵⁸ “[...] a collection of data arranged for ease and speed of search and retrieval.”

⁵⁹ Entendidas de acordo com a ISO/IEC 11179-1 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015a, p. 3, tradução nossa) como “qualquer coisa concreta ou abstrata que exista, existiu ou possa existir, incluindo associações entre essas coisas”. Ainda, de acordo com a referida norma, “Uma entidade existe independentemente de os dados sobre ela estarem disponíveis ou não.”

informação necessários, existentes ou não –, e utilizam codificação sintática, regras, formatos e linguagem que permitam o processamento por máquinas.

Nesse aspecto, o uso de esquema de metadados beneficia a comunicação de dados entre diferentes sistemas de informação, enquanto aqueles criados especificamente para uma instituição, no caso da comunicação entre sistemas, podem necessitar de compatibilização.

Por outro lado, os vocabulários controlados para campos de valor consistem em linguagens documentárias, tesouros, ontologias ou outras estruturas utilizadas para padronizar a forma gramatical e indicar sinônimos e antônimos dos termos empregados, sobretudo na indexação, visando melhorar a recuperação da informação em bases de dados.

Ainda, Rosenfeld e Morville (1998) afirmam que o uso de vocabulário controlado fornece suporte à pesquisa, navegação e criação automática de índices. Já Morville e Rosenfeld (2006) atrelam os vocabulários controlados à pesquisa, à navegação, aos refinamentos e links dinâmicos.

Na perspectiva dos dados, o uso de vocabulário controlado possibilita a semântica formal para a estrutura, a localização, a compreensão, o consumo, o processamento e o (re)uso dos dados.

Assim, na perspectiva dos metadados, criar e manter vocabulários controlados para o campo de valor é tarefa complexa e morosa, enquanto para o campo de atributo, é possível utilizar esquemas e padrões já existentes ou, ainda, criá-los, o que torna necessário fornecer a documentação para o consumo. Nesse contexto, vale destacar que Brandt (2020) apresenta uma metodologia para a modelagem da informação, com enfoque nos metadados de negócios, que pode ser utilizada para a criação e gestão de atributos.

A relevância dos metadados e vocabulários controlados é apontada por Morville e Rosenfeld (2006) como conhecimento necessário aos arquitetos da informação, a fim de compreender como podem ser utilizados nos ambientes informacionais, por exemplo, na geração automática de índices, no uso de links dinâmicos e no estabelecimento de filtros para resultados de busca.

No que tange ao uso de metadados, enquanto campo de valor, há anos é notável a participação de usuários humanos na criação de tags, no processo de indexação e navegação social e, no uso de folksonomias, o que consiste em um processo de cocriação de valor entre o ambiente informacional digital e os usuários e

que contribui de forma ativa com os ambientes informacionais, na adequação dos sistemas de organização, navegação, rotulagem e na criação de índices que beneficiem os sistemas de navegação e de busca.

É notável a imprescindível aproximação das áreas da Ciência da Informação, Ciência da Computação e Sistemas de Informação, visando atuar adequadamente na estrutura das diferentes camadas de sistemas de informação e ambientes informacionais digitais, quer seja na Arquitetura da Informação *top-down* ou *bottom-up*, e até mesmo na Arquitetura de Dados, o que contribui para os processos de localização, acesso, interoperabilidade e (re)uso dos dados.

Com isso, considerando que os ambientes informacionais digitais se constituem em uma teia hipertextual de informações, é possível que o hipertexto ganhe novos contornos, com uma espécie de hyperdata, já utilizada, por exemplo, no contexto da web no *linked data* (BERNERS-LEE, 2006), visando contudo a ligação de bases de dados de diferentes provedores, por meio de protocolos que possibilitem o reuso dos dados a partir de chaves de acesso, que favoreçam o reuso dos dados disponíveis em determinada base de dados, gerando integração e interoperabilidade. Assim, os ambientes informacionais digitais teriam maior aderência à hipermídia, por meio da ligação de diferentes formatos de objetos digitais. Desta forma, as redes podem se expandir, por meio de ambientes informacionais digitais estruturados, tornando o tráfego de dados mais fluido.

3.2.2 Sistema de Rotulagem

O sistema de rotulagem consiste nas formas de representação, por meio de etiquetas ou rótulos, das áreas, conteúdos e conceitos disponíveis no ambiente informacional, cujo objetivo é agrupar e comunicar informações em um espaço reduzido e, condensado. Para Rosenfeld e Morville (1998, p. 61, tradução nossa) “[...] rotular é uma consequência natural da criação de sistemas de organização e navegação [...]”⁶⁰, uma vez que “[...] os rótulos geralmente são as maneiras mais óbvias de mostrar claramente ao usuário [...] [a organização do ambiente informacional digital] e seus sistemas de navegação.”⁶¹

⁶⁰ “[...] labeling is a natural outgrowth of creating organization and navigation systems [...]”

⁶¹ “[...] labels are often the most obvious ways of clearly showing the user your organization and navigation systems.”

Assim como nos demais sistemas, para o sistema de rotulagem é imprescindível considerar como o usuário busca informações no ambiente informacional, para que os rótulos façam sentido. Isso consiste em um desafio para o desenvolvimento da AI. Rosenfeld e Morville (1998) alertam que os usuários não estão dispostos a despender muito tempo para interpretar os rótulos, o que reforça a necessidade da clareza e objetividade na representação. Além disso, segundo Morville e Rosenfeld (2006), os rótulos atuam no relacionamento entre usuários e conteúdos, elementos dinâmicos em um ambiente informacional.

Considerando que o sistema de rotulagem viabiliza a representação nos sistemas de organização e de navegação do ambiente informacional, para desenvolvê-lo é necessário que estes sistemas estejam adequadamente projetados. Além disso, é necessário que os rótulos sejam projetados como um sistema, de forma orgânica, mantenham o contexto e utilizem terminologia condizente com o público que o ambiente pretende atingir. “A rotulagem pobre e pouco profissional pode trair esse investimento e destruir a confiança do usuário em uma organização.”⁶² (ROSENFELD; MORVILLE, 1998, p. 62, tradução nossa).

Assim como no sistema de navegação, o sistema de rotulagem deve ser projetado de forma a manter a coerência dos termos durante a navegação, uma vez que isso auxilia o usuário a reconhecer sua localização e orientação no ambiente informacional. Além disso, alguns domínios, como a web, já possuem alguns termos utilizados por convenção que são facilmente interpretáveis pelos usuários. Seu uso beneficia a familiaridade dos usuários com o ambiente, uma vez que reduz o esforço cognitivo necessário à interpretação.

Ainda, no que tange à coerência, é imprescindível que o rótulo seja utilizado para representar apenas uma área, conteúdo ou conceito disponíveis no ambiente informacional. No que tange aos rótulos dos menus global e local, a repetição de rótulos pode dificultar a compreensão e a localização do usuário. Para auxiliar na compreensão dos rótulos, Rosenfeld e Morville (1998) recomendam que sejam acompanhados de pequenas descrições ou notas explicativas.

Além disso, Morville e Rosenfeld (2006) apontam a necessidade de manter a consistência na rotulagem – lembrando que os rótulos não devem ser pensados de forma independente e sim como um sistema –, uma vez que o processo de navegação

⁶² “*Poor, unprofessional labeling can betray that investment and destroy a user's confidence in an organization.*”

está diretamente atrelado à rotulagem e, com isso, devem auxiliar na manutenção da sequência necessária. “[...] os rótulos devem informar aos usuários por onde começar, para onde ir e que ação estará envolvida em cada passo ao longo do caminho.”⁶³ (MORVILLE; ROSENFELD, 2006, p. 92, tradução nossa). Assim, é imprescindível compreender como o sistema de rotulagem pode interferir no sistema de navegação, beneficiando-o ou ampliando a exposição negativa dos problemas existentes.

Para Rosenfeld e Morville (1998), os rótulos podem ser utilizados na indexação dos conteúdos, o que beneficia, além da navegação a recuperação do conteúdo em sistemas de busca. “Essa abordagem permite que os usuários cortem a granularidade do sistema de organização principal do site e naveguem pelo conteúdo de um modo completamente diferente.”⁶⁴ (ROSENFELD; MORVILLE, 1998, p. 66, tradução nossa).

Esse processo pode ser realizado de forma não visível ao usuário na interface. Assim, além do rótulo indicado de forma visível, é possível incluir, por exemplo, sinônimos e formas variantes em uma camada de programação ou banco de dados, visando ampliar as possibilidades de indexação e recuperação do conteúdo sem, contudo, sobrecarregar a interface ou a camada visível em primeiro plano.

Nesse sentido, considerando o custo – operacional do sistema e/ou cognitivo do usuário – que pode ser gerado por alterações na forma de apresentação dos rótulos no *front-end*, é possível que, a partir dos dados coletados nas interações dos consumidores nos sistemas de busca e de navegação, rótulos alternativos sejam disponibilizados no *back-end*, visando manter a camada visível ao usuário inalterada e, concomitantemente, otimizar a recuperação dos conteúdos com a inserção de múltiplos rótulos.

No que tange à adoção de termos, há a necessidade de manter a consistência do sistema de rotulagem, aplicando-a de forma eficaz e que beneficie a familiaridade do usuário com o ambiente informacional. Essa consistência pode ser utilizada ainda na padronização do projeto do ambiente informacional, por exemplo, no uso de tipos e tamanhos de fontes, cores, posicionamento, espaço em branco.

⁶³ “[...] the labels should tell users where to start, where to go next, and what action will be involved in each step along the way.”

⁶⁴ “This approach allows users to cut across the grain of the site's main organization system and browse the content in a completely different mode.”

Assim, a terminologia, quando pensada no *front-end*, é possível que seja baseada no uso de vocabulários controlados específicos do domínio. E, quando pensada no *back-end*, pode ser estruturada por meio de vocabulários padronizados de dados, como padrões e esquemas de metadados. O uso de vocabulários controlados pode apoiar, ainda, a indexação, visando otimizar os conteúdos para ferramentas, sistemas e motores de busca.

Vale destacar, ainda, com base em Morville e Rosenfeld (2006), que a rotulagem pode se constituir como termos de índice – utilizando inclusive vocabulários controlados – que podem ser invisíveis ao usuário, quando utilizados, por exemplo em elementos que não são apresentados na interface, mas podem ser acionados pela ferramenta de busca, uma vez que estão presentes no banco de dados ou, ainda, incluídos como *tags*. No que tange à recuperação da informação em mecanismos de busca, o uso adequado de termos de índices auxilia na descoberta e no ranqueamento do conteúdo na página de resultados, sendo um objeto de estudo da *Search Engine Optimization* (SEO)⁶⁵.

É possível ainda que haja um sistema de rotulagem secundário, constituído por palavras-chave que descrevem os conteúdos, ou ainda por hiperlinks que os relacionam a outros conteúdos, áreas ou ambientes informacionais. Nesse caso, a rotulagem é visível ao usuário na camada da interface do ambiente informacional.

Assim como no sistema de rotulagem primário ou navegacional, é imprescindível que o sistema de rotulagem secundário seja sistemático, que haja contexto e que a terminologia adotada seja adequada ao ambiente informacional e compreensível aos usuários.

Os sistemas de rotulagem, em geral, são textuais. Por isso, considera-se a relevância da terminologia. Contudo, é possível que sejam utilizados ícones para representar, sobretudo, os sistemas de organização e navegação. Para Rosenfeld e Morville (1998, p. 70, tradução nossa), “O problema com os rótulos icônicos é que eles constituem uma linguagem muito mais limitada do que o texto.”⁶⁶

Neste sentido, deve-se ter em mente que, embora haja algumas convenções quanto ao uso de ícones, sua compreensão pode ser mais complexa que a dos textos. Por outro lado, a utilização de ícones no sistema de rotulagem pode reforçar a

⁶⁵ *Search Engine Optimization* (SEO) consiste em um conjunto de estratégias e técnicas para melhorar o posicionamento de websites em mecanismos de busca (LEDFORD, 2009).

⁶⁶ “The problem with iconic labels is that they constitute a much more limited language than text.”

identidade visual de um ambiente. Para Rosenfeld, Morville e Arango (2015), os rótulos icônicos podem contribuir com os elementos estéticos do ambiente informacional, mas destacam que a forma não deve se sobrepor à função. Assim, tais rótulos não devem comprometer a usabilidade do ambiente. Ainda, os autores esclarecem que, em ambientes informacionais digitais projetados para *desktop*, a área da tela é maior, o que possibilita a prevalência de rótulos de texto que inegavelmente são mais claros, ao passo que, em dispositivos móveis, o espaço é menor e, com isso, os rótulos icônicos podem ser uma melhor opção.

No contexto da Encontrabilidade da Informação (VECHIATO; VIDOTTI, 2014), os ícones podem ser utilizados como *affordances*, cuja utilização é possível em qualquer projeto de Arquitetura da Informação, e os impactos perceptíveis na navegação, por meio do fornecimento de pistas e da orientação espacial (*wayfinding*) ou como *query* para a realização de buscas. Além disso, para os autores, as *affordances* podem se constituir em importante atributo para a usabilidade, utilizado na interface de sistemas de informação e ambientes informacionais.

O advento das tecnologias, das mídias e da própria interação dos usuários com ambientes informacionais digitais alterou o que Rosenfeld e Morville (1998) apontavam sobre a limitação dos rótulos icônicos, que passam a ser amplamente utilizados nos ambientes digitais, inclusive na comunicação interpessoal. Por outro lado, torna-se necessário destacar que, embora os ícones tenham se fortalecido no que tange à rotulagem por humanos, ainda se constituem em um desafio para a compreensão por máquinas. Neste contexto, há também desenvolvimento tecnológico atual, a exemplo das buscas por imagens, que recuperam imagens ou texto, o que deve se fortalecer ao longo dos anos.

Um dos principais desafios do sistema de rotulagem consiste na subjetividade de quem cria e interage com o rótulo. Como qualquer outro elemento da Arquitetura da Informação, a rotulagem deve considerar conteúdo, contexto e usuário. Contudo, os rótulos se utilizam da linguagem para representar informações e, com isso, tornam-se dependentes de interpretações, o que pode dificultar a generalização das convenções, sobretudo em ambientes informacionais digitais que possuem a característica do alcance global ou ainda aqueles com interfaces multilíngues. Além disso, para Morville e Rosenfeld (2006, p. 98, tradução nossa):

A linguagem é simplesmente muito ambígua para você se sentir confiante de que aperfeiçoou um rótulo. Sempre há sinônimos e

homônimos com os quais se preocupar, e contextos diferentes influenciam nossa compreensão do que significa um determinado termo.⁶⁷

Vale lembrar ainda que, para Wurman (1991), a ambiguidade pode gerar ansiedade da informação, e por isso o autor reforça que a informação deve reduzir a incerteza, portanto, deve ser compreensível ao usuário, afirmação essa claramente aplicável aos humanos.

Morville e Rosenfeld (2006) consideram que a ambiguidade pode ser reduzida se o escopo do ambiente informacional for focado, com público e domínio bem definidos. Caso isso não seja possível, recomendam a modularização, que pode resultar em sistemas de rotulagem para diferentes áreas do ambiente informacional.

Qualquer que seja a opção adotada para o sistema de rotulagem, ainda que consista na combinação de termos e ícones, é importante que seu planejamento seja exaustivo e antevêja necessidade de crescimento futuro. Nesse sentido, ainda que alguns rótulos não sejam implementados logo na disponibilização do ambiente, é importante que sejam parte do projeto, para que sejam coerentes e estejam no contexto do sistema proposto. Outro aspecto relevante é considerar a granularidade necessária para o sistema de rotulagem, que pode exigir uma estrutura menor, caso usada para navegação, ou maior, quando usada para conteúdo. A isso, Rosenfeld e Morville (1998) se referem como abrangente em um escopo definido.

3.2.3 Sistema de Navegação

Enquanto o sistema de organização é imprescindível para estruturar a informação em um ambiente informacional, o sistema de navegação é o responsável por permitir flexibilidade e mobilidade, considerando o contexto, as associações, as conexões, os interesses e as necessidades do usuário que definem os caminhos a serem trilhados no uso do ambiente e das informações disponíveis.

Para Rosenfeld e Morville (1998, p. 42, tradução nossa), “Embora um esquema de organização hierárquica bem projetado reduza a probabilidade de os usuários se perderem, um sistema de navegação complementar é frequentemente

⁶⁷ *“Language is simply too ambiguous for you to ever feel confident that you’ve perfected a label. There are always synonyms and homonyms to worry about, and different contexts influence our understanding of what a particular term means.”*

necessário para fornecer contexto e permitir maior flexibilidade de movimento dentro do site.”⁶⁸

Assim, de forma complementar ao sistema de organização, o sistema de navegação deve permitir flexibilidade, movimento e contexto ao usuário no ambiente informacional, contudo, sem sobrecarregá-lo com muitas opções de caminhos a serem trilhados.

Os sistemas de organização e navegação possuem a hierarquia como principal mecanismo ou estratégia para estruturação. Contudo, Rosenfeld e Morville (1998) consideram que esta pode ser uma característica limitante no uso do ambiente e, para tanto, é necessário que haja formas de navegação complementares à hierárquica, a fim de permitir a mobilidade do usuário. No entanto, reforçam os autores que a hierarquia será sempre a base do sistema de navegação.

Além disso, é necessário considerar que o ambiente informacional pode ser parte de outro ambiente. Nesse caso, deve-se levar em conta o sistema de navegação do ambiente macro, para que não sobreponha ou confunda o usuário durante a exploração do ambiente com o qual interage. Nesse sentido, manter a orientação do usuário durante a navegação assegura que ele seja capaz de se movimentar em diferentes sentidos e reconheça sua localização, atentando-se à necessidade de retorno aos diferentes pontos da navegação. Isso ocorre à medida que o sistema de navegação possibilita contextos dentro do todo que compõe o ambiente informacional. “[...] desenvolvemos ferramentas de navegação para evitar que as pessoas se percam.”⁶⁹ (ROSENFELD; MORVILLE, 1998, p. 42, tradução nossa).

Para Rosenfeld, Morville e Arango (2015), tornar claro o contexto do ambiente pode auxiliar na compreensão da informação, o que implica em utilizar linguagem e caminhos claros para a navegação no ambiente.

Morville e Rosenfeld (2006) destacam que o sistema de navegação possui elementos ou subsistemas. Dentre eles, há três principais que estão embutidos no ambiente informacional, designados como global, local e contextual. Cada um deles exerce um papel na navegação do usuário pelo ambiente.

⁶⁸ *“While a well-designed hierarchical organization scheme will reduce the likelihood that users will become lost, a complementary navigation system is often needed to provide context and to allow for greater flexibility of movement within the site.”*

⁶⁹ *“[...] we have developed navigation tools to prevent people from getting lost.”*

O menu global deve ser exibido em todas as páginas que compõem o ambiente informacional digital, uma vez que permite acesso direto às suas áreas principais; o menu local o complementa, uma vez que exhibe as especificidades das áreas que compõe o menu global; enquanto a navegação contextual apresenta os relacionamentos entre conteúdos, por meio de links presentes na página de conteúdo que direcionam para uma página, objeto ou documento, formando uma rede de conexões.

Além deles, tem-se o sistema de navegação suplementar, que consiste em páginas destinadas a exhibir maneiras complementares de acessar a informação, por exemplo, mapa do site e índices. Vale destacar que o sistema de busca pode ser utilizado para navegação suplementar, uma vez que o usuário pode realizar buscas por palavras ou frases e ser direcionado a áreas específicas do ambiente informacional.

Rosenfeld, Morville e Arango (2015) destacam que as camadas superficiais da navegação, nas quais os usuários interagem, são impactadas pelas tecnologias computacionais, de forma que podem necessitar de alteração em seu projeto, considerando as funcionalidades e o dispositivo de acesso. Assim, é possível que os sistemas de navegação se comportem de forma diferente considerando o dispositivo de acesso. Reforçam, contudo, que independentemente disso, devem ser considerados os propósitos do sistema de navegação, fornecer contexto e flexibilidade, visando auxiliar o usuário na orientação espacial e definição de caminhos para a navegação.

Ressalta-se que, em um ambiente informacional digital, a página principal tende a exhibir menus de navegação global, local e contextual, ao passo que, ao adentrar suas páginas secundárias, o menu de navegação local ganha destaque, uma vez que exhibe as especificidades da área que está sendo acessada. Com isso, pode indicar onde o usuário está, o que há nas proximidades e os relacionamentos. Isso pode ser realizado com o uso de *affordances* e *wayfinding*. Nas páginas secundárias, o menu global pode permanecer, contudo, com menor ênfase, apenas para que o usuário possa navegar daquele ponto a outros dentro do ambiente informacional.

De acordo com Morville e Rosenfeld (2006), é possível aplicar aos ambientes informacionais digitais a personalização ou a customização, sendo a primeira configurada a partir de comportamentos, preferências e necessidades potenciais de

determinado usuário ou grupo de usuários, e a segunda definida com o controle do usuário, sobre o conteúdo e as formas de apresentação.

Considerando a característica das obras de Rosenfeld e Morville (1998), Morville e Rosenfeld (2006), Rosenfeld, Morville e Arango (2015), que estão no domínio da web, a navegação em um ambiente informacional digital como os websites ocorre por meio de navegadores de internet que possuem sistemas de navegação próprios, disponíveis para o usuário e cujo funcionamento é independente do sistema de navegação do website. Nesse caso, o sistema de navegação do website deve ser independente do utilizado nos navegadores de internet, além de possibilitar que o usuário encontre caminhos para avançar e retroceder durante a navegação, sem deixar de reconhecer sua localização, ou seja, mantendo a orientação espacial.

Os próprios navegadores de internet possuem algumas funcionalidades que auxiliam o usuário na navegação, como: o acesso ao conteúdo por URL, histórico, favoritos e cores para links acessados ou não (MORVILLE; ROSENFELD, 2006). Já os aplicativos são independentes de navegadores de internet, o que possibilita utilizar convenções de navegação próprias e sem influências do ambiente de exibição de um website, por exemplo. Por outro lado, os aplicativos podem ter a navegação influenciada pelos sistemas operacionais, a exemplo da barra superior padrão no IOS para *desktop* (ROSENFELD; MORVILLE; ARANGO, 2015).

Deve-se considerar ainda que os websites e demais ambientes informacionais digitais disponibilizam links para outros websites e ambientes informacionais, que podem ser motores de busca que direcionam para outros websites e ambientes, e isso também consiste em navegação (MORVILLE; ROSENFELD, 2006). Assim, o usuário pode acessar uma informação a partir de um ambiente informacional externo ao que a armazena, desconsiderando, em princípio, a página principal e o sistema de navegação do ambiente que armazena a informação acessada. Essa flexibilidade na navegação, própria dos ambientes digitais, dificulta a manutenção do contexto e, por isso, é imprescindível que todas as páginas apresentem a localização da informação no ambiente, para auxiliar o usuário na definição dos seus caminhos durante o acesso. Em geral, isso é feito por meio da orientação espacial (*wayfinding*) e de trilhas de navegação (*breadcrumb*).

Vale destacar que as trilhas de navegação, também conhecidas como navegação estrutural, exibem a forma como o sistema de organização dispõe a informação, de forma a apresentar o contexto e o relacionamento entre as páginas e

os conteúdos em um ambiente macro. Essa informação é relevante para que haja clareza de que a trilha de navegação do usuário nem sempre é igual à trilha de navegação exibida pelo ambiente, justamente pelos inúmeros pontos de acesso secundários ao conteúdo, quer seja a partir de uma URL específica, de um outro ambiente informacional ou mesmo de um mecanismo de busca.

Assim, é imprescindível que, independentemente da forma de acesso do usuário à informação, quer seja pela navegação no próprio ambiente ou por acesso a partir de um ambiente externo, a trilha de navegação seja exibida, para que o usuário entenda os relacionamentos da informação que acessou com as demais informações disponíveis no ambiente informacional.

Tal afirmação reforça a dependência dos chamados sistemas da Arquitetura da Informação (ROSENFELD; MORVILLE, 1998; MORVILLE; ROSENFELD 2006), uma vez que o sistema de organização é espelhado na navegação, para a qual o funcionamento o utiliza.

Isso reforça a característica de flexibilidade e mobilidade do sistema de navegação, bem como a relevância da manutenção da orientação espacial, para que o usuário saiba reconhecer onde está na navegação, bem como definir o caminho a trilhar, quer seja para seguir ou retroceder.

A este respeito, conseguir determinar a trilha de navegação do usuário pode ser extremamente complexo, visto que ele pode tomar decisões que o levem a diferentes ambientes informacionais durante a navegação, o que dificulta possibilitar o caminho de volta ao primeiro ambiente informacional utilizado. Essa é uma característica da navegação hipertextual. Assim, destaca-se que, durante a busca por uma informação, caso o usuário tenha navegado e interagido entre diferentes ambientes informacionais, pode haver problemas na sua orientação espacial e na trilha de navegação, uma vez que, em geral, são projetadas para ambientes informacionais individuais e não de forma sistêmica ou ecológica.

De igual maneira, esta premissa pode ser considerada em outros ambientes, seja na web, em diferentes dispositivos, ou ainda em ambientes analógicos. Nesse sentido, é necessário considerar o entorno do ambiente informacional, que pode ser uma estrutura física e/ou computacional, que possui sistemas de navegação próprios a serem considerados ao estabelecer o sistema de navegação do ambiente estruturado.

Essas afirmações encontram reforço em Rosenfeld, Morville e Arango (2015, p. 180):

Ao projetar ambientes informacionais complexos, é particularmente importante fornecer contexto dentro do todo maior. Muitas pistas contextuais disponíveis no mundo físico não existem online. Não há marcos naturais, nem norte e sul. Ao contrário da viagem física, a navegação hipertextual permite que os usuários sejam transportados diretamente para o meio de um sistema desconhecido.⁷⁰

Este conjunto composto pelo ambiente informacional e seu entorno é conceituado por Oliveira (2014) como Ecologia Informacional Complexa, que consiste em uma estrutura composta por ambientes, tecnologias, canais e sujeitos, interligados e conectados entre si de forma sistêmica.

Wurman (1996) enfatiza a necessidade de tornar o complexo claro, ao adotar uma metáfora na representação da linha férrea do metrô de Tokyo, visando torná-la mais compreensível e memorizável, ao utilizar a figura do Yin e Yang, e como ponto de referência o Palácio Imperial. Nesta metáfora, Wurman (1996) destaca a importância da compreensão da essência do que se busca representar, ao indicar o caminho, a sequência de paradas e um ponto turístico amplamente conhecido. Para o usuário, o mapa real é secundário em relação à orientação da localização.

Por outro lado, embora Rosenfeld e Morville (1998) reconheçam que as metáforas são importantes para a compreensão e, por isso, muito úteis, consideram necessário ter cautela ao desenvolver a AI e sua orientação global norteada por metáforas, uma vez que são dependentes da familiaridade dos usuários do ambiente. Assim, deve-se reforçar que a equipe envolvida na estruturação de ambientes informacionais deve ir além dos seus modelos mentais e ter foco na compreensão dos consumidores. Essa atividade, apesar de exaustiva, pode ser realizada a partir da coleta de dados dos usuários durante o consumo.

Pode-se apreender da literatura revisitada nesta tese que a compreensão tem sido priorizada quando realizada por humanos, que possuem estrutura cognitiva, modelos mentais e conhecimento prévio. Contudo, no que tange à compreensão por máquinas, pouco explorada no contexto da Arquitetura da Informação, há a necessidade de semântica formal, vocabulários padronizados que permitem localizar,

⁷⁰ *“When designing complex information environments, it is particularly important to provide context within the greater whole. Many contextual clues available in the physical world do not exist online. There are no natural landmarks, no north and south. Unlike physical travel, hypertextual navigation allows users to be transported right into the middle of an unfamiliar system.”*

acessar, interoperar e utilizar os dados – características desejáveis pelos Princípios FAIR (acrônimo de *Findable, Accessible, Interoperable, Reusable*) –, que podem ainda ser utilizados no contexto e em benefício das técnicas de Inteligência Artificial e Ciência de Dados.

O sistema de navegação auxilia os usuários, quer saibam com exatidão o que buscam ou naveguem pelo ambiente sem a clareza do que pretendem localizar. Neste contexto, “Conforme os usuários se movem por um sistema de navegação bem projetado, eles aprendem sobre produtos, serviços ou tópicos associados ao conteúdo específico que pretendem encontrar.”⁷¹ (ROSENFELD; MORVILLE, 1998, p. 42, tradução nossa). Essa afirmação corrobora os preceitos da descoberta acidental e da serendipidade, presentes na Encontrabilidade da Informação (VECHIATO, 2013; VECHIATO; FARIAS, 2020), por meio das quais, em um processo de busca, o sujeito encontra uma informação acidental ou fortuitamente, que não corresponde à sua necessidade primária, mas sim a uma necessidade em segundo plano que pode, inclusive, não estar relacionada à intencionalidade da busca.

Além disso, no sistema de navegação, é relevante que a funcionalidade e a usabilidade se sobreponham ao design e à estética. Para tanto, devem ser considerados fatores que são parte do processamento do usuário, para que a navegação seja beneficiada.

Destacam Rosenfeld e Morville (1998, p. 50, tradução nossa) que “No projeto de sistemas de navegação para a Web, o contexto é fundamental.”⁷² Essa característica contextual é também extremamente relevante no contexto dos dados, uma vez que sozinhos e por si só podem não fornecer significado relevante aos consumidores. Por isso, devem possuir granularidade e, ao mesmo tempo, variedade e completude para manter as características de representação da entidade e dos seus contextos.

Para Rosenfeld e Morville (1998, p. 60, tradução nossa), projetar o sistema de navegação é um desafio, sobretudo pela vasta disponibilidade de tecnologias disponíveis. Por isso, alertam que não deve haver desvio do objetivo central do

⁷¹ “As users move through a well-designed navigation system, they learn about products, services, or topics associated to the specific content they set out to find.”

⁷² “In designing navigation systems for the Web, context is king.”

sistema de navegação “[...] construir contexto, melhorar a flexibilidade e ajudar o usuário a encontrar as informações de que precisa.”⁷³

Morville e Rosenfeld (2006) reconhecem o sistema de navegação como um elemento da Arquitetura da Informação que se relaciona com design de interação, design de informação, design visual, engenharia de usabilidade e design de experiência do usuário.

É visível a complexidade do projeto e desenvolvimento do sistema de navegação, uma vez que ele se relaciona intimamente com a tríade da Arquitetura da Informação – conteúdo, contexto e usuário –, com os demais sistemas da Arquitetura da Informação (MORVILLE; ROSENFELD, 2006) e, ainda, com aspectos de design e experiência do usuário.

Por outro lado, cumpre-se ressaltar a influência dos vocabulários controlados e da estrutura dos dados para a localização de informações via ferramenta ou mecanismo de busca, visto que o contato inicial com determinada informação pode partir de uma fonte externa, mas caberá ao sistema de navegação guiar o usuário durante sua permanência no ambiente informacional, mesmo que este não o tenha acessado por sua página principal. Tal afirmação reforça a importância da orientação espacial (*wayfinding*) e das trilhas de navegação (*breadcrumb*).

3.2.4 Sistema de Busca

O sistema de busca está fortemente atrelado ao planejamento realizado na implantação do ambiente informacional. Inicialmente, Rosenfeld e Morville (1998), Morville e Rosenfeld (2006) e Rosenfeld, Morville e Arango (2015) afirmam que é necessário planejar a implementação do sistema de busca, tendo em vista os sistemas de organização, rotulagem e navegação, as características dos conteúdos, dos usuários, além da quantidade e dinamicidade de disponibilização de conteúdo. Neste sentido alertam que o sistema de busca pode não se aplicar e não solucionar os problemas existentes na localização de informações em alguns ambientes informacionais.

⁷³ “[...] *building context, improving flexibility, and helping the user to find the information they need.*”

Posteriormente, é necessário ter clareza de que há a necessidade de estruturação dos conteúdos e de configuração adequada do motor de busca, para que a implantação do sistema de busca seja realizada.

Apesar da afirmação dos autores é visível que o advento das tecnologias e o crescimento das fontes de informação, analógicas e digitais, culminaram na ampliação da oferta de motores e ferramentas de busca. Nesse sentido, é importante ponderar que a maioria dos sistemas de informação e ambientes informacionais digitais disponibiliza motores de busca, sobretudo em função da massiva quantidade de dados disponíveis para a pesquisa. Isso fez com que os softwares fossem aprimorados, visando favorecer os processos de indexação, recuperação, ranqueamento e até mesmo recomendação, processos nos quais, mais recentemente, são empregadas técnicas de Inteligência Artificial.

De igual maneira, há o crescimento da oferta de mecanismos de buscas, gerais ou específicos – como os acadêmicos – que realizam a coleta e disponibilização para a consulta de inúmeros sistemas de informação e ambientes informacionais digitais, atuando como uma fonte central de busca, cuja forma de funcionamento do motor de busca é muitas vezes desconhecida pela maioria dos sistemas e ambientes recuperados. Além destes, no contexto da Ciência da Informação, há ainda sistemas de descoberta e entrega, que consistem em sistemas que integram diversas fontes de informação vinculadas ao contexto de implantação e permitem a busca e exibição dos resultados em uma interface única.

Para Rosenfeld e Morville (1998, p. 84, tradução nossa), “[...] sistemas de busca podem e devem variar tanto quanto os sistemas de navegação ou quaisquer outros componentes de websites, porque os usuários não são iguais e a recuperação da informação é muito mais difícil do que a maioria das pessoas imagina.”⁷⁴

Os autores elucidam que os usuários em um processo de busca possuem necessidades e expectativas diferentes, bem como preocupações diversas com a fonte, a qualidade e a quantidade de informações recuperadas. Destaca-se que essa variabilidade está diretamente relacionada ao processo de busca, ao comportamento dos sujeitos, a intencionalidade da busca, à qualidade dos dados e metadados e, por isso, influencia a Recuperação da Informação e a Encontrabilidade da Informação.

⁷⁴ “[...] searching systems can and should vary as much as browsing systems or any other components of web sites do, because all users aren't alike, and information retrieval is much harder than most people realize.”

Além disso, cada usuário possui necessidades de informação e comportamento de busca próprios, que serão preponderantes no uso do ambiente informacional e influenciam na Experiência do Usuário e na Encontrabilidade da Informação. No que tange à necessidade da informação, pode haver tipos de buscas diversos, como a busca por um item conhecido, a busca exploratória, a busca exaustiva, retorno a um resultado de busca anterior, que estão diretamente atreladas ao que o usuário espera encontrar no ambiente. Enquanto o comportamento de busca se utiliza de iteração e integração.

Morville e Rosenfeld (2006) apresentam um modelo de determinação de necessidades de informação designado *too-simple*, composto por 3 passos: uma pergunta inserida pelo usuário, um processamento que ocorre de forma desconhecida pelo usuário e uma resposta apresentada pelo sistema ao usuário. Os autores consideram essa “simplicidade” composta por entrada e saída muito mecanicista, uma vez que desconsidera tudo o que ocorre entre a entrada e a saída, ou seja, o comportamento de busca do usuário, a busca de informações e a descoberta acidental, e, utiliza apenas algoritmos para recuperar dados utilizando tecnologia da informação.

Esse modelo é perigoso porque é construído sobre um equívoco: encontrar informações é um problema direto que pode ser resolvido por uma abordagem algorítmica simples. Afinal, resolvemos o desafio de recuperar dados – que, claro, são fatos e números – com tecnologias de banco de dados como SQL. Então, segue o pensamento, vamos tratar as ideias e conceitos abstratos incorporados em nossos documentos textuais semiestruturados da mesma maneira.⁷⁵ (MORVILLE; ROSENFELD, 2006, p. 32, tradução nossa).

Ao contrário, Morville e Rosenfeld (2006) reconhecem que o modelo *berry-picking* se aproxima ao funcionamento da mente humana. Nele, o usuário insere sua necessidade de informação por meio de uma expressão de busca e, posteriormente se movimenta pelo ambiente, de onde pode extrair informações. Durante o processo, as solicitações podem ser modificadas à medida que internaliza informações e o usuário pode seguir extraindo informações e alterando a busca. Este modelo requer a fácil alternância entre a busca e a navegação.

⁷⁵ “This model is dangerous because it’s built upon a misconception: that finding information is a straightforward problem that can be addressed by a simple, algorithmic approach. After all, we’ve solved the challenge of retrieving data— which, of course, is facts and figures—with database technologies such as SQL. So, the thinking goes, let’s treat the abstract ideas and concepts embedded in our semi-structured textual documents the same way.”

Os autores apresentam ainda o modelo que designam como *pearl-growing*, no qual o usuário parte de um conteúdo do seu interesse e, por meio de sistemas de recomendação ou de citações, pode acessar outros conteúdos relacionados, potencialmente de interesse, pela similaridade, por exemplo, da temática e de palavras indexadas (MORVILLE; ROSENFELD, 2006).

No que tange aos websites corporativos e intranets, Morville e Rosenfeld (2006) destacam o modelo de duas etapas, no qual por meio de uma informação inicial o usuário pode pesquisar ou navegar por links que ligam inúmeros subsites nos quais buscam informações.

Em geral, os usuários costumam alternar suas buscas, entre a navegação e o uso do motor de busca. “Na verdade, os usuários geralmente não sabem se precisam pesquisar ou navegar em primeiro lugar. Portanto, esses respectivos sistemas não devem viver isolados uns dos outros.”⁷⁶ (ROSENFELD; MORVILLE, 1998, p. 90, tradução nossa).

De igual maneira, os usuários podem ter certeza da sua necessidade de informação e com isso recorrer a determinadas fontes e usar expressões e/ou termos de busca específicos, ao passo que é possível que não haja clareza da necessidade de informação, o que leva à realização de buscas exploratórias, seguidas de refinamentos até que cheguem ao resultado esperado ou próximo dele.

Diante do exposto, um dos desafios da AI consiste em “[...] projetar a arquitetura do seu site [ambiente informacional digital] para suportar as abordagens de pesquisa e navegação mais comuns de uma forma harmoniosa e integrada.”⁷⁷ (ROSENFELD; MORVILLE, 1998, p. 86, tradução nossa).

Considerando as afirmações anteriores, o sistema de busca deve ser projetado analisando a tríade da AI – conteúdo, contexto e usuário –, o que requer a presença de múltiplas opções, dentre as quais Rosenfeld e Morville (1998) destacam: tipologias, idiomas, operadores de busca, instruções e ajuda, além de uma interface específica para buscas avançadas. Tudo isso utilizando, preferencialmente, um design limpo e intuitivo, além de considerar as condições de usabilidade e

⁷⁶ “In fact, users often don't know if they need to search or browse in the first place. Therefore, these respective systems shouldn't live in isolation from one another.”

⁷⁷ “[...] design your site's architecture to support the most common searching and browsing approaches in a smooth and integrated way.”

navegabilidade. Adicionalmente, podem ser incluídas opções de filtros para os resultados encontrados, visando permitir o refinamento.

Há de se considerar, ainda, as tecnologias que serão empregadas como motor de busca, o que pode requerer *softwares* específicos para as tarefas principais – coleta, indexação, ordenação e apresentação – considerando que serão recuperados apenas itens armazenados no próprio ambiente. Além disso, podem ser necessários algoritmos para determinar o ranqueamento dos itens recuperados e a recomendação de itens relacionados.

Apesar disso, “As interfaces do mecanismo de pesquisa e, mais importante, os resultados da recuperação devem ter a aparência e o comportamento do restante do seu site [ambiente informacional digital].”⁷⁸ (ROSENFELD; MORVILLE, 1998, p. 91, tradução nossa), o que pode requerer personalizações, ainda que o motor de busca esteja empacotado no próprio ambiente informacional. Neste sentido, é necessário que o sistema de busca seja planejado como parte da interface ou em uma página exclusiva, dependendo das suas opções e dimensões.

Para o funcionamento adequado do sistema de busca, é desejável que as informações disponíveis no ambiente estejam estruturadas. Nesse sentido, quando dados estão estruturados, utilizando metadados padronizados, é possível que todos os campos sejam utilizados nas buscas, ampliando os pontos de acesso aos conteúdos e itens disponíveis no ambiente informacional digital. É essa estruturação que permite que os conteúdos disponíveis nos ambientes informacionais sejam alavancados por meio da Arquitetura da Informação *bottom-up* e que se entende otimizada pela Arquitetura de Dados, conceituada nesta tese.

Morville e Rosenfeld (2006) corroboram desta afirmação “[...] os sistemas de busca geralmente têm melhor desempenho se puderem aproveitar os aspectos dos sistemas de navegação fortes, como os termos de vocabulário controlado usados para marcar o conteúdo.”⁷⁹

É necessário considerar ainda que as buscas podem ser realizadas por consumidores humanos, por meio de uma interface gráfica, no *front-end*, que poderá empregar aspectos cognitivos aos resultados recuperados, propiciando

⁷⁸ “Search engine interfaces, and more importantly, retrieval results, should look and behave like the rest of your site.”

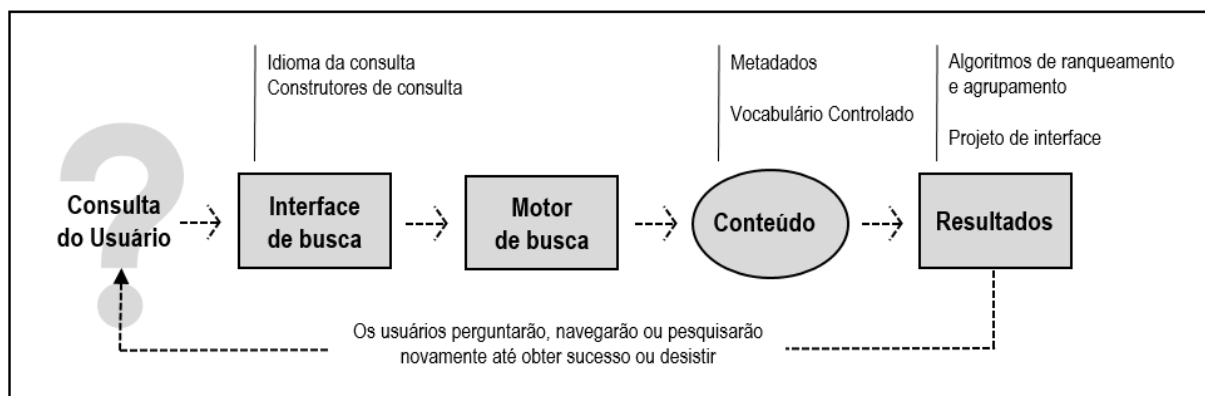
⁷⁹ “[...] search systems often perform better if they can take advantage of aspects of strong navigation systems, such as the controlled vocabulary terms used to tag content.”

Encontrabilidade da Informação. Ao passo que o consumo por máquinas será realizado por meio de uma camada disponível no *back-end*, e o processamento é totalmente dependente da comunicação utilizando estruturas e protocolos de comunicação de dados, que podem propiciar Encontrabilidade de Dados.

Em Morville e Rosenfeld (2006) a anatomia do sistema de busca (Figura 4) é apresentada e os autores elucidam que há uma dinâmica no processo de busca que, em geral, é desconhecida pelo usuário, que ocorre justamente após a consulta por ele efetuada. Esse processo envolve as partes do conteúdo que são indexadas, os metadados, os algoritmos, os *crawler/spider*, os índices, as interfaces de consulta e de resultado, as linguagens de consulta e os construtores de consultas.

“Os projetistas precisam adotar uma abordagem sistêmica que reconheça os papéis desempenhados pela interface de busca, o conteúdo e a apresentação dos resultados.”⁸⁰ (MORVILLE, 2001). Tal afirmação explicita a relevância da anatomia do sistema de busca, para além das “caixas de pesquisa” disponíveis nos ambientes informacionais digitais, mas considerando a complexidade existente entre o conteúdo, seu tratamento, os *softwares*, os algoritmos, e a própria representação da necessidade informacional por parte do consumidor. Toda essa complexidade é resumida na Figura 4, que apresenta a anatomia básica do sistema de busca.

Figura 4 – Sistema de Busca



Fonte: Morville (2001, tradução nossa).

Diante dessa complexidade e considerando que, mais que uma ferramenta, a busca deve ser considerada um sistema, Morville (2001, tradução nossa) alerta:

⁸⁰ “Designers need to take a systemic approach that recognizes the roles played by the search interface, the content and the results presentation.”

Se o seu site [ambiente informacional digital] estiver cheio de ROT [*Redundant, Outdated or Trivial content*] (conteúdo redundante, desatualizado ou trivial), os resultados da busca serão prejudicados. Se o seu algoritmo de classificação enterrar os resultados mais relevantes, os usuários não conseguirão encontrá-los. Como muitos outros sistemas complexos, a busca é tão boa quanto seu link mais fraco.

E precisamos projetar para a natureza iterativa e interativa do comportamento de busca de informações dos usuários, reconhecendo que as pessoas precisam se mover com fluidez entre os modos de pesquisa, navegação e pergunta.⁸¹

No que tange à apresentação dos resultados, Rosenfeld e Morville (1998) afirmam que é necessário considerar como o conteúdo está estruturado e as expectativas dos usuários e, destacam que o ideal é possibilitar que o usuário configure a página de acordo com seus interesses, considerando, por exemplo, os campos, a quantidade e a forma de classificação dos registros recuperados, além da resolução da tela. Acrescenta-se a isso que dispositivos móveis podem requerer personalizações específicas em função do tamanho da tela.

Na apresentação dos resultados via interface, para Morville e Rosenfeld (2006), há três questões preponderantes, a forma de apresentação, quais elementos de conteúdo exibir e a quantidade de registros exibidos por consulta. E, o ranqueamento dos resultados ganha destaque

[...] independentemente de quantas maneiras você indicar há mais resultados do que cabem em uma tela, muitos (se não a maioria) usuários nunca se aventurarão além dessa primeira tela. Portanto, não exagere ao fornecer muito conteúdo por resultado, pois os primeiros resultados podem obscurecer o restante da recuperação.⁸² (MORVILLE; ROSENFELD, 2006, p. 163, tradução nossa).

Além disso, considerando os ambientes informacionais digitais, afirmam que “Os usuários geralmente assumem que os primeiros resultados são os melhores.”⁸³ (MORVILLE; ROSENFELD, 2006, p. 168, tradução nossa). Essa afirmação, também presente em Rosenfeld, Morville e Arango (2015, p. 240), reforça a necessidade de estruturação do conteúdo disponibilizado no ambiente informacional, para que o

⁸¹ “If your site is bloated with ROT (*Redundant, Outdated or Trivial content*), search results will suffer. If your ranking algorithm buries the most relevant results, users will fail to find them. Like many other complex systems, search is only as good as its weakest link.

And we need to design for the iterative, interactive nature of users’ information seeking behavior, recognizing that people need to move fluidly between searching, browsing and asking modes.”

⁸² “[...]regardless of how many ways you indicate that there are more results than fit on one screen, many (if not most) users will never venture past that first screen. So don’t go overboard with providing lots of content per result, as the first few results may obscure the rest of the retrieval.”

⁸³ “Users will generally assume that the top few results are best.”

mecanismo de busca possa localizá-lo, reconhecê-lo como relevante e contextual e então exibi-lo. Além disso, é a estruturação do conteúdo que possibilita determinar que informações relacionadas a cada conteúdo recuperado serão exibidas ao usuário na página de resultados, importantes para auxiliá-lo na definição dos resultados que são relevantes para a sua necessidade de informação.

Nesse contexto, também pode ser relevante considerar que, para a exibição dos resultados na página de resultado de busca, podem ser aplicados diferentes métodos. Rosenfeld, Morville e Arango (2015) mencionam a ordenação e a classificação, que podem ser alfabéticas, cronológicas, por relevância, por popularidade, por avaliação ou por anúncio pago (*pay for placement* ou P4P), e cuja aplicação pode, inclusive, ser escolhida pelo usuário. Em todos os casos, para a ordenação ou classificação, o conteúdo precisa estar estruturado com metadados. No caso da relevância, o ranqueamento é realizado por algoritmos que possuem funcionamentos diversos, mas que se utilizam da estruturação do conteúdo e de sua representação para determinar a relevância em relação à expressão de busca empregada. Assim, além de estruturar adequadamente os metadados, pode ser relevante se atentar às técnicas de *Search Engine Optimization* (SEO), caso o conteúdo seja geral ou de *Academic Search Engine Optimization* (ASEO) para conteúdo acadêmico⁸⁴.

Em ambientes analógicos, o processo de busca pode ser mediado por profissionais capacitados para o serviço de referência, dos quais destacam-se os bibliotecários, que podem auxiliar o usuário, por exemplo, na definição dos termos e fontes para a consulta (ROSENFELD; MORVILLE, 1998). Por outro lado, àquela época, os autores enfatizaram que um ambiente informacional não pode automatizar algumas atividades que são essencialmente desenvolvidas por humanos, sobretudo, no que tange aos “contextos”.

Atualmente, alguns ambientes informacionais digitais que disponibilizam motor de busca possuem ferramentas para auxiliar o usuário, com uma espécie de serviço de referência digital. Neste sentido, inicialmente isso ocorria com atendimento de humanos, via *chat* em tempo real ou agendamento para atendimento remoto. Atualmente, o processo tem sido automatizado por aplicações baseadas em

⁸⁴ Nesse contexto, os produtores de conteúdo, seja ele geral ou acadêmico, incluindo os textos científicos, devem também se atentar às técnicas de SEO e ASEO para melhorar a relevância do conteúdo para os algoritmos de classificação de relevância e, com isso, melhorar o ranqueamento na página de resultados de busca.

Inteligência Artificial (IA), a exemplo do *chatbot*, que pode se constituir em um serviço de referência automatizado. É necessário esclarecer, contudo, que as aplicações computacionais, ainda que utilizem IA, são baseadas em programação e necessitam de uma base para o treinamento/aprendizagem de máquina, visando criar contextos em ambientes digitais, similares aos que ocorrem em ambientes analógicos e com atendimento humano.

Assim, é imprescindível que, no momento da programação, da escolha dos algoritmos e das bases de treinamento/aprendizagem de máquina, sejam considerados conteúdos, contextos e usuários, para que as informações sejam adequadamente representadas e recuperadas. Esse processo necessita de amplo conhecimento do conteúdo armazenado e dos usuários potenciais e reais do ambiente e, para tanto, pode utilizar a coleta de dados dos usuários, humanos ou máquinas, durante o consumo.

Para Rosenfeld e Morville (1998, p. 101, tradução nossa) “[...] a busca funciona muito melhor quando o espaço de informações é definido de forma restrita e contém conteúdo homogêneo.”⁸⁵ Além disso, destacam que o desempenho da busca está relacionado à forma de indexação, que pode ser realizada por todo o conteúdo ou de forma mais seletiva, ao permitir, por exemplo, que o consumidor defina os tipos, metadados e áreas do ambiente que deseja pesquisar. Podem, ainda, refletir o sistema de organização utilizado no ambiente informacional.

Morville e Rosenfeld (2006) e Rosenfeld, Morville e Arango (2015), embora não tivessem objetivo de se aprofundarem no processo de recuperação da informação, destacam que os algoritmos de recuperação são o centro dos motores e mecanismos de busca, e que cada um possui uma especificidade, o que dificulta conceber que o sistema de busca seja capaz de atender a todas as expectativas dos usuários do ambiente informacional.

Rosenfeld, Morville e Arango (2015) retomam os algoritmos para apresentar aqueles que podem ser utilizados para auxiliar os usuários na construção da estratégia de busca. Nesse sentido, discorrem sobre os verificadores ortográficos, as ferramentas fonéticas, de derivação e de processamento de linguagem natural e, ainda, os vocabulários controlados. Consideram ainda que o perfil dos usuários do ambiente informacional é preponderante para identificar que algoritmos são

⁸⁵ “[...] searching works much better when the information space is defined narrowly and contains homogeneous content.”

necessários no sistema de busca. Para tanto, é possível, inclusive, analisar os logs de consultas realizadas.

No contexto da presente tese, é reforçada a relevância dos algoritmos de recuperação da informação e do tratamento dos objetos informacionais como insumos para os processos de Recuperação da Informação e Encontrabilidade da Informação. Lembrando que tanto os algoritmos quanto a estrutura dos dados influenciam a precisão e a revocação no processo de recuperação da informação – o que reforça a necessidade de estruturação das informações por meio de metadados, para que sejam adequadamente consumidas pelo próprio sistema de busca do ambiente e beneficiem a recuperação e a Encontrabilidade da Informação.

O conteúdo, uma vez estruturado por metadados e representado de maneira exaustiva, pode se valer de tecnologias computacionais para enriquecer a busca e a recuperação da informação por meio, inclusive, da descoberta acidental e, com isso, contribuir, por exemplo, para o desenvolvimento e aprimoramento de sistemas de recomendação, para a criação de bases de treinamento para aprendizagem de máquina, para o aprimoramento do processamento de linguagem natural (PLN) e para o enriquecimento semântico de termos.

Além disso, a utilização adequada de metadados, tanto no conteúdo/objeto informacional quanto na sua representação, propicia que se tornem FAIR. Nesse sentido, é preponderante destacar que a FAIRificação do ambiente informacional, embora altamente desejável, é insuficiente, uma vez que ambientes informacionais FAIR podem tornar os conteúdos/objetos informacionais localizáveis e, em certa medida acessíveis. Contudo, para que os conteúdos/objetos informacionais sejam realmente acessíveis, interoperáveis e reutilizáveis, é imprescindível que eles estejam estruturados por meio de metadados do domínio, para que sejam facilmente acionados e processados. Os Metadados e Princípios FAIR são abordados com mais detalhes, respectivamente, nas seções 4.4.1.3 e 4.4.1.5.3 desta tese.

Ainda, Morville e Rosenfeld (2006) reconhecem que a configuração do mecanismo de busca é melhor desempenhada por profissionais da área da computação. Contudo, destacam que a sua implantação ocorre para atender às necessidades do usuário e, neste sentido, alertam que o Arquiteto da Informação deve atuar também no projeto e desenvolvimento desse sistema, considerando a inter-relação dele com os demais sistemas e a prevalência da funcionalidade ao *software* ou linguagem de programação.

A interação entre o usuário e o sistema de busca ocorre por meio da interface gráfica, parte integrante do escopo da Arquitetura da Informação. Nesse sentido, Morville e Rosenfeld (2006) e Rosenfeld, Morville e Arango (2015) afirmam que a interface de busca disponibilizada ao usuário evoluiu juntamente com o crescimento de usuários da web, sendo, nos primórdios, necessário conhecimento especializado acerca de operadores de busca, e, atualmente esses operadores incluídos automaticamente pelo algoritmo de consulta.

Assim, em geral, o que se apresenta na página principal do motor de busca é uma caixa para busca básica e a página de resultados exibe opções de filtro, enquanto a busca avançada fica disponível em uma página secundária, acessível a partir da página de busca básica, por meio da qual outros operadores e filtros podem ser selecionados pelo próprio usuário.

Morville e Rosenfeld (2006, p. 182, tradução nossa) reconhecem que os usuários geralmente iniciam a pesquisa simplesmente digitando um conjunto de palavras ou uma frase, sem despender tempo na construção de uma estratégia de busca: “Você certamente poderia fornecer uma página de ‘ajuda’ que explicasse como criar consultas mais avançadas e precisas, mas os usuários raramente visitam esta página.”⁸⁶

Com isso, reconhecem que a interface do mecanismo de busca deve ser simples e que, à medida que o usuário necessite filtrar os resultados, seja possível que ele aprenda com o funcionamento do motor de busca. Apesar disso, destacam que deve haver investimento na interface de busca avançada, ainda que ela tenha uma quantidade menor de usuários.

Talvez uma boa regra seja descobrir as várias funções de busca pesadas de seu mecanismo de busca na página avançada para os poucos usuários que desejam experimentá-las, mas projete seu sistema de busca com o objetivo de tornar desnecessário para a grande maioria dos pesquisadores a necessidade de ir para a página de pesquisa avançada.⁸⁷ (MORVILLE; ROSENFELD, 2006, p. 185, tradução nossa).

Vale lembrar que Wurman (1996) já afirmava que os modelos mentais e conhecimentos prévios são essenciais para a compreensão e o aprendizado.

⁸⁶ “You certainly could provide a “help” page that explains how to create more advanced, precise queries, but users may rarely visit this page.”

⁸⁷ “Perhaps a good rule of thumb is to unearth your search engine’s various heavy-duty search functions on the advanced page for those few users who want to have a go at them, but design your search system with the goal of making it unnecessary for the vast majority of searchers to ever need to go to the advanced search page.”

No contexto da Arquitetura da Informação, é importante reforçar o relacionamento entre os diferentes sistemas que compõem sua anatomia. No sistema de busca, os demais sistemas são claramente visualizados: por exemplo, o sistema de organização possibilita gerar buscas por áreas específicas do ambiente informacional, o sistema de rotulagem contribui para as buscas por tópicos específicos, o sistema de navegação possibilita que o usuário trace os caminhos a partir dos resultados, enquanto os tesouros, vocabulários controlados e metadados fornecem a estrutura necessária para a indexação, recuperação e ranqueamento dos resultados de busca.

Para que isso ocorra, a informação deve estar corretamente estruturada e armazenada na camada de dados dos sistemas de informação utilizados pelos ambientes informacionais, visando, por meio de aplicações e ferramentas computacionais, exibir os resultados na interface gráfica, que consiste na camada superior dos ambientes informacionais digitais. É justamente neste aspecto que a Arquitetura da Informação *bottom-up* se relaciona diretamente com a Arquitetura de Dados no contexto da Ciência da Informação.

3.2.5 Tesouros, Vocabulários Controlados e Metadados

A abordagem sistêmica da Arquitetura da Informação (ROSENFELD; MORVILLE, 1998) inicialmente não tratava de forma aprofundada o uso de tesouros, vocabulários controlados e metadados. Contudo, na obra de Morville e Rosenfeld (2006), a relevância de tesouros, vocabulários controlados e metadados ganhou destaque em um capítulo da obra, característica mantida em Rosenfeld, Morville e Arango (2015).

É possível vislumbrar essa abordagem em função da sua relevância para a constituição dos sistemas que compõem a Arquitetura da Informação (ROSENFELD; MORVILLE, 1998) – sistema de organização, sistema de rotulagem, sistema de navegação e sistema de busca – que, claramente, podem ser beneficiados pelo uso de tesouros, vocabulários controlados e metadados, que nesta tese se denominam Sistema de Representação da Informação, conforme apresentado por Oliveira e Vidotti (2012). Vale destacar que um Sistema de Representação da Informação, sob o olhar da CI, possui outros elementos que não foram tratados pelos autores, como por exemplo taxonomia e ontologias.

Esta denominação, como sistema de representação da informação, deve-se à abordagem dada, em Morville e Rosenfeld (2006) e Rosenfeld, Morville e Arango (2015), aos tesouros, vocabulários controlados e metadados, a partir da Ciência da Informação, reforçando o emprego na representação da informação como suporte à recuperação da informação. Além destes, é possível vislumbrar ainda o uso nos sistemas de recomendação, bastante presentes em ambientes digitais, e cujo uso em ambientes informacionais digitais tem sido ampliado.

Em Rosenfeld, Morville e Arango (2015), tesouros, vocabulários controlados e metadados, são considerados “[...] sistemas “invisíveis” que ajudam a moldar o ambiente informacional nos bastidores.”⁸⁸. Tais sistemas designados pelos autores como “invisíveis” tornam-se visíveis na Arquitetura de Dados.

Nesse sentido, destaca-se, a partir de Morville e Rosenfeld (2006, p. 216, tradução nossa), que “[...] precisamos fundamentar nosso trabalho no objetivo final de aprimorar a capacidade de nossos usuários de localizar o que precisam.”⁸⁹

Na obra de Morville e Rosenfeld (2006), tesouros, vocabulários controlados e metadados são definidos e suas aplicações exemplificadas. Nesta seção da tese, dedicada a revisitar os autores clássicos da Arquitetura da Informação, não há a pretensão de aprofundar a abordagem de tais conceitos, que serão tratados apenas sob o prisma dos referidos autores.

Assim, visando compreender como esses conceitos são tratados nas obras de Morville e Rosenfeld (2006) e Rosenfeld, Morville e Arango (2015), eles são revisitados e discutidos nesta seção da tese, visando igualmente compreender seu papel na Arquitetura da Informação.

No que tange aos metadados, Morville e Rosenfeld (2006) iniciam afirmando que são dados sobre dados. A clássica e generalista definição é seguida de uma explicação um pouco mais técnica e os autores continuam afirmando “As tags de metadados são usadas para descrever documentos, páginas, imagens, software, arquivos de vídeo e áudio e outros objetos de conteúdo para fins de navegação e recuperação aprimoradas.”⁹⁰ Morville e Rosenfeld (2006, p. 194).

Enquanto vocabulário controlado é compreendido como:

⁸⁸ “[...] “invisible” systems that help shape the information environment behind the scenes.”

⁸⁹ “[...] we need to ground our work in the ultimate goal of enhancing the ability of our users to find what they need.”

⁹⁰ “Metadata tags are used to describe documents, pages, images, software, video and audio files, and other content objects for the purposes of improved navigation and retrieval.”

Em sua forma mais vaga, um vocabulário controlado é qualquer subconjunto definido de linguagem natural. Na sua forma mais simples, um vocabulário controlado é uma lista de termos equivalentes na forma de um anel de sinônimos ou uma lista de termos preferidos na forma de um arquivo de autoridade. Defina relacionamentos hierárquicos entre os termos (por exemplo, mais amplo, mais restrito) e você terá um esquema de classificação. Modele relacionamentos associativos entre conceitos (por exemplo, ver, ver também) e você está trabalhando em um tesouro.⁹¹ (MORVILLE; ROSENFELD, 2006, p. 194, tradução nossa).

Para Morville e Rosenfeld (2006, p. 203, tradução nossa)

[...] tesouro é uma rede semântica de conceitos, conectando palavras a seus sinônimos, homônimos, antônimos, termos mais amplos e mais restritos e termos relacionados.

[...] Seu objetivo mais importante é o gerenciamento de sinônimos – o mapeamento de muitos sinônimos ou variantes de palavras em um termo ou conceito preferido – para que as ambiguidades da linguagem não impeçam as pessoas de localizar o que precisam.⁹²

Rosenfeld, Morville e Arango (2015, p. 282, tradução nossa) reforçam a afirmação anterior, e a complementam:

No entanto, nosso tesouro assume a forma de um banco de dados online, totalmente integrado à interface do usuário de um produto ou serviço digital. E embora o tesouro tradicional ajude as pessoas a passar de uma palavra para muitas palavras, nosso tesouro faz o oposto. Seu objetivo mais importante é o gerenciamento de sinônimos – o mapeamento de muitos sinônimos ou variantes de palavras em um termo ou conceito preferido – para que as ambiguidades da linguagem não impeçam as pessoas de encontrar o que precisam.⁹³

Ainda, segundo Rosenfeld, Morville e Arango (2015), os tesouros possuem três tipos básicos de relacionamentos semânticos: equivalência, hierárquica e associativa. Essa rede semântica tem como centro o termo preferido, também chamado de termo autorizado, cabeçalho de assunto e descritor; a relação de equivalência consiste no gerenciamento de sinônimos, chamados de termos

⁹¹ “At its most vague, a controlled vocabulary is any defined subset of natural language. At its simplest, a controlled vocabulary is a list of equivalent terms in the form of a synonym ring, or a list of preferred terms in the form of an authority file. Define hierarchical relationships between terms (e.g., broader, narrower) and you’ve got a classification scheme. Model associative relationships between concepts (e.g., see also, see related) and you’re working on a thesaurus.”

⁹² “[...] thesaurus is a semantic network of concepts, connecting words to their synonyms, homonyms, antonyms, broader and narrower terms, and related terms.

[...] Its most important goal is synonym management—the mapping of many synonyms or word variants onto one preferred term or concept—so the ambiguities of language don’t prevent people from finding what they need.”

⁹³ “However, our thesaurus takes the form of an online database, tightly integrated with the user interface of a digital product or service. And though the traditional thesaurus helps people go from one word to many words, our thesaurus does the opposite. Its most important goal is synonym management—the mapping of many synonyms or word variants onto one preferred term or concept—so the ambiguities of language don’t prevent people from finding what they need.”

variantes; a hierárquica possibilita estabelecer categorias e subcategorias, designados como termo geral e termo específico; e a associativa estabelece conexões que não se enquadram nos relacionamentos anteriores, considerados termos relacionados.

Os tesouros e vocabulários controlados são especialmente úteis para o sistema de rotulagem, uma vez que geralmente são criados e mantidos por especialistas em representação da informação e no domínio. De igual maneira, são importantes para a navegação e para a recuperação da informação.

Para Morville e Rosenfeld (2006), o ideal é que sejam utilizados tesouros e vocabulários controlados específicos do domínio do conteúdo a ser representado, o que beneficia a compreensão dos usuários dada a terminologia utilizada. Porém, alertam que nem todos os domínios possuem tesouros e vocabulários controlados, o que gera a necessidade de uso de um domínio genérico ou ainda a criação de um sistema de rotulagem.

No que tange ao sistema de busca, quando utilizados como construtores de busca, os tesouros e vocabulários controlados, por meio da semântica, podem incluir automaticamente na consulta, por exemplo, sinônimos e termos relacionados (MORVILLE; ROSENFELD, 2006). Enquanto, para a indexação, favorecem o estabelecimento de uma forma padronizada de descrição de atributos das entidades representadas, a partir da semântica, gerando consistência nas informações armazenadas em uma base de dados.

Metadados e vocabulários controlados apresentam uma lente fascinante através da qual podemos ver a rede de relacionamentos entre sistemas. Em muitos grandes websites orientados por metadados, vocabulários controlados tornaram-se a cola que mantém os sistemas juntos. Um dicionário de sinônimos no back-end pode permitir uma experiência de usuário mais perfeita e satisfatória no front-end.⁹⁴ (MORVILLE; ROSENFELD, 2006, p. 193, tradução nossa).

Assim, tesouros, vocabulários controlados e metadados, inicialmente utilizados como ferramentas de apoio ao trabalho técnico de representação da informação pela Ciência da Informação, sobretudo em catálogos para a representação de conteúdos informacionais, passaram a ganhar espaço na web. Morville e Rosenfeld

⁹⁴ *“Metadata and controlled vocabularies present a fascinating lens through which we can view the network of relationships between systems. In many large metadatadriven web sites, controlled vocabularies have become the glue that holds the systems together. A thesaurus on the back end can enable a more seamless and satisfying user experience on the front end.”*

(2006, p. 227, tradução nossa) em 2006 já percebiam o uso de ferramentas da Ciência da Informação, no contexto mais genérico, na representação da informação na web: “Metadados, vocabulários controlados e tesouros estão se tornando cada vez mais os blocos de construção da maioria dos principais websites e intranets. As soluções de taxonomia única estão dando lugar a abordagens mais flexíveis e facetadas.”⁹⁵

É evidente que o uso dessas ferramentas por usuários que não as dominam pode gerar algum transtorno inicial. Contudo, assim como discutido anteriormente, no sistema de busca é possível criar mecanismos para familiarizar os usuários, considerando, sobretudo, os benefícios que serão alcançados.

Em Rosenfeld, Morville e Arango (2015), são abordadas as relações semânticas existentes nos tesouros, o que possibilita compreender o papel do sistema de representação na recuperação da informação. Evidentemente, o emprego do sistema de representação no campo de valor dos dados torna o processo de recuperação da informação mais eficiente e enriquecido, se consideradas as relações semânticas possíveis. Isso colabora com a descoberta e Encontrabilidade da Informação por humanos.

Por outro lado, embora as discussões de Morville e Rosenfeld (2006) e Rosenfeld, Morville e Arango (2015) se concentrem no que se designa nesta tese como sistema de representação da informação, constituído por tesouros, vocabulários controlados e metadados, como ferramentas para representação e recuperação da informação, visando o benefício de usuários humanos, é possível destacar seus usos enquanto estruturas de dados, utilizadas igualmente para representação e recuperação da informação, contudo, somadas de integração e, interoperabilidade, geradas a partir da interoperabilidade semântica, em benefício de consumidores, sejam eles humanos, aplicações ou agentes computacionais.

Além disso, quando a informação é adequadamente armazenada em um banco de dados, utilizando estrutura de dados, é possível a aplicação de ontologias e outras ferramentas que possibilitam gerar inclusive bases de conhecimento, com aportes e em benefício da Inteligência Artificial e da Ciência de Dados.

Com isso, destaca-se que os benefícios gerados pelo sistema de representação da informação para usuários-máquinas, em primeira instância, devem ser compreendidos como mais uma forma de beneficiar usuários humanos. É notório

⁹⁵ “*Metadata, controlled vocabularies, and thesauri are increasingly becoming the building blocks of most major web sites and intranets. Single-taxonomy solutions are giving way to more flexible, faceted approaches.*”

que, a estrutura adequada da informação, por meio dos dados, utilizando metadados e vocabulários controlados, favorece o processamento automático desses dados em ambientes informacionais digitais únicos e em Ecologias Informacionais Complexas, possibilitando ainda o enriquecimento semântico e a ligação dos dados. E essa estruturação beneficia o uso e o reuso dos dados por aplicações e agentes computacionais, utilizados, em última instância para benefício de usuários humanos.

Desta forma, pode-se afirmar que o contexto dado por Morville e Rosenfeld (2006) e Rosenfeld, Morville e Arango (2015) aos tesouros, vocabulários controlados e metadados, enquanto ferramentas para a representação da informação, utilizadas sobretudo no campo de valor dos metadados, deve igualmente ser adotado de forma padronizada e sistemática na camada de dados dos sistemas de informação e ambientes informacionais digitais, possibilitando a adoção de semântica formal para os dados, enquanto atributo de metadados, que culminará em estruturas de representação mais robustas e globalmente acionáveis. Essa discussão é ampliada na seção 4.4.1.4 desta tese.

Essa estrutura de representação, ou estrutura de dados, é fundamental para a Arquitetura da Informação *bottom-up* e se consolida por meio da Arquitetura de Dados, cujo conceito e elementos são apresentados nesta tese.

3.3 ARQUITETURA DA INFORMAÇÃO: ABORDAGEM INFORMACIONAL

A abordagem da Arquitetura da Informação, designada por Oliveira (2014) como informacional, possui todos os elementos da abordagem sistêmica, e a eles incorpora o arcabouço teórico-conceitual da Ciência da Informação, área da qual os precursores de ambas as abordagens são oriundos.

Resmini e Rosati (2011), designam essa abordagem como Ciência da Informação e destacam a atuação dos precursores:

Rosenfeld e Morville, e aqueles que seguem sua visão inicial de LIS [*Library and Information Science*], devem ser creditados por trazer muitas das metodologias centrais usadas para o projeto de navegação, rotulagem e estrutura do site. Eles ofereceram à florescente comunidade de prática uma abordagem extremamente empírica e prática e, sozinhos, trouxeram a pesquisa do usuário e a engenharia de usabilidade para o núcleo das principais ferramentas de AI. Como Rosenfeld gosta de dizer, eles “certamente abraçaram outras disciplinas”. Ao longo dos anos, sua visão sobre o assunto evoluiu, mas até agora, a visão clássica de AI de Rosenfeld e Morville como o projeto de taxonomias, menus e mapas do site ainda

representa a visão dominante e mais credenciada, especialmente para aqueles fora do campo.⁹⁶ (RESMINI; ROSATI, 2011, p. 28, tradução nossa).

Nesse sentido, já no início da obra *Information architecture for the world wide web*, de Rosenfeld e Morville (1998, tradução nossa), os autores destacam: “Este livro mostra como aplicar princípios de arquitetura e biblioteconomia para projetar websites e intranets coesos que sejam fáceis de usar, gerenciar e expandir.”⁹⁷ E continuam, afirmando que a obra “[...] trata da aplicação dos princípios da arquitetura e da biblioteconomia ao projeto de websites.”⁹⁸

Inegavelmente, Rosenfeld e Morville possuem contribuições imprescindíveis a esta abordagem, sobretudo pela expertise aplicada ao campo da Arquitetura da Informação.

Descobrimos que nossa formação em ciência da informação e biblioteconomia tem se mostrado muito útil para lidar com as relações entre as páginas e outros elementos que compõem um website inteiro. Por definição, os bibliotecários lidam com a organização e o acesso à informação nos sistemas de informação e são capacitados para trabalhar com tecnologias de busca, navegação e indexação. Bibliotecários com visão de futuro (recentemente descritos como *cybrarians*) vêem que sua experiência se aplica a novas áreas não relacionadas ao fornecimento de acesso a informações impressas armazenadas em bibliotecas tradicionais. Portanto, a biblioteconomia é uma disciplina importante a se recorrer para obter experiência em arquitetura da informação.⁹⁹ (ROSENFELD; MORVILLE, 1998, p. 16, tradução nossa).

A afirmação dos autores explicita relacionamentos diretos entre a Biblioteconomia e Ciência da Informação e a Arquitetura da Informação, sobretudo no que tange aos aspectos de organização, representação, navegação, acesso e uso da

⁹⁶ “Rosenfeld and Morville, and those many following along their initial LIS view, must be credited for bringing in many of the core methodologies used for the design of navigation, labeling, and site structure. They offered the blooming community of practice an extremely empirical and practical approach, and they single-handedly brought user research and usability engineering into the core of mainstream IA tools. As Rosenfeld is fond of saying, they “certainly embraced other disciplines.” Through the years their view on the subject evolved, but so far, Rosenfeld and Morville’s classical view of IA as the design of taxonomies, menus, and site maps still represents the mainstream and most accredited view, especially for those outside the field.”

⁹⁷ “This book shows how to apply principles of architecture and library science to design cohesive Websites and intranets that are easy to use, manage, and expand.”

⁹⁸ “[...] about applying the principles of architecture and library science to web site design.”

⁹⁹ “We’ve found that our backgrounds in information science and librarianship have proven very useful in dealing with the relationships between pages and other elements that make up a whole site. By definition, librarians deal with organization of and access to information within information systems and are trained to work with searching, browsing, and indexing technologies. Forward-looking librarians (recently described as *cybrarians*) see that their expertise applies in new arenas unrelated to providing access to printed information stored in traditional libraries. So librarianship is an important discipline to turn to for information architecture expertise. Just remember that librarians are also prone to get lost in details, a weakness that can distract from determining the big picture of a web site.”

informação, em diferentes ambientes informacionais, inclusive nos digitais, e com a atuação voltada ao usuário, parte da tríade da AI.

Oliveira (2014, p. 104) utiliza a Teoria do Conceito (DAHLBERG, 1978b) para identificar os núcleos conceituais da Arquitetura da Informação sob a ótica de Morville e Rosenfeld (2006): “[...] propriedade, formato, estrutura, metadados, volume e dinamismo servem para descrever os conteúdos de forma significativa e facilitar a recuperação dos recursos disponíveis nos ambientes de informação digital.”, e depreende do índice da obra os seguintes núcleos conceituais:

[...] biblioteca, catálogo, Biblioteconomia, Ciência da Informação, gestão da informação, arquivos de autoridade, Classificação Decimal de Dewey, classificação facetada, tesouros, vocabulários controlados, organização da informação, nuvens de informação, modelos informacionais, mapeamento de informações, tecnologia da informação, entre outros. (OLIVEIRA, 2014, p. 106).

Com isso, Oliveira (2014) enfatiza o diálogo da Arquitetura da Informação de Morville e Rosenfeld com aspectos da organização e da representação da informação, parte do arcabouço da Ciência da Informação e Biblioteconomia.

Vidotti, Cusin e Corradi (2008, p. 175) destacam a característica multidisciplinar da Arquitetura da Informação, ao afirmarem

A Arquitetura da Informação é uma metodologia que unifica os métodos de organização, representação, recuperação, acesso e disseminação de informação advindos da Biblioteconomia, com a estruturação espacial da Arquitetura, utilizando-se de Tecnologias de Informação e Comunicação [cuja base advém da Ciência da Computação] para o planejamento de ambientes hipermídia informacionais digitais.

Vidotti, Cusin e Corradi (2008) incorporam aos sistemas de organização, rotulagem, navegação e busca, elementos estruturantes para alicerçá-los, cujos termos foram selecionados por Oliveira (2014, p. 102) utilizando a Teoria do Conceito “[...] organização de conteúdos, armazenamento, preservação, representação, descrição, classificação, metadados, tesouro, vocabulário controlado, recuperação, acesso e uso [...]”.

Esses elementos estruturantes são conceitos utilizados pela Biblioteconomia e Ciência da Informação, o que se constitui na abordagem informacional da Arquitetura da Informação, que possui como norteadores os sistemas da abordagem sistêmica da AI.

Neste sentido, Vidotti, Cusin e Corradi (2008, p. 182) partem dos sistemas contidos na Arquitetura da Informação de Rosenfeld e Morville (1998), consideram outros elementos presentes na anatomia da Arquitetura da Informação em Morville e Rosenfeld (2006) e fazem incorporações e relacionamentos sob o prisma da Ciência da Informação:

[...] no contexto da Ciência da Informação, a Arquitetura da Informação enfoca a organização de conteúdos informacionais e as formas de armazenamento e preservação (sistemas de organização), representação, descrição e classificação (sistema de rotulagem, metadados, tesouro e vocabulário controlado), recuperação (sistema de busca), objetivando a criação de um sistema de interação (sistema de navegação) no qual o usuário deve interagir facilmente (usabilidade) com autonomia no acesso e uso do conteúdo (acessibilidade) no ambiente hipermídia informacional digital. (VIDOTTI; CUSIN; CORRADI, 2008, p. 182).

Vale destacar que Rosenfeld e Morville (1998) já apontavam o uso de vocabulários controlados, sobretudo para a representação dos campos de valor em bancos de dados e para a estruturação do sistema de rotulagem. Em ambos os casos, tratam dos vocabulários controlados e tesouros como sendo fontes úteis e criadas por profissionais da informação, mas os retratam de forma simplista como: “Um vocabulário controlado é simplesmente uma lista de termos predeterminados que descrevem um tópico, como arte ou ciência da computação.” (p. 73).

Essa visão foi ampliada em Morville e Rosenfeld (2006), com o acréscimo de metadados, mas permaneceu tendo ênfase nos processos de representação da informação, no campo de valor dos atributos de metadados.

Vidotti, Cusin e Corradi (2008) enfatizam ainda que a Arquitetura da Informação como metodologia para a estruturação de ambientes informacionais digitais deve torná-los flexíveis e customizáveis, visando possibilitar ao usuário a representação e o relacionamento das informações durante o uso do ambiente. Essa compreensão corrobora a teoria da Ciência de Serviços, aplicada no sentido de permitir ao sujeito a cocriação de valor ao ambiente e à informação, durante a interação, considerando que, durante o consumo da informação, ele também fornece elementos capazes de auxiliar na melhoria do ambiente informacional e dos conteúdos, ao mesmo tempo que pode fornecer informações úteis ao ambiente informacional.

Oliveira (2014) também pode ter suas análises incluídas na abordagem informacional da Arquitetura da Informação, ao discutir, inclusive, as problemáticas relacionadas ao tratamento da informação:

Porém, nos diversos ambientes analógicos e ambientes de informação digital, nem sempre é possível encontrar a informação que se deseja ou realizar a atividade que se objetiva em função de uma inadequada organização e representação da informação, comprometendo a experiência de interação nesses produtos do 'engenho tecnológico'. (OLIVEIRA, 2014, p. 19).

Além destes, Oliveira (2014) aponta que, durante a navegação em um ambiente informacional, a desorientação e o transbordamento cognitivo podem ser atrelados a problemas na Arquitetura da Informação; ao passo que a melhoria do acesso e uso da informação é possibilitada por Arquiteturas da Informação bem projetadas.

Essa discussão apresentada por Oliveira (2014) reforça a necessidade de um novo olhar para a organização e a representação da informação, considerando sempre o usuário, quer seja humano ou, mais recentemente, máquina, para que a Arquitetura da Informação nos ambientes informacionais digitais favoreça o processo de recuperação e Encontrabilidade da Informação, visando acesso e (re)uso.

Para tanto, é evidente que o processamento dos dados representacionais possui papel fundamental, quer seja pela compatibilidade de padrões e aspectos semânticos dos atributos e valores utilizados, ou ainda por protocolos e padrões que possibilitem a localização, o acesso e o (re)uso, acionáveis via interface (*front-end*) ou banco de dados (*back-end*) – considerando aspectos de segurança.

Destaca-se, nesse sentido, a relevância dos Princípios FAIR na estrutura dos dados, para que sejam localizáveis, acessíveis, interoperáveis e reutilizáveis para consumo por aplicações e agentes computacionais.

Nesse sentido, é necessário reforçar a relevância que as TIC alcançaram no contexto da Biblioteconomia e Ciência da Informação, não apenas como suporte à informação, mas como elemento central na coleta, processamento e disseminação. Isso posto, destaca-se a necessidade da atuação da área em aspectos relacionados à estrutura de dados, suas formas de coleta, tratamento, armazenamento, integração, interoperabilidade, uso e reuso, o que se faz por meio da utilização das TIC no nível granular da informação, quer seja nas estruturas de representação ou, ainda, na própria estruturação dos dados, formato e linguagem dos objetos digitais.

Essa emergência da aplicação das TIC na área de Biblioteconomia e Ciência da Informação fora explicitada por Santos e Vidotti (2009), que consideram que a perspectiva de investigação e compreensão das TIC na Ciência da Informação era, até então, muito relacionada a ferramentas, regras e algoritmos, altamente relacionados à Ciência da Computação. Assim, apresentam uma nova perspectiva, baseada na verticalidade, inter e transdisciplinaridade, visando potencializar sua aplicação por meio do “[...] diálogo com os diversos campos do conhecimento e com seu centro no humano.” Assim,

O foco da abordagem [...] está no processo tríplice de investigação de mecanismos de cooperação e colaboração, de neutralização e de ajuste que se estabelecem na percepção das necessidades estruturais para a recepção das representações das informações, tendo como base a percepção da cultura e da sociedade e as projeções tecnológicas. (SANTOS; VIDOTTI, 2009).

Isso posto, esclarece-se que, ainda que as investigações se relacionem às TIC, no contexto da Ciência da Informação e na abordagem informacional da Arquitetura da Informação, o usuário é um elemento central e preponderante da investigação, sendo ele causa e efeito do tratamento informacional.

3.4 ARQUITETURA DA INFORMAÇÃO: ABORDAGEM PERVASIVA

A Arquitetura da Informação Pervasiva tem como precursores Resmini e Rosati (2011), que iniciam uma discussão a fim de ampliar o olhar da Arquitetura da Informação, até então ampla e largamente discutida por Rosenfeld e Morville, sobretudo no domínio da web, considerando os aspectos tecnológicos emergentes, que podem ser utilizados em ambientes informacionais analógicos e digitais. Para tanto, consideram, além da tríade – conteúdo, contexto e usuário –, que a informação pode ser pervasiva e ubíqua, e necessitar de diferentes ambientes e mídias para suportá-la em atendimento às necessidades dos sujeitos que interagem com ela.

No prefácio da obra de Resmini e Rosati (2011), Morville elucida que, nos anos 1990, início do desenvolvimento da internet, o que se esperava era que ela pudesse mudar o mundo, o que voltou a acontecer, na perspectiva dele, àquela época, por meio da computação ubíqua: “Estamos criando experiências do usuário

multicanal, multiplataforma, transmídia e físico-digital que derrubam as barreiras entre as categorias.”¹⁰⁰ (p. xi, tradução nossa).

Morville ressalta ainda a contribuição feita por Rosenfeld e Morville aos arquitetos da informação, na compreensão da web, por meio da Arquitetura da Informação e dos sistemas de organização, rotulagem, navegação e busca, além do que, nesta tese, designa-se sistema de representação da informação. Embora essas questões continuem relevantes, o autor reconhece que novas questões emergiam naquele momento, a partir dos relacionamentos possíveis entre a vivência no mundo físico e por meio da web. Nesse contexto, é lançada a obra de Resmini e Rosati (2011, p. xii, tradução nossa) que, segundo Morville, é um “[...] livro revigorante sobre o projeto de ecossistemas para orientação e compreensão [que] promove uma abordagem holística da arquitetura da informação e da experiência do usuário que extrai insights de várias disciplinas e contextos históricos.”¹⁰¹

Nesse sentido, Resmini e Rosati (2011, p. 52, tradução nossa) afirmam:

Nós, IAs [Arquitetos da Informação], temos que ter um pouco mais de Wurman em nossos bolsos e ir além do livro Urso Polar: à medida que as informações vazam para dispositivos móveis e espaços físicos, a arquitetura da informação não é apenas para a World Wide Web, mas ajuda a projetar todos os espaços informacionais compartilhados, lugares, serviços, e processos que tornam a experiência do usuário possível em primeiro lugar.¹⁰²

A Arquitetura da Informação Pervasiva discutida pelos autores está diretamente relacionada às alterações ocorridas a partir dos anos 2000, quando o advento das tecnologias móveis começou a alterar as formas de viver, produzir e consumir dados, informações e outros produtos e serviços, bem como as relações de trabalho, lazer e relacionamentos.

Deixe para trás as vastas implicações da disseminação da Web social e considere apenas a computação móvel: ser capaz de ser conectado em qualquer lugar significa que não há certeza do contexto físico em que determinada informação é consumida, o que, por sua vez, é um grande desafio de design. A maneira como interagimos, os dados de que precisamos, como nos distraímos com as informações que recebemos e a urgência ou o momento dos avisos ou lembretes

¹⁰⁰ “We’re creating multichannel, cross-platform, transmedia, physiodigital user experiences that tear down the walls between categories”

¹⁰¹ “[...] refreshing book about the design of ecosystems for wayfinding and understanding promotes a holistic approach to information architecture and user experience that draws insights from multiple disciplines and historical contexts.”

¹⁰² “We IAs have to have a little more Wurman in our pockets and move beyond the Polar Bear Book: as information bleeds out to mobile devices and physical spaces, information architecture is not just for the World Wide Web, but helps design all shared informational spaces, places, services, and processes that render the user experience possible in the first place.”

mudam o tempo todo. Mas eles estão sempre lá: nós não simplesmente desligamos o computador e saímos.¹⁰³ (RESMINI; ROSATI, 2011, p. 35, tradução nossa).

Por isso, o que Resmini e Rosati (2011) propõem é uma alteração na Arquitetura da Informação como se conhece, para projetos abrangentes, ecológicos e holísticos que favoreçam a experiência do usuário entre ambientes, lugares, canais, dispositivos,

[...] precisamos reformular a arquitetura da informação para melhor atender a nós e às nossas necessidades em constante mudança. Um grande desafio, de fato, mas onde há um desafio há uma oportunidade. Na verdade, não achamos que essa nova arquitetura de informação seja grande ou pequena: achamos que a arquitetura de informação pervasiva é ampla.¹⁰⁴ (RESMINI; ROSATI, 2011, p. 35, tradução nossa).

Assim, de acordo com Resmini e Rosati (2011, p. 52)

As arquiteturas de informação tornam-se ecossistemas. Quando diferentes mídias e diferentes contextos estão intimamente interligados, nenhum artefato pode permanecer como uma entidade única e isolada. Cada artefato se torna um elemento em um ecossistema maior. Todos esses artefatos têm vários links ou relacionamentos entre si e devem ser projetados como parte de um único processo contínuo de experiência do usuário.¹⁰⁵

No Brasil, Oliveira (2014) discute o que designa abordagens da Arquitetura da Informação e, a partir de Resmini e Rosati (2011), propõe contribuições teórico-conceituais sob a ótica da Ciência da Informação. Para o autor,

Emerge um movimento tecnológico de integração de ambientes físicos com ambientes de informação digital, por meio da criação de camadas informacionais que os intersecciona, facilitando a experiência do sujeito ao percorrê-los. Temos visto ainda um fenômeno ecológico onde uma mesma informação necessita estar acessível em um ambiente analógico, em um site web, em um aplicativo para tablets, em um aplicativo para smartphone ou em uma televisão digital. Compreendemos que essa informação necessita amoldar-se ao contexto e ao dispositivo usado para acessá-la, o que complexifica as

¹⁰³ "Leave the vast implications of the social Web going pervasive behind and consider mobile computing alone: being able to be plugged in on the go means that there is no certainty of the physical context in which a certain piece of information is consumed, which, in turn, is one huge design challenge. The way we interact, the data we need, how we allow ourself to be distracted by the information we receive, and the urgency or timing of warnings or reminders change all the time. But they are always there: we do not simply switch the computer off and walk out."

¹⁰⁴ "[...] we need to reshape information architecture to better serve us and our changing needs. A huge challenge indeed, but where there is a challenge there is an opportunity. If anything, we do not think this new information architecture is big, or little: we think pervasive information architecture is broad."

¹⁰⁵ "Information architectures become ecosystems. When different media and different contexts are intertwined tightly, no artifact can stand as a single, isolated entity. Every artifact becomes an element in a larger ecosystem. All of these artifacts have multiple links or relationships with each other and have to be designed as part of one single seamless user experience process."

questões arquiteturais da informação em cada ambiente, em cada dispositivo e gera a necessidade de repensar a experiência tecnológica para torná-la mais integrada e holística. (OLIVEIRA, 2014, p. 18).

Oliveira (2014), após revisitar autores como Ronda-León (2008) e Resmini e Rosati (2011), apresenta sua visão acerca da Arquitetura da Informação e suas bases inter e transdisciplinares. Assim, entende que Arquitetura, Design, Sistemas de Informação, Ciência da Informação, Computação Ubíqua e Design de Serviços¹⁰⁶ são núcleos conceituais consolidados que norteiam a Arquitetura da Informação. Contudo, Oliveira (2014, p. 84) destaca “[...] a Ciência da Informação tem um papel fundamental na abordagem pervasiva pois, a pervasividade é um atributo da informação [...]”

Para Oliveira (2014), a Arquitetura da Informação Pervasiva (AIP) emerge em um momento caracterizado pela crescente de soluções tecnológicas marcadas por pervasividade e ubiquidade, cujo tratamento por meio das abordagens anteriores da Arquitetura da Informação – arquitetural, sistêmica e informacional – se tornou insuficiente. Para além disso, Resmini e Rosati (2011) ponderam uma alteração no protagonismo do sujeito em ambientes informacionais digitais, cujas experiências necessitam de alicerces na computação ubíqua e no design de serviços.

Os usuários tornam-se intermediários. Os usuários agora estão contribuindo como participantes desses ecossistemas e produzem ativamente novos conteúdos ou corrigem o conteúdo existente por meio de links, *mash-ups*, comentários ou críticas. A distinção tradicional entre autores e leitores, ou produtores e consumidores, torna-se tênue a ponto de ser inútil e desprovida de qualquer sentido.¹⁰⁷ (RESMINI; ROSATI, 2011, p. 53).

Relevante destacar que o protagonismo dos sujeitos passa pela alteração de um papel de usuário para um processo de cocriação, uma vez que o mesmo sujeito que acessa e usa a informação em ambientes informacionais digitais também pode criar informações novas, inclusive a partir do (re)uso. Destaca-se ainda a presença de um usuário não humano, que pode atuar no consumo e fornecimento de dados, bem como no enriquecimento semântico dos dados e informações já disponíveis, utilizando

¹⁰⁶ Resmini e Rosati (2011, p. 34, tradução nossa) definem Design de Serviços, a partir da Wikipedia, como “A atividade de planejar e organizar pessoas, infraestrutura, comunicação e componentes materiais de um serviço para melhorar sua qualidade, a interação entre o provedor de serviços e os clientes e a experiência do cliente.”

¹⁰⁷ “Users become intermediaries. Users are now contributing participants in these ecosystems and actively produce new content or remediate existing content by ways of linking, mash-ups, commentary, or critique. The traditional distinction between authors and readers, or producers and consumers, becomes thin to the point of being useless and void of all meaning.”

estruturas padronizadas, protocolos de comunicação e, quando necessário, conversão de registros.

O advento das tecnologias, sobretudo as computacionais, desde o surgimento da Arquitetura da Informação permanece em aceleração, assim como o fornecimento e consumo de dados e informações que inicialmente se faziam prioritariamente por humanos e atualmente por máquinas, cujo processamento é mais rápido e eficaz, mas que, para a compreensão, dependem da adequada estrutura dos dados que serão processados. Por meio dessas tecnologias, é possível, inclusive, criar ambientes informacionais digitais completamente automatizados no que se refere à coleta e disponibilização dos conteúdos, a partir de múltiplas fontes.

Isso posto, Oliveira (2014, p. 109) considera a Arquitetura da Informação Pervasiva “[...] uma abordagem atual que pondera, entre outros aspectos, os processos de hibridização dos espaços humanos onde os sujeitos vivem, trabalham e divertem-se aos ambientes de informação digital.”

Para o autor, o advento das tecnologias culminou em uma alteração no comportamento dos sujeitos, favorecida pelo acesso a ambientes informacionais digitais inclusive por meio de dispositivos móveis, o que possibilita a interação com ambientes analógicos por meio de ambientes digitais ou com ambientes digitais a partir de ambientes analógicos. Isso se aplica a diferentes aspectos, desde a busca e acesso à informação, até as relações de consumo e mais recentemente, nas relações humanas, podendo, inclusive, serem mediados por agentes computacionais, a exemplo dos *chatbots*, dos sistemas de recomendação, dos estudos de comportamento de consumo ou preferências dos sujeitos.

Para Resmini e Rosati (2011, p. 53), a dinâmica dos ambientes informacionais deu lugar à hibridização, na qual “[...] novas arquiteturas abrangem diferentes domínios (físico, digital e híbrido), diferentes tipos de entidades (dados, itens físicos e pessoas) e diferentes mídias.”¹⁰⁸ Isso possibilita que as experiências sejam multicanais (*multichannel*) ou entre canais (*cross channel*).

Por mais que as fronteiras que separam produtores e consumidores fiquem tênues, o mesmo acontece com as diferentes mídias e gêneros. Todas as experiências podem abranger diferentes ambientes.

¹⁰⁸ “These new architectures embrace different domains (physical, digital, and hybrid), different types of entities (data, physical items, and people), and different media.”

No que tange à Arquitetura da Informação Pervasiva, sob a ótica de Oliveira (2014), cumpre destacar a tríade a ela relacionada:

- a) pervasividade: “[...] a informação digital invadiu a sociedade e a cultura, de modo que está presente nos espaços, ambientes analógicos e digitais, lugares web e não-web, em diversos tipos de dispositivos e modifica a vida os sujeitos [...]” (p. 124); “[...] pervasivo se refere em primeiro lugar à informação e lhe fornece a qualidade de ser pervasiva, de ser penetrante, de ser extensível, de alastrar-se nos espaços, ambientes, dispositivos tecnológicos da ecologia e incorporar-se aos comportamentos dos sujeitos.” (p. 136);
- b) ubiquidade: “A informação é ubíqua e está incorporada aos múltiplos espaços, ambientes e comportamentos das pessoas.” (p. 125);
- c) responsividade: “[...] a informação digital que penetra nos mais diversos produtos tecnológicos da pós-modernidade - Notebook, Netbook, Tablets, smarthphones, painéis digitais, televisão digital, outdoor digital, entre outros - dependendo das características do dispositivo e da capacidade de seus ambientes de informação se moldarem ao contexto e a informação, os sujeitos poderão utilizar melhor a informação digital fazendo pontes entre esses dispositivos e seus ambientes [...]” (p. 125).

Nesse sentido, “A ideia de pervasivo engloba a ideia de ubíquo quando nos referimos à informação e a ideia de responsivo se refere [...] às tecnologias.” (OLIVEIRA, 2014, p. 125).

Diante disso, de forma divergente das abordagens arquitetural, sistêmica e informacional, cujos elementos norteadores se aplicam à informação e à estrutura dos ambientes informacionais projetados, a abordagem pervasiva apresentada por Oliveira (2014) atrela a pervasividade à informação e a responsividade ao sistema de informação/ambiente informacional.

Nesse sentido, é relevante destacar que, na Arquitetura da Informação Pervasiva, a informação é considerada pervasiva, o que torna possível a fluidez necessária para a constituição de Ecologias Informacionais Complexas.

Oliveira (2014, p. 166) conceitua Arquitetura da Informação Pervasiva como:

[...] uma abordagem teórico-prática da disciplina científica pós-moderna Arquitetura da Informação, trata da pesquisa científica e do projeto de ecologias informacionais complexas. Busca manter o senso de localização do usuário na ecologia e o uso de espaços, ambientes

e tecnologias de forma convergente e consistente. Promove a adaptação da ecologia à usuários e aos novos contextos, sugerindo conexões no interior da ecologia e com outras ecologias. Facilita a interação com conjuntos de dados e informações ao considerar os padrões interoperáveis, a acessibilidade, a usabilidade, as qualidades semânticas e a encontrabilidade da informação, portanto deve buscar bases na Ciência da Informação.

A Arquitetura da Informação Pervasiva atua no tratamento de objetos e fenômenos que designa como estrutura informacional ecológica, referindo-se “[...] a um conjunto de espaços, ambientes, canais, mídias, tecnologias e sujeitos, com seus comportamentos, todos interligados e conectados de maneira holística pela informação.” (OLIVEIRA, 2014, p. 134), e para a qual destaca o caráter sistêmico e complexo. Assevera ainda que um dos maiores desafios dessa Ecologia Informacional Complexa é a interoperabilidade, que possibilita o funcionamento sistêmico, dinâmico e integrado indispensável à Arquitetura da Informação Pervasiva.

Vale lembrar que a AI se alicerça na tríade conteúdo, contexto e usuário (MORVILLE; ROSENFELD, 2006; ROSENFELD; MORVILLE; ARANGO, 2015). Isso torna extremamente relevante conhecer o contexto de aplicação para que a AI seja alinhada, única e unívoca, de forma a criar uma identidade. Na Ecologia Informacional Complexa, é necessário se atentar ao conteúdo, entendido de forma ampla, como o conjunto de informações, oriundas de diferentes fontes, com diferentes volumes, estruturas e representações, que precisam ser trabalhadas para que estejam disponíveis e acessíveis no ambiente. E o usuário, cujas necessidades, comportamentos e experiências devem ser compreendidos e considerados, de forma a serem refletidos no ambiente.

Para Oliveira (2014, p. 140), “O contexto na Arquitetura da Informação Pervasiva é necessariamente múltiplo, fluido, flexível, mutável e complexo.” (OLIVEIRA, 2014, p. 140). Tal afirmação considera o princípio de que o todo é maior que a soma das partes e pode considerar que “[...] ao projetar arquiteturas pervasivas é preferível sacrificar detalhes locais e precisão local para uma melhor experiência global do que vice-versa, pois a imprecisão local pode se tornar a precisão global.”¹⁰⁹ (RESMINI; ROSATI, 2011, p. 47).

Ainda, Oliveira (2014, p. 126) afirma que “A Arquitetura da Informação Pervasiva deve utilizar os aparatos tecnológicos de modo que se tornem invisíveis

¹⁰⁹ *“in designing pervasive architectures it is preferable to sacrifice local details and local precision for a better global experience than it is to do vice versa, as local imprecision might become global precision.”*

numa ecologia informacional”. Com isso, a informação se sobrepõe à tecnologia e, de maneira complementar, ao suporte, formato, e demais atributos a ela relacionados. A informação se sobrepõe, ainda, ao próprio ambiente informacional que a disponibiliza, o que a coloca, de fato, como elemento central da Arquitetura da Informação Pervasiva, imprescindível para a constituição de Ecologias Informacionais Complexas.

[...] percebemos que a Arquitetura da Informação Pervasiva encontra-se num estado de emergência que fragiliza qualquer laboração teórica e epistêmica que lhe forneça uma identidade científica autônoma, própria ou desvinculada, ao contrário, só faz sentido refletir sobre o status científico da Arquitetura da Informação Pervasiva percebendo-a no interior da história da Arquitetura da Informação. Em contrapartida, parece ser razoável afirmar que a partir das demandas informacionais e tecnológicas que surgem após anos 2000, a Arquitetura da Informação Pervasiva surge como uma abordagem vinculada à Arquitetura da Informação que goza do status de disciplina científica pós-moderna. (OLIVEIRA, 2014, p. 132).

Diante do exposto nesta seção da tese, tem-se como achado da investigação o reposicionamento da Arquitetura da Informação com abordagem Pervasiva, em função da sua aproximação dos elementos estruturantes da Arquitetura da Informação com abordagem informacional, o que se apresenta como uma morfologia.

3.4.1 Arquitetura da Informação: Reposicionamento da Abordagem Pervasiva para a Informacional

Considerando o percurso investigativo trilhado na seção 3 desta tese, na qual foram revisitados os precursores considerados autores clássicos do campo da Arquitetura da Informação e que se constitui como aparato teórico, tem-se como achado da pesquisa que se ampara no polo teórico e se apresenta como resultado no polo morfológico do estudo, o reposicionamento da abordagem Pervasiva da Arquitetura da Informação, proposta por Oliveira (2014).

Considerando o estudo de Oliveira (2014), verifica-se que a Arquitetura da Informação Pervasiva possui dois elementos centrais:

- a) a informação pervasiva, que transita de forma fluida entre os diferentes ambientes informacionais, que podem ser acessados de forma física e/ou digital, por diferentes dispositivos, utilizando diferentes sistemas operacionais;

- b) a Ecologia Informacional Complexa, que se refere “[...] ao conjunto de espaços, ambientes, canais, mídias, tecnologias e sujeitos com seus comportamentos, todos interligados e conectados de maneira holística pela informação.” (OLIVEIRA, 2014, p. 134).

Assim, no que Oliveira (2014) entende como Arquitetura da Informação Pervasiva, a informação disponível em ambientes informacionais – digitais, analógicos ou híbridos – pode se tornar pervasiva à medida que transita e se propaga na Ecologia Informacional Complexa.

Diante do exposto, e tendo em vista o papel preponderante dos sujeitos enquanto consumidores e fornecedores de informação, apreende-se que a Ecologia Informacional Complexa se utiliza da experiência *cross channel* (RESMINI; ROSATI, 2011) ao possibilitar que o sujeito interaja com a informação em diferentes dispositivos, mídias e canais. Para tanto, utiliza como elo os diferentes sistemas da Arquitetura da Informação Sistêmica que possibilita o acesso dos usuários à informação na Ecologia Informacional Complexa, independente do ambiente, canal ou mídia utilizados para disponibilização e acesso. Permite, ainda, que o usuário reconheça sua localização e mantenha sua orientação na Ecologia durante o acesso e uso da informação.

Neste sentido, pode-se afirmar que, uma vez que a Arquitetura da Informação Pervasiva possui como elementos centrais a informação pervasiva e as Ecologias Informacionais Complexas, ela não se constitui em uma nova abordagem da Arquitetura da Informação. Ao contrário, está posicionada na abordagem Informacional que se utiliza da abordagem Sistêmica, e a incrementa com aspectos tecnológicos emergentes.

Considera-se, assim, que os aspectos tecnológicos emergentes e a característica pervasiva da informação possibilitam a constituição de Ecologia Informacional Complexa.

Essa perspectiva pode ser embasada ainda em Floridi (2010, p. 16, tradução nossa):

O limite entre aqui (analógico, baseado em carbono, off-line) e lá (digital, baseado em silício, online) está rapidamente se tornando indistinto, mas isso é tanto vantajoso para o último quanto para o primeiro. O digital está se espalhando pelo analógico e se fundindo com ele. Esse fenômeno recente é conhecido como 'Computação

Ubíqua', 'Inteligência Ambiental', 'Internet das Coisas' ou coisas aumentadas pela Web'.¹¹⁰

Um projeto ampliado de Arquitetura da Informação contempla uma Ecologia Informacional Complexa e requer diferentes tecnologias que abarquem os ambientes analógico e/ou digital de forma a torna-los híbridos. A característica de unicidade do projeto permite que a Arquitetura da Informação se expanda aos diferentes ambientes, mídias e canais que compõem a Ecologia Informacional Complexa, acessíveis por qualquer dispositivo, de modo a manter o usuário no mesmo projeto de Arquitetura da Informação.

Nesse sentido, Rosenfeld, Morville e Arango (2015, p. 20, tradução nossa) ponderam que

Você não pode projetar produtos e serviços que funcionem de forma eficaz e coerente em vários canais de interação se você não entender como eles influenciam e interagem uns com os outros e com vários outros sistemas que os afetam. [...] cada canal de interação traz à mistura diferentes limitações e possibilidades que devem informar o todo. Uma compreensão abrangente e de alto nível do ecossistema pode ajudar a garantir que seus elementos constituintes funcionem juntos para apresentar experiências coerentes aos usuários.¹¹¹

Assim, o aspecto ecológico trazido pela Ecologia Informacional Complexa, tida como o todo constituído por suas partes – tecnologias, ambientes, comportamentos dos sujeitos – torna o projeto de uma Arquitetura da Informação mais amplo, capaz de tratar a complexidade e o hibridismo dos ambientes, considerando a pervasividade da informação. Destaca-se nesse sentido que, no contexto da Arquitetura da Informação, tendo a informação a característica pervasiva, a Ecologia Informacional é Complexa justamente por se ocupar do conglomerado de ambientes, mídias, sujeitos, tecnologias, com enfoque dinâmico e fluido na informação pervasiva.

Isso se relaciona diretamente com Oliveira (2014, p. 135)

[...] a Arquitetura da Informação Pervasiva deve voltar-se para o tratamento arquitetural de Ecologias Informacionais Complexas, ou seja, novas ecologias informacionais que integram holisticamente

¹¹⁰ "The threshold between here (analogue, carbon-based, off-line) and there (digital, silicon-based, online) is fast becoming blurred, but this is as much to the advantage of the latter as it is of the former. The digital is spilling over into the analogue and merging with it. This recent phenomenon is variously known as 'Ubiquitous Computing', 'Ambient Intelligence', 'The Internet of Things', or 'Web-augmented things'."

¹¹¹ "You can't design products and services that work effectively and coherently across various interaction channels if you don't understand how they influence and interact with one another and with various other systems that affect them. [...] each interaction channel brings to the mix different limitations and possibilities that should inform the whole. A high-level, comprehensive understanding of the ecosystem can help ensure that its constituent elements work together to present coherent experiences to users. As a discipline, information architecture is ideally suited to this task."

espaços, ambientes, tecnologias e os sujeitos com seus comportamentos por meio da informação.

Essa afirmação esclarece que a Arquitetura da Informação Pervasiva (OLIVEIRA, 2014) se refere à arquitetura de Ecologias Informacionais Complexas, e o mesmo ocorre na discussão dos elementos essenciais da AIP. Assim, considerando a essência da Arquitetura da Informação Pervasiva, conforme proposta por Oliveira (2014), é notório que alguns elementos já estavam presentes nas abordagens sistêmica e informacional, tendo estes novas roupagens em função da evolução das tecnologias computacionais. Nesse sentido, o real incremento consiste na Ecologia Informacional Complexa que, tendo como *core* a informação, se integra à Arquitetura da Informação com abordagem Informacional.

O reposicionamento da Arquitetura da Informação Pervasiva dialoga ainda com Resmini e Rosati (2011, p. 52, tradução nossa)

Uma vez que a arquitetura da informação se baseia em princípios que são amplamente independentes de qualquer meio específico - afinal, ela está preocupada com a estruturação do espaço de informação tanto quanto a arquitetura está preocupada com a estruturação do espaço físico - ela fornece um modelo conceitual flexível, mas sólido para o projeto de *crosscontext* e experiências de usuário em vários canais que abrangem diferentes mídias e ambientes [...]. Ao abordar essas questões estruturais, é capaz de fornecer a todos os atores uma estrutura cognitiva constante e coerente ao longo de todo o processo. É importante enfatizar que não se trata de design de interface ou design de interação. Essas são peças valiosas e necessárias do quadro geral, mas geralmente estão relacionadas a pontos de contato únicos, um de cada vez.¹¹²

A proposta de realocação da Arquitetura da Informação Pervasiva encontra lugar ainda em Lacerda (2015, p. 203)

Recomenda-se, ainda, a busca pela consolidação da subdisciplina Arquitetura da Informação Pervasiva, ou até mesmo a incorporação total das teorias e modelos desta na Arquitetura da Informação, quando não houver mais necessidade de diferenciação entre projetos para ecossistemas de informação e projetos para espaços isolados, ao tornarem-se aqueles a realidade presente e inconteste.

¹¹² “Since information architecture relies on principles that are largely independent from any specific medium - after all it is concerned with the structuring of information space as much as architecture is concerned with structuring physical space - it provides a flexible but solid conceptual model for the design of *crosscontext* and *cross-channel* user experiences which span different media and environments [...]. By addressing these structural issues, it is capable of providing all actors with a constant, coherent cognitive framework throughout the whole process. It is important to emphasize that this is not interface design or interaction design. These are both valuable and necessary pieces of the general picture, but they are usually concerned with single touch points, one at a time.”

E, em Rosenfeld, Morville e Arango (2015, p. 59, tradução nossa) para os quais os ambientes informacionais podem apresentar diferentes manifestações, considerando, sobretudo, os dispositivos utilizados para o acesso e a interação. Contudo, os autores alertam que os elementos estruturais e semânticos devem ser mantidos. “A coerência entre diferentes instâncias da arquitetura [da informação] é alcançada pelo uso consistente da linguagem e pelo estabelecimento de uma relação particular, ou ordem, entre os elementos linguísticos que a compõem.”¹¹³

A discussão apresentada na seção 3.3 desta tese aponta que a abordagem informacional da Arquitetura da Informação se utiliza da abordagem sistêmica, e a ela são acrescidos aspectos concernentes ao tratamento e à representação da informação, próprios da área de Biblioteconomia e Ciência da Informação.

Considerando os elementos da Arquitetura da Informação Pervasiva apresentados por Oliveira (2014) é possível verificar uma relação estreita com a Arquitetura da Informação com abordagem Informacional (Quadro 2), com isso, o reposicionamento é possível, e inclui nessa última alguns elementos, ao passo que realiza um processo de inovação incremental de outros já existentes.

Quadro 2 – Comparativo dos Elementos da Arquitetura da Informação Pervasiva e da Arquitetura da Informação

Elemento	Arquitetura da Informação Pervasiva (AIP)	Arquitetura a Informação (AI)	Incremento
Status Científico	Oliveira (2014)	Rosenfeld e Morville (1998) Morville e Rosenfeld (2006) Vidotti, Cusin e Corradi (2008) Oliveira (2014) Rosenfeld, Morville e Arango (2015)	Oliveira (2014) realiza uma releitura da Arquitetura da Informação conforme já consolidada na área da Ciência da Informação e posiciona a necessidade de um olhar atualizado sob o prisma das Tecnologias de Informação e Comunicação.
Ecologia informacional Complexa	Oliveira (2014)	Morville e Rosenfeld (2006)	Morville e Rosenfeld (2006) utilizam o conceito de ecologia da informação para alicerçar a tríade – conteúdo, contexto e usuário – também

¹¹³ “Coherence between different instances of the architecture is achieved by consistent use of language, and by establishing a particular relationship, or order, between the linguistic elements that comprise it.”

Elemento	Arquitetura da Informação Pervasiva (AIP)	Arquitetura a Informação (AI)	Incremento
			<p>considerados, mas de forma expandida em Oliveira (2014).</p> <p>Oliveira (2014) demonstra, por meio da sua investigação, a premência do planejamento macro da Arquitetura da Informação, considerando conteúdo, contexto e usuário, contudo, incorporando mídias, canais, tecnologias, múltiplas formas de acesso e a necessidade de um olhar holístico atualizado para o projeto de AI, que pode se materializar em um ambiente informacional, seja analógico, digital ou híbrido, ou ainda em uma Ecologia Informacional Complexa.</p>
Pervasividade	<p>Resmini e Rosati (2011)</p> <p>Oliveira (2014)</p>		<p>Resmini e Rosati (2011) vislumbram que a informação não é dependente do contexto físico a partir do qual se tem acesso e apontam a necessidade de desenvolver projetos de AI amplos, com vistas à experiência do usuário em diferentes ambientes e canais.</p> <p>Em Oliveira (2014), informação, parte da tríade da AI, é qualificada como pervasiva, considerando sua característica de propagação em diferentes canais, mídias, sistemas, tecnologias e interações. Assim, a informação pervasiva pode se alastrar e se incorporar aos diferentes espaços, ambientes, tecnologias e sujeitos que compõem e/ou interagem com um ambiente informacional e/ou uma Ecologia Informacional Complexa.</p>

Elemento	Arquitetura da Informação Pervasiva (AIP)	Arquitetura a Informação (AI)	Incremento
Ubiquidade	Resmini e Rosati (2011) Oliveira (2014)		A ubiquidade é entendida como onipresença. Assim, a informação permeia a Ecologia Informacional Complexa, constituída por espaços, ambientes, tecnologias e comportamentos dos sujeitos.
Everyware	Resmini e Rosati (2011) Oliveira (2014)	Rosenfeld e Morville (1998) Morville e Rosenfeld (2006) Oliveira (2014) Rosenfeld, Morville e Arango (2015)	Em Oliveira (2014), a convergência se refere à característica de integração holística a sistemas complexos, de forma que o todo seja maior que a soma das partes. Destaca-se que, nos sistemas da AI de Morville e Rosenfeld (2006), a convergência é abordada nos diferentes sistemas, visando assegurar aspectos como navegação e senso de localização. Assim, a tecnologia empregada pela AIP, proveniente da computação ubíqua, pode se refletir na AI, de forma incremental.
<i>Place-making</i> ou senso de localização	Resmini e Rosati (2011) Oliveira (2014)	Rosenfeld e Morville (1998) Morville e Rosenfeld (2006) Oliveira (2014) Rosenfeld, Morville e Arango (2015)	O senso de localização é elemento presente na AI, sobretudo nos sistemas de rotulagem e navegação (ROSENFELD; MORVILLE, 1998; MORVILLE; ROSENFELD, 2006). O incremento dado por Resmini e Rosati (2011) e Oliveira (2014) consiste na expansão do senso de localização de um ambiente informacional para uma Ecologia Informacional Complexa.
Consistência	Resmini e Rosati (2011) Oliveira (2014)	Rosenfeld e Morville (1998) Morville e Rosenfeld (2006)	Consiste no estabelecimento da lógica que determina as categorizações em toda a navegação na Ecologia Informacional Complexa,

Elemento	Arquitetura da Informação Pervasiva (AIP)	Arquitetura a Informação (AI)	Incremento
		Rosenfeld, Morville e Arango (2015)	considerando todos os espaços, ambientes e tecnologias. De acordo com Oliveira (2014, p. 143) “[...] consistência dialoga com os sistemas de rotulagem e sistemas de representação – metadados, tesouros e vocabulários controlados - discutidos por Morville e Rosenfeld (2006).”
Resiliência	Resmini e Rosati (2011) Oliveira (2014)	Morville e Rosenfeld (2006) Vechiato (2013)	A resiliência consiste na capacidade da Arquitetura da Informação se moldar aos usuários e suas necessidades específicas, bem como às buscas contextuais. A base da AI, em Morville e Rosenfeld (2006 p. 4, tradução nossa), a define como “A arte e ciência de moldar produtos e experiências de informação para suportar a usabilidade e a localização.”, cuja implementação é otimizada pelas tecnologias computacionais recentes. Além disso, favorece a descoberta acidental (VECHIATO, 2013).
Redução	Resmini e Rosati (2011) Oliveira (2014)	Rosenfeld e Morville (1998) Morville e Rosenfeld (2006) Rosenfeld; Morville; Arango (2015)	A Ecologia Informacional Complexa se relaciona diretamente com a consistência, visando reduzir o estresse gerado no uso de múltiplos canais e plataformas. Potencializa o uso dos sistemas de organização e rotulagem, visando o agrupamento adequado e a possibilidade de recomendações.
Correlação	Resmini e Rosati (2011)	Rosenfeld e Morville (1998)	Conecta sujeitos e objetos em todas as partes da

Elemento	Arquitetura da Informação Pervasiva (AIP)	Arquitetura a Informação (AI)	Incremento
	Oliveira (2014)	Morville e Rosenfeld (2006) Vechiato (2013) Rosenfeld, Morville, Arango (2015)	<p>Ecologia Informacional Complexa. É potencializada pelas relações hierárquicas.</p> <p>Constitui-se em um incremento ao Sistema de Organização (ROSENFELD; MORVILLE, 1998), ao enriquecer os relacionamentos na AI por meio das Ecologias Informacionais Complexas, independente do dispositivo ou canal utilizado.</p> <p>Tem como efeito a serendipidade e a descoberta, objetivando Encontrabilidade da Informação (VECHIATO, 2013).</p>
Interoperabilidade	Oliveira (2014)	Rosenfeld e Morville (1998) Morville e Rosenfeld (2006) Vidotti, Cusin e Corradi (2008) Oliveira (2014) Rosenfeld, Morville e Arango (2015)	<p>As Ecologias Informacionais Complexas são sistêmicas e necessitam de funcionamento integrado, o que requer o intercâmbio de dados e informações entre suas partes.</p> <p>A interoperabilidade favorece a manutenção do conteúdo, de forma dinâmica, e potencializa a descoberta, a partir do sistema de busca (ROSENFELD; MORVILLE, 1998) e se realiza por meio da estrutura de representação, sobretudo pelos metadados (MORVILLE; ROSENFELD, 2006; VIDOTTI, CUSIN, CORRADI, 2008).</p>
Semântica e web semântica	Oliveira (2014)	Morville e Rosenfeld (2006) Vechiato (2013)	De acordo com Oliveira (2014, p. 150-151), “[...] as questões semânticas estiveram presentes ao longo da história da Arquitetura da Informação, principalmente em Morville e Rosenfeld (2006) que utilizam redes semânticas para o

Elemento	Arquitetura da Informação Pervasiva (AIP)	Arquitetura a Informação (AI)	Incremento
			<p>delineamento de contexto, conteúdos e comportamentos dos usuários e para dar suporte na sistematização da navegação, da organização, da busca, da rotulagem e da representação em um ambiente de informação digital.”</p> <p>De forma complementar, a semântica permite o enriquecimento da informação a partir de conteúdos do próprio ambiente informacional e da interoperabilidade com outros ambientes e Ecologias Informacionais Complexas, e favorece a recuperação e a Encontrabilidade da Informação.</p>
Acessibilidade	Oliveira (2014)	Vidotti, Cusin e Corradi (2008)	<p>A acessibilidade foi apresentada por Vidotti, Cusin e Corradi (2008), que destacam a relevância para que os ambientes informacionais sejam inclusivos e possibilitem autonomia ao usuário. Oliveira (2014) expande essa visão, de ambiente para a Ecologia Informacional Complexa.</p> <p>É necessário apenas repensar que as Ecologias Informacionais Complexas podem ser acessadas por indivíduos de qualquer localidade, de forma que as recomendações globais (e não apenas as nacionais) de acessibilidade devem ser atendidas.</p>
Usabilidade	Oliveira (2014)	Wurman (1996) Rosenfeld e Morville (1998)	A usabilidade está presente nas bases da AI, quando seus precursores já pensavam em aproximar ambientes urbanos e

Elemento	Arquitetura da Informação Pervasiva (AIP)	Arquitetura a Informação (AI)	Incremento
		Morville e Rosenfeld (2006) Vidotti, Cusin e Corradi (2008) Camargo e Vidotti (2011)	informacionais da arquitetura para, dentre outros fatores, melhorar a usabilidade. Na definição de Morville e Rosenfeld (2006, p. 4, tradução nossa), AI é “A arte e ciência de moldar produtos e experiências de informação para suportar a usabilidade e a localização.”
Encontrabilidade	Oliveira (2014)	Morville e Rosenfeld (2006) Vechiato (2013)	Morville e Rosenfeld (2006) apresentam o conceito de <i>findability</i> , como a facilidade de ser localizável, enquanto Vechiato (2013) define Encontrabilidade da Informação, como sendo a intersecção entre as funcionalidades de um ambiente informacional e as necessidades do sujeito. Considera-se que a Encontrabilidade, nos dois sentidos apresentados, dialoga com o objetivo central da AI.

Fonte: Autoria própria.

Conforme elucida o Quadro 2, dos quinze elementos essenciais à Arquitetura da Informação Pervasiva apresentados por Oliveira (2014), treze elementos já eram discutidos na Arquitetura da Informação nas abordagens clássicas, sobretudo nas abordagens sistêmica e informacional. Os elementos pervasividade e ubiquidade são os novos conceitos incorporados à discussão por Resmini e Rosati (2011) e Oliveira (2014).

Ainda de acordo com Oliveira (2014), as abordagens da Arquitetura da Informação podem coexistir, de forma que novas abordagens não extinguem as anteriores. Contudo, a chamada Arquitetura da Informação Pervasiva (OLIVEIRA, 2014) tem como fenômenos a Ecologia Informacional Complexa, cujo elemento central é a informação, sendo esta última o fenômeno da abordagem Informacional da Arquitetura da Informação elucidando que a informação pode ser pervasiva. Tais

afirmações corroboram Lacerda (2015), que tem a AIP como uma especialização da disciplina de AI.

A pervasividade da informação possibilita que ela transite e se propague entre diferentes ambientes, tornando-se mais relevante que o sistema de informação e/ou ambiente informacional que a disponibiliza. Assim, é possível que a informação (conteúdo) requeira múltiplos canais e tecnologias para ser adequadamente disponível e acessível ao usuário, mantendo um único contexto.

Neste sentido, os estudos da presente tese culminaram no reposicionamento da Arquitetura da Informação Pervasiva, não como uma nova abordagem da Arquitetura da Informação, mas como parte da Arquitetura da Informação com abordagem Informacional, considerando que a informação pode ser pervasiva e ubíqua.

Além disso, a maioria dos chamados elementos essenciais da Arquitetura da Informação Pervasiva (OLIVEIRA, 2014) podem ser considerados oriundos das abordagens sistêmica e informacional da Arquitetura da Informação, e aqueles que não estão presentes podem ser incorporados, de forma a atualizá-las a contextos tecnológicos atuais discutidos pelo autor.

Isso possibilita que a Arquitetura da Informação seja expandida para o projeto de Ambientes Informacionais analógicos e/ou digitais e, quando necessário, para o projeto de Ecologias Informacionais Complexas, constituídas pelo conjunto de ambientes e tecnologias necessárias à disponibilização, acesso e uso da informação, por múltiplas plataformas e canais, atendendo às necessidades dos usuários (público-alvo). Esse fator favorece o planejamento de uma Arquitetura da Informação única, composta por diferentes ambientes e tecnologias, considerando a necessidade de manter seus sistemas operantes e consistentes para que a Ecologia Informacional Complexa seja holística e transparente ao usuário.

O que é necessário é uma abordagem sistemática, abrangente e holística para estruturar as informações de uma maneira que facilite a localização e a compreensão, independentemente do contexto, canal ou meio que o usuário emprega para acessá-las. [...] A arquitetura da informação pode ser usada como uma lente para ajudar equipes e indivíduos a obter essa perspectiva.¹¹⁴ (ROSENFELD; MORVILLE; ARANGO, 2015, p. 16-17, tradução nossa).

¹¹⁴ *“What is needed is a systematic, comprehensive, holistic approach to structuring information in a way that makes it easy to find and understand—regardless of the context, channel, or medium the user employs to access it. [...] Information architecture can be used as a lens to help teams and individuals gain this perspective.”*

Na visão de Resmini e Rosati (2011, p. 19, tradução nossa), a Arquitetura da Informação possui um papel imprescindível no desenvolvimento de estruturas globais e holísticas:

Trabalhar na estrutura geral entre os canais permite que os projetistas introduzam padrões cognitivos (ou interações) globais e constantes para que artefatos únicos, pontos de contato únicos no processo, possam ser usados e explorados sem forçar os usuários a aprender ou reaprender vários comportamentos divergentes, mesmo quando as interfaces individuais se diferem.¹¹⁵

Evidencia-se com isso que, quando determinado processo ou demanda está centrado na informação, é necessário que haja um fluxo, integrado, coerente e consistente – inclusive no uso de terminologia, layout e padrões. Com isso, o usuário não necessita explorar diferentes modos de realizar uma mesma ação, o que o torna mais confiante e seguro no acesso à informação, independente do dispositivo, mídia ou canal (RESMINI; ROSATI, 2011). Inegavelmente, esse tratamento favorece ainda a compreensão e reduz a ansiedade da informação.

Rosenfeld, Morville e Arango (2015) destacam que o avanço das tecnologias possibilitou que a informação fosse dissociada do artefato que a continha, da mesma forma que alterou os contextos de uso. Assim, nos ambientes digitais, é possível ainda coletar dados no momento do uso de determinada informação e, com isso, possibilitar melhorias no que é oferecido, bem como criar novas ofertas. Assim, para os autores, as organizações devem considerar que o acesso à informação acontece em diversos contextos, utilizando diferentes dispositivos que possibilitam formas interação próprias, e que os usuários “[...] obviamente vão querer que essas experiências sejam consistentes e coerentes, independentemente de onde e como a informação está sendo acessada.”¹¹⁶ (p. 16, tradução nossa).

Resmini e Rosati (2011) elucidam ainda que os princípios da Arquitetura da Informação são independentes do meio e se referem ao ambiente informacional, o que possibilita o planejamento de experiências do usuário entre canais, contextos, mídias e ambientes. Nesse sentido, vale ressaltar que o tratamento adequado da informação por meio da abordagem informacional, bem como dos sistemas contidos na abordagem sistêmica, possibilita que a informação pervasiva transite por esses

¹¹⁵ “Working on the general structure across channels allows designers to introduce global, constant cognitive (or interaction) patterns that single artifacts, single touch points in the process, can use and exploit without forcing users to learn or relearn multiple diverging behaviors even when individual interfaces differ.”

¹¹⁶ “[...] obviously want these experiences to be consistent and coherent regardless of where and how the information is being accessed.”

canais, contextos, mídias e ambientes, que compõem a Ecologia Informacional Complexa, mantendo a consistência necessária para a adequada experiência do usuário.

Para tanto, o projeto da Arquitetura da Informação deve considerar todos os ambientes informacionais, canais e mídias necessários – considerando sempre conteúdo, contexto e usuário – de forma a estabelecer o fluxo adequado, para que o usuário se mantenha na Ecologia Informacional Complexa (independente do canal, mídia ou ambiente), reconheça os mesmos elementos de organização, rotulagem, busca e tenha a navegação favorecida entre os ambientes e canais que compõem a Ecologia.

Quando dizemos que a arquitetura da informação precisa estruturar o processo, subimos um degrau na escada da abstração, onde a arquitetura da informação é menos um conjunto específico de ferramentas para, digamos, projeto da Web e mais um conector de projeto entre canais e contextos.¹¹⁷ (RESMINI; ROSATI, 2011, p. 52, tradução nossa).

Essas conexões podem ser estruturadas por meio de fluxos de informação, uma vez que, de acordo com Resmini e Rosati (2011, p. 61, tradução nossa), “A arquitetura de informação pervasiva é uma metodologia [...] que se concentra no projeto dos fluxos de informação subjacentes às ecologias ubíquas.”¹¹⁸, considerando o processamento da informação não apenas em um ambiente informacional, mas espalhada, de forma pervasiva.

Ainda, como já mencionado, Rosenfeld, Morville e Arango (2015) alertam que a Arquitetura da Informação deve prever estruturas semânticas que possam ser instanciadas em diferentes canais e assegurar que a experiência de uso seja coerente. “Embora as capacidades e limitações de cada canal sejam diferentes, as estruturas semânticas empregadas em cada um deles devem ser familiares e consistentes.”¹¹⁹ (p. 18, tradução nossa).

Entendendo a relevância da informação pervasiva no contexto da(s) Ecologia(s) Informacional(is) Complexa(s), a estruturação da informação, seja nos

¹¹⁷ “When we say that information architecture needs to structure the process, we move one step up the ladder of abstraction, where information architecture is less of a specific set of tools for, say, Web design and more of a design connector between channels and contexts.”

¹¹⁸ “Pervasive information architecture is a heuristic methodology that focuses on the design of the information flows that underlie ubiquitous ecologies.”

¹¹⁹ “While the capabilities and limitations of each channel are different, the semantic structures employed in each of them should be familiar and consistent.”

metadados e/ou no objeto digital, é imprescindível para favorecer a pervasividade entre diferentes canais, mídias e tecnologias, bem como na interligação de diferentes Ecologias Informacionais Complexas, por meio de informações, uma vez que isso se realiza por meio da estrutura granular da informação, os dados.

Com isso, vislumbra-se que, por meio da Arquitetura de Dados, seja possível tornar os dados localizáveis, acessíveis, interoperáveis e reutilizáveis, favorecendo a aplicação de estruturas computacionais que realizem o (re)uso de dados para fornecimento e o consumo, bem como para a constituição de Ecologias Informacionais Complexas.

De igual maneira, na infraestrutura da camada de dados dos ambientes informacionais digitais e dos sistemas de informação, necessita-se de estruturação para que os dados sejam acionáveis, interoperáveis e integráveis, o que constitui Ecologias Complexas de Dados, conforme apresentada nesta tese.

3.5 ECOLOGIA INFORMACIONAL COMPLEXA

Embora a Ecologia Informacional Complexa, sob a ótica de Oliveira (2014), esteja contida na Arquitetura da Informação Pervasiva como um dos elementos essenciais, sua relevância para esta pesquisa requer destaque, de forma que é abordada em seção a ela dedicada.

A abordagem da Arquitetura da Informação realizada por Rosenfeld e Morville (1998), Morville e Rosenfeld (2006) e Rosenfeld, Morville e Arango (2015) considera que os ambientes informacionais não são estáticos e que, ao contrário, são dinâmicos e orgânicos, apresentados como sistemas complexos e adaptáveis, tendo em vista os fluxos de informação que atuam interna e externamente.

Nesse contexto, recorrem à Ecologia da Informação de Davenport (1998) e afirmam que essa é composta por conteúdo, contexto e usuário, e que, por isso, quando empregada no desenvolvimento do projeto de AI, precisa equilibrar a complexidade da tríade. Além disso, destacam que, assim como o projeto de Arquitetura da Informação, cada Ecologia da Informação é única.

No contexto desta tese, é importante posicionar que, para Davenport (1998) há muitos problemas relacionados ao fluxo e à gestão da informação, para os quais acredita que as tecnologias da informação sejam a solução, muitas vezes, deixando

de se atentar às necessidades dos usuários. Tal afirmação é justamente o oposto do que se apregoa na Arquitetura da Informação.

Diferentemente dessa abordagem, de acordo com Davenport (1998, p. 14), “[...] a ecologia da informação baseia-se na maneira como as pessoas criam, distribuem, compreendem e usam a informação.”, de forma que a tecnologia passa a ter um papel secundário, sendo considerada “[...] apenas um dos componentes do ambiente de informação e frequentemente não se apresenta como meio adequado para operar mudanças.”

Assim como fez Wurman, Davenport (1998, p. 21) se utilizou de uma metáfora ao utilizar o termo Ecologia da Informação, uma vez que considera a ecologia como “[...] a ciência de compreender e administrar todos os ambientes [...]”.

Quando começamos a pensar nas muitas relações entrecruzadas de pessoas, processos, estruturas de apoio e outros elementos do ambiente informacional de uma empresa, obtemos um padrão melhor para administrar a complexidade e a variedade do uso atual da informação. Também poderíamos descrever a ecologia da informação como *administração holística da informação* ou *administração informacional centrada no ser humano*. O ponto essencial é que essa abordagem devolve o homem ao centro do mundo da informação, banindo a tecnologia para seu devido lugar, na periferia. (DAVENPORT, 1998, p. 21).

Assim, para Davenport (1998), a Ecologia da Informação requer uma compreensão holística, constituída por quatro atributos-chave: integração de diferentes tipos de informação; reconhecimento de mudanças evolutivas; ênfase na observação e na descrição; e ênfase no comportamento pessoal e informacional.

Para Hagedorn (2000, p. 5, tradução nossa) ecologia da informação é “A rede de relacionamentos que compõe um espaço de informação. As partes de uma ecologia da informação são os conteúdos, as ferramentas criadas para alavancar os conteúdos, o contexto do conteúdo e os usuários que acessam.”¹²⁰

Morville e Rosenfeld (2006, p. 24, tradução nossa) consideram que os websites e intranets – que na atualidade generalizamos como ambientes informacionais digitais – são dinâmicos, justamente pela natureza orgânica. Nesse sentido, alicerçam a Arquitetura da Informação na tríade conteúdo, contexto e usuário, e afirmam que “Usamos o conceito de uma “ecologia da informação” composta de

¹²⁰ “The network of relationships that makes up an information space. The pieces of an information ecology are the content, the tools created to leverage the content, the context of the content and the users who access the content.”

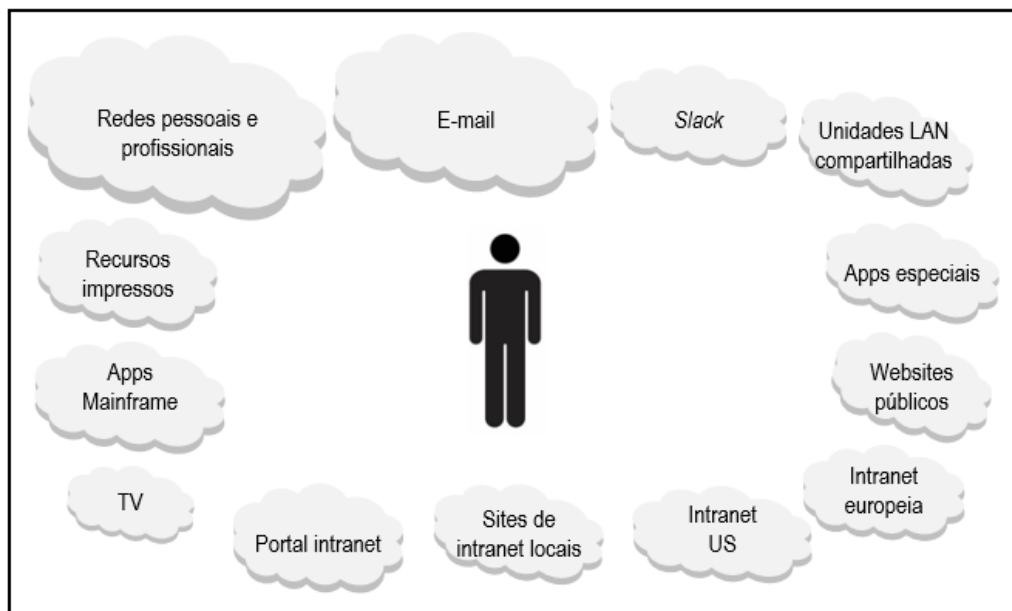
usuários, conteúdo e contexto para lidar com as complexas dependências que existem.”¹²¹.

Em Rosenfeld, Morville e Arango (2015, p. 32), essa afirmação é reforçada e os autores se referem ao diagrama apresentado na Figura 3 para representar os relacionamentos existentes na tríade da AI, na qual há interdependência entre conteúdo, contexto e usuário no que chamam de ecologia de informação complexa e adaptativa.

Assim, considerando que cada projeto de AI é único, e também em razão da tríade, Rosenfeld, Morville e Arango (2015, p. 32) entendem que cada ecologia informacional também é única. Na visão dos autores, tais ecologias informacionais são constituídas por facetas distintas, compostas por: propriedade, formato, estrutura, metadados, volume e dinamismo.

A Figura 5, constante da obra de Rosenfeld, Morville e Arango (2015), representa visualmente em um diagrama conceitual como os funcionários veem a ecologia informacional de uma empresa. Desta forma, o elemento central do diagrama é o próprio funcionário.

Figura 5 – Diagrama conceitual de como os funcionários veem a Ecologia da Informação da empresa



Fonte: Rosenfeld, Morville e Arango (2015, p. 373, tradução nossa).

¹²¹ “We use the concept of an “information ecology” composed of users, content, and context to address the complex dependencies that exist.”

No referido diagrama, cada nuvem corresponde à importância que o elemento possui na visão dos funcionários, o que justifica os diferentes tamanhos. O principal elemento de uma ecologia informacional de uma empresa, na visão dos funcionários, são as redes pessoais e profissionais. Tal afirmação reforça o posicionamento de que o humano está no centro da ecologia informacional, apresentado por Davenport (1998).

Para Oliveira (2014, p. 134),

A estrutura informacional ecológica diz respeito ao conjunto de espaços, ambientes, canais, mídias, tecnologias e sujeitos com seus comportamentos, todos interligados e conectados de maneira holística pela informação. De forma complementar, a ecologia composta pelos componentes supracitados é organizada de forma sistêmica, pois o que chamamos de ecologia é um todo articulado pelo conglomerado das partes. Considerando a heterogeneidade das partes e que os sujeitos atuam em/sobre cada parte, as relações intra e extra-ecológicas são complexas no sentido moriniano do termo, ou seja, a ecologia é uma estrutura sistêmica complexa.

As Ecologias Informacionais Complexas, considerando a abordagem de Oliveira (2014) e Oliveira e Vidotti (2016), são percebidas como um tecido plural e diversificado que exige uma profunda reflexão, uma vez que possibilitam uma nova forma de projetar a Arquitetura da Informação, não apenas como um ambiente, mas como um conjunto de ambientes que dialogam e usam tecnologias que possibilitam se conectar, e que podem ter a mesma informação representada e estruturada individualmente, sendo, contudo, possível integrá-la.

A Ecologia Informacional Complexa se sustenta quando há diferentes ambientes informacionais e canais que os interligam. Do ponto de vista visual, na camada da interface, isso ocorre por meio da padronização da identidade visual – por exemplo com folhas de estilo, tipologias, localização de marcas, cores –, mas ao mesmo tempo com elementos que os diferenciem, para que o usuário reconheça que permanece na mesma Ecologia Informacional Complexa e identifique elementos que possibilitem o reconhecimento da sua localização, ainda que a navegação alterne entre os ambientes informacionais. Além disso, deve possibilitar ao usuário, durante a navegação, retornar ao conteúdo ou ambiente informacional de origem e, ainda, sair da Ecologia.

Evidencia-se, assim, que a Ecologia Informacional Complexa, como um todo constituído de partes, precisa estar perceptível em qualquer ponto da navegação. Neste ponto, cumpre destacar que o reposicionamento da Arquitetura da Informação

Pervasiva para a Arquitetura da Informação com abordagem Informacional se reflete na Ecologia Informacional Complexa, que assume lugar na Arquitetura da Informação. Tal reposicionamento integra o polo morfológico desta tese.

Nesse reposicionamento da Arquitetura da Informação Pervasiva para a Arquitetura da Informação com abordagem Informacional, levam-se em conta os aspectos tecnológicos emergentes e a característica pervasiva da informação.

Vale lembrar que, para Rosenfeld, Morville e Arango (2015), o projeto de Arquitetura da Informação se aplica a diferentes mídias, dispositivos e canais, o que reforça que a Ecologia Informacional Complexa se faz por meio de um projeto de Arquitetura da Informação ampliado, de forma a considerar todos os ambientes, canais, mídias e tecnologias necessários para conectar o usuário ao conteúdo, no contexto atual de crescente disponibilidade de informação e tecnologias.

Assim, no projeto de Arquitetura da Informação, quando a informação for pervasiva, há a necessidade de um projeto ampliado de AI, que se constitui em uma Ecologia Informacional Complexa que, segundo Oliveira (2014), sustenta-se no conjunto de ambientes, canais, mídias, tecnologias, considerando sujeitos e seus comportamentos, para que se interliguem de forma holística. Nesse sentido, a teia plural que possibilita a interligação holística entre ambientes, canais, mídias, tecnologias e sujeitos é o projeto ampliado de Arquitetura da Informação, por meio do qual as características do projeto serão refletidas em toda a Ecologia Informacional Complexa, de forma a manter o usuário no mesmo projeto de Arquitetura da Informação.

Esse breve percurso teórico evidencia a complexidade na constituição de Ecologias Informacionais Complexas, ainda que atue apenas na interface. Oliveira (2014) aponta como ponto nevrálgico da Ecologia Informacional Complexa a interoperabilidade, uma vez que a informação está armazenada em diferentes sistemas de informação.

As ecologias informacionais são sistêmicas e complexas, de modo que um dos maiores desafios da Arquitetura da Informação Pervasiva é interoperar os blocos sistêmicos complexos promovendo o funcionamento dinâmico e integrado da ecologia. (OLIVEIRA, 2014, p. 148).

Isso posto, há uma lacuna que consiste na estruturação da camada de dados dos ambientes informacionais e sistemas de informação, de forma que os dados possam ser acessados, integrados e interoperados, para que propiciem o consumo

para a constituição de Ecologia(s) Informacional(is) Complexa(s). Isso requer um tratamento da camada de dados dos ambientes informacionais e sistemas de informação para que os dados possam estar adequadamente estruturados. A este processo, cuja estruturação é um dos fenômenos desta pesquisa, conceitua-se como Arquitetura de Dados.

Neste sentido, a adequada estruturação da camada de dados dos ambientes informacionais e sistemas de informação ocorre por meio da Arquitetura de Dados, que favorece a interoperabilidade de dados entre ambientes informacionais e sistemas de informação para a constituição de Ecologia(s) Complexa(s) de Dados.

Assim, a estrutura adequada da camada de dados dos ambientes informacionais e sistemas de informação, utilizando Arquitetura de Dados e/ou Ecologia Complexa de Dados, favorece o consumo dos dados para a estruturação de projetos de Arquitetura da Informação e de Ecologia Informacional Complexa concebidas a partir do consumo da camada de dados, o que ocorre para a apresentação do conteúdo e envolve todos os sistemas da Arquitetura da Informação – organização, rotulagem, navegação, busca e representação – na perspectiva da *AI bottom-up*.

Tais afirmações se relacionam e materializam a afirmação de Morville e Rosenfeld (2006, p. 193, tradução nossa):

Um website é uma coleção de sistemas interconectados com dependências complexas. Um único link em uma página pode fazer parte simultaneamente da estrutura, organização, rotulagem, navegação e sistemas de busca do site. É útil estudar esses sistemas de forma independente, mas também é crucial considerar como eles interagem. O reducionismo não nos dirá toda a verdade.¹²²

Diante do exposto, a discussão realizada acerca das Ecologias Informacionais Complexas no âmbito da Arquitetura da Informação contribui para investigar os relacionamentos entre Arquitetura da Informação e Arquitetura de Dados, conforme se apresenta na seção 5, em atendimento ao objetivo específico d.

¹²² “A web site is a collection of interconnected systems with complex dependencies. A single link on a page can simultaneously be part of the site’s structure, organization, labeling, navigation, and searching systems. It’s useful to study these systems independently, but it’s also crucial to consider how they interact. Reductionism will not tell us the whole truth.”

3.6 ARQUITETURA DA INFORMAÇÃO: REFLEXÕES

As interfaces dos ambientes informacionais são responsáveis pela conexão entre o usuário e o conteúdo disponível, conexão essa cujo projeto e implementação ocorre por meio da Arquitetura da Informação *top-down*. Assim, o projeto e a estrutura do ambiente objetivam suportar a interação entre usuários e conteúdos em ambientes informacionais.

A Arquitetura da Informação *top-down* é

O processo de desenvolvimento de uma arquitetura da informação com base na compreensão do contexto, do conteúdo e das necessidades do usuário. Isso envolve a determinação do escopo do site e a criação de plantas e maquetes detalhando o agrupamento e rotulagem das áreas de conteúdo.¹²³ (HAGEDORN, 2000, p. 7, tradução nossa).

Rosenfeld, Morville e Arango (2015) esclarecem que, no início da AI, os projetos de ambientes informacionais eram realizados previamente à sua disponibilização, de forma que a AI *top-down* fornecia grande contributo. Contudo, a dinâmica dos ambientes, conteúdos e mecanismos de busca destacou a relevância da AI *bottom-up*.

A Arquitetura da Informação *bottom-up* é definida como

O processo de desenvolvimento de uma arquitetura de informação com base na compreensão do conteúdo e das ferramentas usadas para alavancar esse conteúdo (por exemplo, pesquisa, índices). Isso envolve a criação de componentes, os bancos de dados para contê-los e os procedimentos para sua manutenção.¹²⁴ (HAGEDORN, 2000, p. 2, tradução nossa).

No contexto da Arquitetura da Informação *bottom-up*, os componentes consistem em metadados – atributo e representação –, objetos e conteúdos. Destacam Rosenfeld, Morville e Arango (2015, p. 85, tradução nossa) que “O próprio conteúdo pode ter arquitetura da informação incorporada a ele.”¹²⁵ e que sua estrutura, sequenciamento e marcação ajudam na recuperação e no uso do conteúdo,

¹²³ “The process of developing an information architecture based on an understanding of the context of the content and the user needs. This involves determining the scope of the site and the creation of blueprints and mockups detailing the grouping and labeling of content areas.”

¹²⁴ “The process of developing an information architecture based on an understanding of the content and the tools used to leverage that content (e.g., search, indexes). This involves the creation of building blocks, the databases to contain them and the procedures for their maintenance.”

¹²⁵ “Content itself can have information architecture embedded within it.”

uma vez que os usuários passam a realizar buscas em mecanismos de busca externos e não no próprio ambiente informacional.

Como mencionado, Morville e Rosenfeld (2006) reconhecem que, inicialmente, a Arquitetura da Informação possuía maior ênfase na abordagem *top-down*, por meio do planejamento de soluções estáveis. Contudo, a consolidação do campo levou a avanços e apontou a necessidade de maior atenção à abordagem *bottom-up*, pensando em soluções emergentes e adaptáveis. Nesse sentido, os autores traçaram um quadro comparativo (Quadro 3) entre o que designam Arquitetura da Informação Clássica (*top-down*) e Arquitetura da Informação Moderna (*bottom-up*).

Quadro 3 – Arquitetura da Informação Clássica e Arquitetura da Informação Moderna

AI Clássica	AI Moderna
Prescritiva	Descritiva
Top-Down	Bottom-Up
Planejada	Emergente
Estável	Adaptativa
Centralizada	Distribuída

Fonte: Morville e Rosenfeld (2006, p. 388, tradução nossa).

Esse quadro comparativo permite vislumbrar que a Arquitetura da Informação clássica é insuficiente para alicerçar ambientes informacionais digitais dinâmicos, e ainda mais limitada quando a informação, por sua característica de pervasividade, é colocada no centro da Ecologia Informacional Complexa, formada por canais, ambientes, tecnologias e sujeitos (OLIVEIRA, 2014) ou no reuso das estruturas de representação para a criação de ambientes informacionais (TORINO; CONEGLIAN; VIDOTTI, 2020).

Tem-se aqui um aspecto relevante no reposicionamento da Arquitetura da Informação Pervasiva para a abordagem Informacional da Arquitetura da Informação, uma vez que o projeto de Arquitetura da Informação estará focado em conteúdo, contexto e usuário, com proeminência da informação em relação ao sistema de informação ou ambiente informacional, o que possibilita, por meio da análise da tríade, considerar a necessidade de desenvolvimento de um ambiente informacional digital ou de uma Ecologia Informacional Complexa, composta por ambientes, canais e

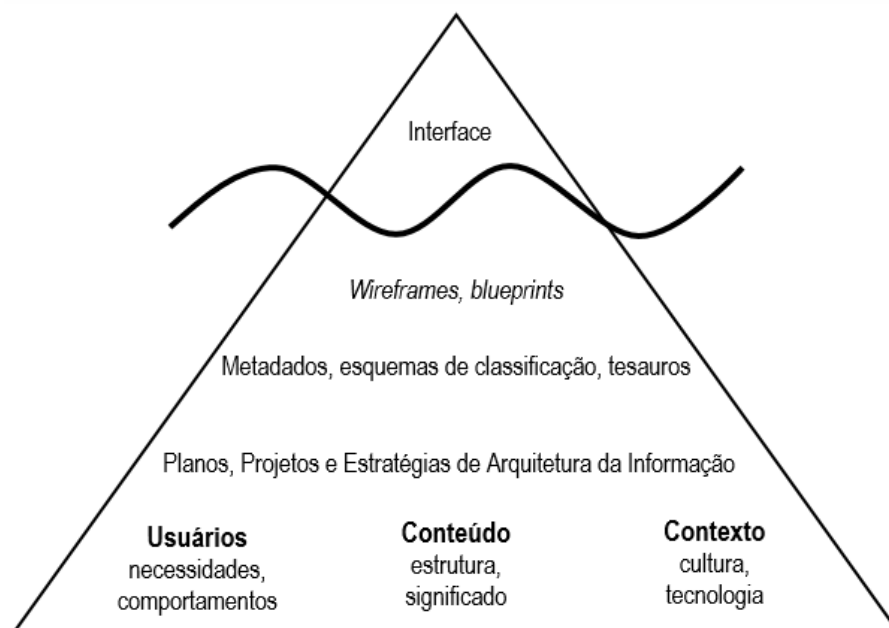
tecnologias para favorecer o acesso e uso da informação por diferentes sujeitos, considerando suas necessidades e interações.

Por isso, é importante que o desenvolvimento da Arquitetura da Informação *top-down* e o da Arquitetura da Informação *bottom-up* sejam integrados, favorecendo a navegação no próprio ambiente e em ambientes relacionados.

Neste contexto, um ambiente informacional desenvolvido a partir de um projeto de Arquitetura da Informação favorece a localização da informação independente de uma necessidade de informação que gera busca intencional ou a descoberta acidental. Além disso, considerando a AI *top-down* e a AI *bottom-up*, isso pode ocorrer na navegação ou busca no próprio ambiente informacional, ou a partir de uma busca ou navegação em outro ambiente informacional que direcione o usuário para a informação localizada.

Para Morville e Rosenfeld (2006), o iceberg da Arquitetura da Informação (Figura 6) mostra que a parte visível da Arquitetura da Informação é a menor parte, visível aos usuários humanos por meio da interface. Por outro lado, toda a parte que alicerça a Arquitetura da Informação permanece submersa e invisível aos usuários humanos.

Figura 6 – Iceberg da Arquitetura da Informação



Fonte: Morville e Rosenfeld (2006, p. 390, tradução nossa).

Na visão dos autores, a parte invisível do iceberg da Arquitetura da Informação é constituída por sua tríade – usuários, conteúdo e contexto – estratégias e planejamento, sistema de representação da informação e as ferramentas que culminam nos entregáveis da Arquitetura da Informação.

Há de se considerar que esse iceberg é incompleto na representação de toda a complexidade existente na estruturação de um ambiente informacional digital, justamente por tratar como invisível apenas os elementos que são parte do projeto e dos fluxos informacionais, que farão parte da sua camada visível, a interface. Contudo, há ainda uma camada mais baixa que é necessária para a estruturação e o armazenamento dos dados que serão acionados pela camada superior.

Importante lembrar que, no contexto da Arquitetura da Informação, são considerados usuários os humanos que acessam e utilizam os ambientes, sobretudo, a partir de uma interface gráfica. E, embora a AI utilize em seus sistemas a organização, a representação, a rotulagem e a busca, o trabalho da AI se concentra no planejamento e estruturação para a entrega de informações em uma interface para usuários humanos.

Neste sentido, diante dos avanços e tecnologias atuais, de crescente disponibilização de dados, no fenômeno conhecido como *Big Data*, bem como dos sistemas de informação e ambientes informacionais digitais que atendem a finalidades específicas, tem-se amplos volumes de dados gerados, armazenados e disponibilizados, o que torna crescente a relevância de planejar a estruturação da camada de dados dos ambientes, visando, de forma análoga ao que se faz na Arquitetura da Informação, trabalhar a arquitetura na camada dos dados, por meio do que se conceitua nesta tese como Arquitetura de Dados.

Essa preocupação com a Arquitetura de Dados traz inegavelmente benefícios para usuários humanos e para o desenvolvimento da AI, ao passo que beneficia, ainda, usuários máquinas (aplicações e agentes computacionais) ao favorecer o (re)uso de dados para diferentes finalidades, por meio de ligações sintáticas e semânticas que podem, inclusive, contribuir para o surgimento de ambientes que não necessitem da intervenção humana para disponibilização de conteúdo. Essas ligações de dados de múltiplos ambientes ocorrem por meio do que se designa nesta tese como Ecologia Complexa de Dados.

As discussões apresentadas na seção 3 desta tese permitem compreender que a Arquitetura da Informação *bottom-up* não é suficiente para estruturar a camada

de dados dos ambientes informacionais digitais. Tal fato reforça a necessidade de investigar e compreender a Arquitetura de Dados no contexto da Ciência da Informação, a fim de aproximá-la da Arquitetura da Informação já consolidada na área.

Além disso, na obra de Morville e Rosenfeld (2006), alguns aspectos relacionados ao processamento da informação são trabalhados, sobretudo no sistema de busca e em tesouros, vocabulários controlados e metadados. Nesse sentido, é perceptível que, à medida que a área de Ciência da Informação avança, as discussões acerca das tecnologias se tornam mais presentes e, considerando que as tecnologias são hoje indissociáveis da sociedade como meio para acesso e uso da informação, essa apropriação pela área de CI se faz não apenas útil, como necessária.

Ainda, no que tange à Arquitetura da Informação e suas práticas, Morville e Rosenfeld (2006) posicionam a necessidade de profissionais da Ciência da Informação no planejamento, implantação, uso e avaliação de sistemas de informação e ferramentas tecnológicas. Contudo, deixam claro que há especificidades que são melhor trabalhadas por especialistas de áreas correlatas, como a Ciência da Computação, no caso da estrutura dos bancos de dados.

Tais afirmações abrem um leque de possibilidades de atuação para os referidos profissionais, com destaque para os da Ciência da Informação em aspectos relacionados à informação e tecnologia – e, com destaque, nesta tese, ao tratamento dos dados.

4 ARQUITETURA DE DADOS

O delineamento metodológico desta tese realiza um percurso teórico iniciado na Arquitetura da Informação (seção 3), campo no qual é encontrada a lacuna tratada nesta pesquisa, que se refere à estruturação da camada de dados dos ambientes informacionais digitais. Desta forma, na seção 4, a discussão acerca dessa lacuna é realizada, ainda compondo o polo teórico, mais especificamente a partir de tessituras acerca de dados e informação, antecedentes da Arquitetura de Dados e da Teoria do Conceito, sendo essa última utilizada como técnica para a criação do conceito de Arquitetura de Dados no contexto da Ciência da Informação – que possui Elementos, Entregáveis, Tríade e Fluxo – e que, juntamente com os conceitos operacionais de Dados e de Ecologia Complexa de Dados, que se apresentam como parte do polo morfológico do estudo.

Um dos precursores da Arquitetura da Informação, Wurman, em sua obra *Information Anxiety 2*, afirma que “Compreender a informação é poder.”¹²⁶ (WURMAN, 2000, p. 286, tradução nossa). Neste sentido, esclarece que não basta acessar a informação, é necessário compreendê-la.

A compreensão da informação por humanos ocorre por processamento cognitivo (de forma altamente dependente dos conceitos já assimilados), cuja capacidade de averiguação é subjetiva. Por outro lado, a compreensão por máquinas ocorre por meio do processamento dos dados, do uso de algoritmos, bibliotecas e até mesmo de bases de treinamento, no contexto da Inteligência Artificial. Assim, o processo de compreensão por máquinas depende da estruturação adequada dos dados para que, ao serem localizados, possam ser facilmente acessados e processados, o que possibilita (re)uso em diferentes contextos, e permite que a averiguação seja objetiva.

Para tanto, faz-se necessário atuar no processamento da informação, na sua forma de expressão por meio dos dados na sua forma bruta, visando estruturá-los adequadamente para favorecer a localização, o acesso, a coleta, o processamento, a integração e/ou interoperabilidade, o armazenamento, o uso e a disponibilização para uso futuro.

¹²⁶ “*Understanding information is power.*”

Nesse sentido, Wurman (2000, p. 286, tradução nossa) expressou o que para ele, nos anos 2000, consistia em um sonho, “Estamos à beira do casamento da tecnologia da informação e arquitetura de informação. Nossa extraordinária capacidade de armazenar e transmitir dados fará desse sonho um sonho acordado.”¹²⁷.

Nesta tese, tem-se como pressuposto que o sonho de Wurman, se materializa por meio do que se conceitua como Arquitetura de Dados.

Antes de mais nada, recorre-se a outros precursores da Arquitetura da Informação, mencionados anteriormente – Rosenfeld e Morville (1998); Morville e Rosenfeld (2006) e Rosenfeld, Morville e Arango (2015) – que afirmam haver complexidade em definir Arquitetura da Informação, sobretudo em razão dos desafios da linguagem e da representação. Contudo, apresentam algumas ponderações que são úteis à compreensão das limitações do campo, dentre as quais, para o presente estudo, destacam-se:

Usamos o termo informação para distinguir a arquitetura da informação da gestão de dados e conhecimento. Dados são fatos e números. Os bancos de dados relacionais são altamente estruturados e produzem respostas específicas para perguntas específicas. Conhecimento é o que está na cabeça das pessoas. Os gerentes do conhecimento desenvolvem ferramentas, processos e incentivos para encorajar as pessoas a compartilhar essas coisas. A informação existe no meio confuso. Com os sistemas de informação, geralmente não há uma única resposta "certa" para uma determinada pergunta. Estamos preocupados com informações de todas as formas e tamanhos: websites, documentos, aplicativos de software, imagens e muito mais. Também estamos preocupados com metadados: termos usados para descrever e representar objetos de conteúdo, como documentos, pessoas, processos e organizações.¹²⁸ (MORVILLE; ROSENFELD, 2006, p. 5, tradução nossa).

Embora essas afirmações estejam posicionadas em uma das edições da obra com enfoque primordial em websites, elas também estão presentes na última edição, de Rosenfeld, Morville e Arango (2015, p. 25), cuja discussão extrapola os ambientes

¹²⁷ “We are at the cusp of the marriage of information technology and information architecture. Our extraordinary ability to store and transmit data will make this dream a waking dream.”

¹²⁸ “We use the term information to distinguish information architecture from data and knowledge management. Data is facts and figures. Relational databases are highly structured and produce specific answers to specific questions. Knowledge is the stuff in people’s heads. Knowledge managers develop tools, processes, and incentives to encourage people to share that stuff. Information exists in the messy middle. With information systems, there’s often no single “right” answer to a given question. We’re concerned with information of all shapes and sizes: web sites, documents, software applications, images, and more. We’re also concerned with metadata: terms used to describe and represent content objects such as documents, people, processes, and organizations.”

web, em função da evolução e complexidade dos canais, mídias, dispositivos e tecnologias utilizados em ambientes informacionais digitais.

Na discussão, os autores esclarecem que, na Arquitetura da Informação, o tratamento se dá à informação disponível em ambientes informacionais, sejam eles analógicos ou digitais, acessíveis pela web ou outras plataformas computacionais. Esse tratamento está relacionado ao processo de representação da informação, cujas bases teórico-práticas emergem na Ciência da Informação e se apoiam em estruturas tecnológicas para o armazenamento.

Ainda, apreende-se das afirmações que a Arquitetura da Informação notadamente se ocupa da informação, ainda que na AI *bottom-up* as bases de dados sejam superficialmente abordadas. É compreensível a menção a bases de dados, visto que a informação digital, quando representada ou disponibilizada, necessita de um ambiente digital de armazenamento, por meio do qual será realizado o acesso e o uso.

Arango (2019) destaca “[...] o papel central dos ambientes de informação nas sociedades atuais e observa que projetar efetivamente para esses ambientes exige pensar sob a superfície das estruturas subjacentes a eles.”¹²⁹

Considerando o percurso teórico-conceitual da Arquitetura da Informação, conforme discutido na seção 3 desta tese, intitulada Arquitetura da Informação, é possível afirmar que a Arquitetura da Informação *bottom-up* necessita de uma camada de tratamento e estruturação de dados, para que as informações representadas e os dados armazenados nos sistemas de informação e ambientes informacionais digitais possam ser adequadamente estruturados e acionados quando requeridos, quer seja por usuários humanos ou máquinas. Dessa forma, a Arquitetura da Informação *bottom-up* necessita de uma estrutura formal de alto nível¹³⁰. Além disso, é necessário considerar que a estruturação dos dados se faz por meio da Arquitetura de Dados – conceituada nesta tese –, e todos os elementos que a compõem.

Neste sentido, Morville e Rosenfeld (2006, p. 47, tradução nossa) afirmam que

[...] se você olhar de perto o suficiente, você pode ver a arquitetura da informação mesmo quando ela está incorporada ao seu conteúdo. De fato, ao apoiar a pesquisa e a navegação, a estrutura inerente ao conteúdo permite que as respostas às perguntas dos usuários

¹²⁹ “[...] the central role of information environments in today’s societies and notes that designing effectively for such environments calls for thinking beneath the surface to the structures that underlie them.”

¹³⁰ Neste contexto, nos referimos ao uso de metadados estruturados e de semântica formal, sobretudo nos atributos de metadados, cujo aprofundamento se faz nas seções 4.4.1.3 e 4.4.1.4 desta tese.

“subam” à superfície. Esta é a arquitetura de informação *bottom-up*; estrutura de conteúdo, sequenciamento e marcação [...].¹³¹

Apreende-se da afirmação que a Arquitetura da Informação *bottom-up* é considerada, pelos autores, conforme conceituada por Hagedorn (2000) na perspectiva das ferramentas utilizadas para que os conteúdos sejam acionáveis pelo usuário via interface do ambiente informacional e/ou sistema de informação.

Destaca-se que, para Morville e Rosenfeld (2006, p. 48, tradução nossa)

A arquitetura da informação *bottom-up* é importante porque os usuários estão cada vez mais propensos a ignorar a arquitetura da informação *top-down* do seu site. Em vez disso, eles estão usando ferramentas de pesquisa em toda a Web, como o Google, clicando em anúncios e clicando em links enquanto leem seu conteúdo por meio de seus agregadores para se aprofundarem em seu site. Uma vez lá, eles vão querer pular para outro conteúdo relevante em seu site sem aprender como usar a estrutura *top-down*. Uma boa arquitetura de informação é projetada para antecipar esse tipo de uso [...].¹³²

Ainda, é notório que o conteúdo é o que leva o usuário a determinada área do ambiente informacional. Assim, “O mapeamento de conteúdo é onde a arquitetura da informação *top-down* encontra a *bottom-up*.”¹³³ (MORVILLE; ROSENFELD, 2006, p. 313, tradução nossa), por isso, a representação, presente na Arquitetura da Informação *bottom-up* é imprescindível, e a estruturação dos dados representacionais ocorre na Arquitetura de Dados.

Nesse sentido, Rosenfeld e Morville (1998, p. 37, tradução nossa) afirmam que, normalmente, a organização da informação hierárquica pode demandar o uso de banco de dados, e que isso requer que, na Arquitetura da Informação *top-down*, o conteúdo e a estrutura dos tipos de registros sejam identificados. No que tange à implantação e estruturação dos bancos de dados, os autores apontam a necessidade de envolver um programador. Enquanto, na representação da informação, apresentam a possibilidade do uso de vocabulário controlado, entendido como “Um vocabulário controlado especifica os termos aceitáveis para uso em um determinado

¹³¹ “[...] if you look closely enough, you can see information architecture even when it’s embedded in the guts of your content. In fact, by supporting searching and browsing, the structure inherent in content enables the answers to users’ questions to “rise” to the surface. This is bottom-up information architecture; content structure, sequencing, and tagging [...]”

¹³² “Bottom-up information architecture is important because users are increasingly likely to bypass your site’s top-down information architecture. Instead, they’re using webwide search tools like Google, clicking through ads, and clicking links while reading your content via their aggregators to find themselves deep in your site. Once there, they’ll want to jump to other relevant content in your site without learning how to use its top-down structure. A good information architecture is designed to anticipate this type of use [...]”

¹³³ “Content mapping is where top-down information architecture meets bottom-up.”

campo. Também pode empregar notas de escopo que definem cada termo.”¹³⁴ (ROSENFELD; MORVILLE, 1998, p. 38, tradução nossa), ou ainda como “Um vocabulário controlado é simplesmente uma lista de termos predeterminados que descrevem um tópico, como arte ou ciência da computação.”¹³⁵ (ROSENFELD; MORVILLE, 1998, p. 73, tradução nossa).

A afirmação dos autores indica o uso de vocabulários controlados para a representação do campo de valor do dado, inclusive, ao considerar que seu uso gera consistência e beneficia a navegação e a busca; bem como para a estruturação do sistema de rotulagem.

No que tange à representação da informação, os sistemas da Arquitetura da Informação são fortemente dependentes de estruturação, o que se faz por meio de metadados e demais elementos que compõem a Arquitetura de Dados.

Considerando que os metadados são compostos por atributos codificados e campo de valor – sendo os atributos relacionados à estrutura dos dados e o campo de valor correspondente à representação da entidade – é relevante considerar que, no domínio da AI, os atributos são “Aspectos da informação sobre um objeto de conteúdo. Os atributos podem ser campos, tags e meta-tags.”¹³⁶ (HAGEDORN, 2000, p. 2, tradução nossa). Enquanto valores de atributos, consistem em “Dados descritivos sobre o objeto de conteúdo. Os valores dos atributos podem ser controlados (de acordo com um vocabulário) ou não controlados. Os valores dos atributos podem ser metainformação.”¹³⁷ (HAGEDORN, 2000, p. 2, tradução nossa).

Esclarece-se, assim, que a representação da informação necessita de uma camada de estruturação, constituída por atributos ou elementos codificados a ela atrelados que estão estruturados na camada de dados. Neste sentido, os metadados, entendidos como atributo codificado e campo de valor, possuem estruturação sintática e semântica próprias, visando o uso por humanos e/ou máquinas, o que permite que sejam não-estruturados ou estruturados.

Morville e Rosenfeld (2006, p. 5, tradução nossa, grifo do autor) elucidam que, no contexto da Arquitetura da Informação,

¹³⁴ “A controlled vocabulary specifies the acceptable terms for use in a particular field. It may also employ scope notes that define each term.”

¹³⁵ “A controlled vocabulary is simply a list of predetermined terms that describe a topic, such as art or computer science.”

¹³⁶ “Aspects of information about a content object. Attributes may be fields, tags and meta-tags.”

¹³⁷ “Descriptive data about the content object. Attribute values can be controlled (conforming to a vocabulary) or uncontrolled. Attribute values may be meta information.”

A estruturação envolve determinar os níveis apropriados de granularidade para os "átomos" de informações em seu site e decidir como relacioná-los entre si. Organizar envolve agrupar esses componentes em categorias significativas e distintivas. Rotular significa descobrir como chamar essas categorias e a série de links de navegação que levam a elas.^{138, 139}

Sant'Ana (2016, p. 132) destaca que os dados coletados e armazenados precisam ser disponibilizados. Assim,

É necessário prever meios que permitam que estes dados estejam acessíveis e ainda incorporar semântica para que sejam passíveis de interpretação, preferencialmente automatizada. É necessário possibilitar que a base de dados contenha elementos que permitam e facilitem, também, a sua localização.

No que tange à camada de dados de um sistema de informação e/ou ambiente informacional digital, é necessário que, de forma análoga ao que se faz na Arquitetura da Informação, haja uma arquitetura – conceituada nesta tese como Arquitetura de Dados – que, considerados os elementos, estabelece o projeto e o fluxo de dados, considerando as formas de localização, coleta/consumo, tratamento, armazenamento, definição das condições de uso, disponibilização/fornecimento, acesso e (re)uso.

O percurso da investigação realizada nesta tese permite afirmar que a Arquitetura de Dados não consiste em uma nova abordagem da Arquitetura da Informação. Assim, busca-se esclarecer como a Arquitetura de Dados se materializa e se relaciona à Arquitetura da Informação, cuja conexão se faz por meio da *AI bottom-up*.

Diante do exposto, e considerando os objetivos específicos (b, c, e d) desta tese – especialmente no que concerne em investigar e compreender a Arquitetura de Dados e a Ecologia Complexa de Dados, estabelecer os elementos conceituais essenciais à Arquitetura de Dados no contexto da Ciência da Informação e investigar as relações entre Arquitetura da Informação e Arquitetura de Dados no contexto da Ciência da Informação –, tem-se a necessidade de buscar na literatura as bases conceituais da Arquitetura de Dados (AD), de forma a compor o segundo capítulo do polo teórico da pesquisa, preponderante para atingir ao objetivo geral desta tese.

¹³⁸ "Structuring involves determining the appropriate levels of granularity* for the information "atoms" in your site, and deciding how to relate them to one another. Organizing involves grouping those components into meaningful and distinctive categories. Labeling means figuring out what to call those categories and the series of navigation links that lead to them."

¹³⁹ Os átomos aos quais os autores se referem são dados.

Considerando as discussões realizadas nesta tese, faz-se necessário investigar como dado e informação são compreendidos na Ciência da Informação, visando contribuir para a discussão dos dados no contexto da Arquitetura de Dados.

Neste sentido, de acordo com DAMA-DMBOK¹⁴⁰ (2017, p. 20, tradução nossa) “Reconhecer que dados e informações precisam ser preparados para diferentes propósitos conduz a um princípio central de gerenciamento de dados: tanto os dados quanto as informações precisam ser gerenciados.”¹⁴¹

Além disso, faz-se necessário compreender como a Arquitetura de Dados tem sido conceituada em diferentes áreas do conhecimento, quais os elementos que a compõem e os entregáveis que possui, visando contribuir para alicerçar o conceito de Arquitetura de Dados (AD) no contexto da Ciência da Informação, objetivo geral desta tese.

4.1 DADO E INFORMAÇÃO: TESSITURAS TEÓRICAS

As discussões realizadas nas seções anteriores desta tese buscaram na literatura científica da Arquitetura da Informação aspectos relacionados à informação e à camada de dados dos ambientes informacionais digitais ou sistemas de informação. E, muito embora a camada de dados – entendida por Morville e Rosenfeld (2006) como invisível – seja perceptível, ela não é abordada de forma consistente e com o rigor necessário na AI *bottom-up* sob a ótica da representação, estruturação, tratamento, modelo e fluxo de dados, restringindo-se à menção de metadados, vocabulários controlados e tesouros, que são denominados nesta tese como elementos do sistema de representação da informação, imprescindíveis ao processo de representação da informação.

Nesse contexto, há três termos que são recorrentes na discussão: arquitetura, dado e informação. Assim, considerando a Arquitetura da Informação já consolidada na Ciência da Informação e o objetivo desta tese, que consiste em investigar e

¹⁴⁰ O *Data Management Body of Knowledge (DAMA-DMBOK, 2017)* é uma publicação da *Data Management Association*, também conhecida como *DAMA Internacional*, uma associação global sem fins lucrativos, composta por diferentes agentes que se dedicam ao avanço dos conceitos e práticas de gestão de dados e informações. “O objetivo principal da *DAMA Internacional* é promover a compreensão, o desenvolvimento e a prática do gerenciamento de dados e informações como ativos corporativos essenciais para apoiar a organização.” (*DAMA Internacional, 2022, tradução nossa*).

¹⁴¹ “Recognizing data and information need to be prepared for different purposes drives home a central tenet of data management: Both data and information need to be managed.”

compreender a Arquitetura de Dados a fim de conceituá-la no contexto da Ciência da Informação e relacioná-la com a Arquitetura da Informação, é necessário buscar aspectos que possibilitem compreender os conceitos de dado e informação, cuja relação tão estreita por vezes leva ao uso dos termos como sinônimos.

Neste ponto, considera-se que a discussão a fim de estabelecer os limites entre os conceitos de dado e informação é complexa. Porém, torna-se necessário buscar elementos que esclareçam como tais conceitos são adotados neste estudo – sobretudo para conceituar Arquitetura de Dados no contexto da Ciência da Informação – esclarecendo que não há pretensão de esgotar a temática.

Iyamu (2019, p. 1, tradução nossa) corrobora desse entendimento ao afirmar que:

Embora esteja bem documentado no domínio acadêmico, incluindo vários dicionários em inglês, que dados e informações não são a mesma coisa, sendo o último um refinamento do primeiro, muitas pessoas dentro de organizações e ambientes acadêmicos continuam a usar esses termos de forma vaga e intercambiável.¹⁴²

O autor continua, afirmando que esse é um desafio que impacta as organizações, especialidades e atuação profissional, bem como a obtenção de dados e informações, envolvendo requisitos, desenvolvimentos e implementações de sistemas e tecnologias de informação. A discussão de Iyamu (2019) é importante neste estudo, pois aponta os impactos da falta de clareza no uso dos termos dado e informação no contexto da Arquitetura da Informação e de Dados, empregada nos sistemas de informação, de forma que entendemos que há convergência com o que se busca discutir nesta tese.

Para Sayão e Sales (2020, p. 32) “A relação entre dado e informação desencadeia uma infinidade de debates que conectam aspectos tecno-sociais com abordagens epistemológicas, que enriquecem os estudos e debates em Ciência da Informação.”

Considerando o posicionamento desta tese na Ciência da Informação, iniciamos a discussão acerca de dado e informação a partir de dicionários específicos do domínio.

¹⁴² “Even though it is well documented in the academic domain, including various English dictionaries, that data and information are not the same thing, with the latter being a refinement of the former, many people within organisations and academic environments continue to loosely and interchangeably use these terms.”

De acordo com o Dicionário brasileiro de terminologia arquivística do Arquivo Nacional (Brasil) (2005, p. 61) dado é a “Representação de todo e qualquer elemento de conteúdo cognitivo, passível de ser comunicada, processada e interpretada de forma manual ou automática.” (p. 62), ao passo que informação é o “Elemento referencial, noção, idéia ou mensagem contidos num documento.” (p. 107).

No Dictionary of Information Science and Technology (KHOSROW-POUR, 2007, p. 152, tradução nossa, grifo nosso) a entrada para dados é:

Dados 1: Uma coleção de atributos (numéricos, alfanuméricos, figuras, imagens) sobre entidades (coisas, eventos, atividades). Os dados espaciais representam recursos tangíveis (entidades). Além disso, os dados espaciais são geralmente um atributo (descriptor) do recurso espacial. (Al-Hanbali & Sadoun, 2006) 2: Um conjunto de fatos discretos e objetivos sobre eventos. (Joia, 2005) 3: Representações binárias (digitais) de fatos atômicos, principalmente de transações financeiras. Os dados também podem ser textos, gráficos, imagens mapeadas por bits, som ou segmentos de vídeo ao vivo analógicos ou digitais. Dados estruturados são a matéria-prima para análise usando um DSS orientado a dados. Os dados são fornecidos por produtores de dados e são usados por consumidores de **informações** para criar **informações**. (Power, 2005) 4: Combinação de fatos e significados que são processados em **informação**. (Yoon *et al.*, 2005) 5: Os dados são portadores de conhecimento e **informação**. Eles consistem principalmente em sinais e são a matéria-prima a ser processada. Os dados representam observações ou fatos fora de contexto que não são diretamente significativos. Tanto a **informação** quanto o conhecimento são comunicados por meio de dados. (Haghirian, 2006) 6: Conhecimento altamente explícito derivado dos dados em bancos de dados e data warehouses usados para tomada de decisões estratégicas após resumir, analisar, minerar e assim por diante. (Kulkarni & Freeze, 2006) 7: Algo dado ou admitido como fato no qual uma inferência pode ser baseada. Observações simples do mundo, muitas vezes quantificadas e facilmente estruturadas, capturadas em máquinas e transferidas. O número de “baby boomers” nascidos em um determinado ano são dados. (Mockler & Dologite, 2005) 8: O termo genérico para sinais, símbolos e imagens. Os dados podem ser salvos, processados, impressos e assim por diante. Eles não estão vinculados a indivíduos. (Hofer, 2006) 9: A matéria-prima que alimenta o processo de geração de **informação**. (Hoxmeier, 2005) 10: O conjunto de amostras, fatos ou casos em um repositório de dados. Como exemplo de amostra, considere os valores de campo de uma aplicação de crédito específica em um banco de dados bancário. (de Carvalho *et al.*, 2005) 11: Frequentemente definido como os fatos brutos e dados de **informação**. O conhecimento é sobre a aplicação de dados e **informações** para uma determinada tarefa, de modo que essa tarefa possa ser efetivamente executada. (R. Zhang, 2005).¹⁴³

¹⁴³ “Data 1: A collection of attributes (numeric, alphanumeric, figures, pictures) about entities (things, events, activities). Spatial data represent tangible features (entities). Moreover, spatial data are usually an attribute (descriptor) of the spatial feature. (Al-Hanbali & Sadoun, 2006) 2: A set of discrete and objective facts concerning events. (Joia, 2005) 3: Binary (digital) representations of atomic facts, especially from financial transactions. Data may also be text, graphics, bit-mapped images, sound, or analog or digital live-video segments. Structured data

Ainda, no Dictionary of Information Science and Technology (KHOSROW-POUR, 2007, p. 152, tradução nossa, grifo nosso) a entrada para informação consiste em:

Informação 1: Um processo com **dados** como entrada e conhecimento como saída. Um indivíduo é o sujeito que transforma os **dados** em conhecimento. As relações entre dois dispositivos técnicos são processos de troca de **dados**. As relações entre dois ou mais indivíduos são processos de comunicação. O processo de informação inversa é chamado de processo de documentação (por exemplo, escrever um artigo). (Hofer, 2006) 2: **Dados** com contexto e utilidade. (Hoxmeier, 2005) 3: Unidade comparativa de cognição que define uma mudança entre o estado anterior e o presente dos sistemas naturais, artificiais ou semióticos. (Targowski, 2005) 4: Uma mensagem, ou um **dado**, que faz a diferença. A informação tem significado e se torna conhecimento quando a pessoa a internaliza. (Huotari & livonen, 2005) 5: Comumente conhecido como uma coleção de fatos ou **dados**. Em Ciência da Computação, refere-se a **dados** processados, armazenados ou transmitidos. Na Gestão do Conhecimento, refere-se ao conhecimento codificado. (Theng, 2005) 6: **Dados** contextualizados que podem ser analisados e aplicados às circunstâncias de tomada de decisão. (Taylor, 2005) 7: **Dados** que estão associados a algum sistema que permite que o significado seja derivado por alguma entidade. (Benyon, 2006) 8: **Dados** com atributos de relevância e propósito. (Joia, 2005) 9: No uso de inteligência, **dados** não processados de toda descrição que podem ser usados na produção de inteligência. (Melkonyan, 2005) 10: Símbolos interpretados e estruturas de símbolos que reduzem tanto a incerteza quanto o equívoco ao longo de um período de tempo definido. (Hirji, 2005) 11: Conhecimento adquirido por experiência ou estudo. (Drake, 2006) 12: Conhecimento derivado da leitura, observação ou instrução, às vezes consistindo em fatos ou **dados** não organizados ou não relacionados. **Dados** dotados de relevância e finalidade, como, por exemplo, o balanço patrimonial e a demonstração de resultados de uma empresa. (Mockler & Dologite, 2005) 13: Um termo que se refere a detalhes sobre um evento ou situação no passado ou simplesmente um fato científico. A informação pode ser considerada como uma peça de conhecimento de um tipo objetivo. Ela resulta da colocação de **dados** em algum contexto significativo, geralmente na forma de uma mensagem. É puramente descritivo e explícito, não possibilita

are the raw material for analysis using a data-driven DSS. The data are supplied by data producers and are used by information consumers to create information. (Power, 2005) 4: Combination of facts and meanings that are processed into information. (Yoon et al., 2005) 5: Data are carriers of knowledge and information. They consist mostly of signs and are the raw material to be further processed. Data represent observations or facts out of context that are not directly meaningful. Both information and knowledge are communicated through data. (Haghirian, 2006) 6: Highly explicit knowledge derived from the data in databases and data warehouses used for strategic decision making after summarizing, analyzing, mining, and so forth. (Kulkarni & Freeze, 2006) 7: Something given or admitted as a fact on which an inference may be based. Simple observations of the world, which are often quantified, and easily structured, captured on machines, and transferred. The number of "baby boomers" born in a given year is data. (Mockler & Dologite, 2005) 8: The generic term for signs, symbols, and pictures. Data can be saved, processed, printed, and so on. They are not bound to individuals. (Hofer, 2006) 9: The raw material that feeds the process of information generation. (Hoxmeier, 2005) 10: The set of samples, facts, or cases in a data repository. As an example of a sample, consider the yield values of a particular credit application in a bank database. (de Carvalho et al., 2005) 11: Often defined as the raw facts and information data. Knowledge is about the application of data and information for a given task so that the given task can be effectively performed. (R. Zhang, 2005)"

decisões ou ações, nem desencadeia novas questões. (Haghirian, 2006).¹⁴⁴ (KHOSROW-POUR, 2007, p. 327-328, tradução nossa, grifo nosso).

O Dicionário de Biblioteconomia e Arquivologia, de Cunha e Cavalcanti (2008) apresenta os termos de forma mais detalhada, utilizando, inclusive, qualificadores que possibilitam ampliar a compreensão dos conceitos a partir das suas adjetivações. No caso de dados, os autores apresentam uma definição geral no contexto da Estatística e Informática e uma no contexto da Biblioteconomia, Comunicação e Informática, seguidas de entradas específicas, relativas a adjetivações de dados. Para os autores, no contexto da Estatística e Informática, dados são

[...] a menor representação convencional e fundamental de uma **informação** (fato, noção, objeto, nome próprio, número, estatística, etc.) sob forma analógica ou digital passível de ser submetida a processamento manual ou automático (roufon). (CUNHA; CAVALCANTI, 2008, p. 112-113, grifo nosso).

Ainda, no contexto da Estatística e Informática, os autores supracitados adjetivam o termo em dado estruturado e factual. Enquanto no contexto da Biblioteconomia, Comunicação e Informática, dados:

Em sentido mais amplo, abrange toda **informação** quantificável (números, letras, gráficos, imagens, sons ou uma combinação desses tipos). 1.1 Sinais ou códigos usados para alimentação, armazenamento, processamento e produção de um resultado. 1.2 Representação de um acontecimento ou conceito, sob uma forma susceptível de comunicação, de interpretação ou de tratamento, quer manualmente, quer por meios automáticos. 2. Grupo de caracteres alfabéticos, numéricos, alfanuméricos ou quaisquer outros, que representam uma condição ou valor específico. Os dados são, na realidade, os blocos construtivos da **informação**. 2.1 Uma referência não-elaborada, algo não-interpretado, não-classificado, não-

¹⁴⁴ "Information 1: A process with data as input and knowledge as output. An individual is the subject who transforms the data into knowledge. Relations between two technical devices are data exchange processes. Relations between two or more individuals are communication processes. The reverse information process is called documentation process (e.g., writing an article). (Hofer, 2006) 2: Data with context and utility. (Hoxmeier, 2005) 3: A comparative unit of cognition that deines a change between the previous and present state of the natural, artiicial, or semiotic systems. (Targowski, 2005) 4: A message, or data, which makes a difference. Information has meaning, and becomes knowledge when a person internalizes it. (Huotari & livonen, 2005) 5: Commonly known as a collection of facts or data. In Computer Science, it refers to processed, stored, or transmitted data. In Knowledge Management, it refers to codiied knowledge. (Theng, 2005) 6: Contextualized data that can be analyzed and applied to decision-making circumstances. (Taylor, 2005) 7: Data that is associated with some system that enables meaning to be derived by some entity. (Benyon, 2006) 8: Data with attributes of relevance and purpose. (Joa, 2005) 9: In intelligence usage, unprocessed data of every description which may be used in the production of intelligence. (Melkonyan, 2005) 10: Interpreted symbols and symbol structures that reduce both uncertainty and equivocality over a deined period of time. (Hirji, 2005) 11: Knowledge acquired through experience or study. (Drake, 2006) 12: Knowledge derived from reading, observation, or instruction, at times consisting of unorganized or unrelated facts or data. Data endowed with relevance and purpose, for example, a irm's balance sheet and income statement. (Mockler & Dologite, 2005) 13: A term referring to details about an event or situation in the past or simply a scientific fact. Information can be regarded as a piece of knowledge of an objective kind. It results from placing data within some meaningful context, often in the form of a message. It is purely descriptive and explicit, does not enable decisions or actions, nor does it trigger new questions. (Haghirian, 2006)"

estruturado, não-ajustado a um contexto (baseado em matto). 2.2 Informação em forma codificada. 3. "Fatos, noções ou instruções representados de uma forma conveniente para um processo de comunicação, uma interpretação ou um processamento quer humano, quer através de meios automáticos. Os dados analógicos são representados por funções contínuas, enquanto que os dados digitais são representados por funções discretas" (afnor). (CUNHA; CAVALCANTI, 2008, p. 113, grifo nosso).

Neste contexto, em Cunha e Cavalcanti (2008), as adjetivações se referem a dados alfabéticos, bibliográficos, catalográficos, brutos, de áudio, de censo, de entrada, discretos, espaciais, estatísticos, geoespaciais, factuais, gráficos, legíveis por computador, não-processados, não-recuperados, não-trabalhados, numéricos, originais, pessoais, processáveis por máquina, temáticos, temporais, textuais.

Ao tratar de informação, Cunha e Cavalcanti (2008), de igual maneira, apresentam uma definição geral, seguida de adjetivações específicas. Para os autores, informação é definida como:

1. BIB 1.1 Registro de um conhecimento que pode ser necessário a uma decisão. A expressão 'registro' inclui não só os documentos tipográficos, mas também os reprográficos, e quaisquer outros suscetíveis de serem armazenados visando sua utilização. 1.2 Informação, na sua definição mais ampla, é uma prova que sustenta ou apóia um fato. 1.3 Registro de um conhecimento para utilização posterior. 1.4 **Dados** numéricos alfabéticos ou alfanuméricos processados por computador. 2. BIB COMN INF com a informação podem-se realizar diversas operações, tais como: criação, transmissão, armazenamento, recuperação, recepção, cópia (em diferentes formas), processamento e destruição. A transmissão da informação é feita numa grande variedade de formas, entre as quais se incluem: luz, som, ondas de rádio, corrente elétrica, campos magnéticos e marcas sobre o papel. 3. COMN coleção de símbolos que possuem significados. 4. ENG INF TEL uma informação pode ser descrita em termos de sua manifestação física: o meio que a transporta, a exatidão, a quantidade que é transmitida ou recebida. 5. LING a informação pode ser descrita em termos do seu objeto de referência, seu significado e estrutura. i. (assimilação) => assimilação da informação, i. (entropia) entropia, i. (medida) => medida da informação, i. (necessidade) => necessidade de informação, i. (processamento) processamento da informação, processamento de dados, i. (processo) => processo de informação, i. (suporte físico/material) => suporte de informação, i. (teoria) => teoria da informação, i. (tratamento) => processamento da informação, i. (vício) => viciado em informação. (CUNHA; CAVALCANTI, 2008, p. 201-202, grifo nosso).

E as adjetivações para informação são: adicional, administrativa, atualizada, bibliográfica, biológica, biomédica, bruta, bursátil, catalográfica, científica, i. científica e tecnológica, com valor agregado, comercial, complementar, comunitária,

confidencial, corporativa, corrente, de acompanhamento, de agência de notícias, de atualização, de empresas, de texto completo, desatualizada, digital, direta, documental, e sociedade, econômica, eletrônica, em arte, em ciência e tecnologia, em linha, em química, em saúde, em tempo real, errônea, especializada, esportiva, estatística, estratégica, factual, financeira, geográfica, gerencial, governamental, gráfica, imobiliária, impertinente, industrial, industrial e tecnológica, inválida, irrelevante, jurídica, legal, legível mecanicamente, legível por máquina, local, mecanizável, médica, não-estratégica, não-linear, não-orgânica, numérica, nutricional, oficial, oral, orgânica, organizacional, para a indústria, para a saúde, para as minorias, para negócios, parasita, pictórica, pré-empacotada, primária, pública, quântica, química, rápida, reempacotada, recuperada, registrada, relevante, secreta, secundária, sobre bolsas de valores, sobre economia, sobre empresas, sobre esportes, sobre fatos, sobre imóveis, sobre localização e acervo, sobre nutrição, sobre patentes, sobre preços, sobre viagens, suplementar, técnica, tecnológica, terciária, terminológica, textual, transmutada, utilitária (CUNHA; CAVALCANTI, 2008).

Vale destacar que, no conjunto de definições do termo dado apresentado em em Cunha e Cavalcanti (2008), há 14 ocorrências da palavra informação/ões, ao passo que, nas definições de informação, há 30 menções ao termo dado/s. O mesmo ocorre em Khosrow-Pour (2007): na definição do termo dado, há 8 ocorrências da palavra informação/ões, ao passo que, na definição de informação, há 14 menções ao termo dado/s.

Evidencia-se assim, nas definições de dados apresentadas por Khosrow-Pour (2007) e Cunha e Cavalcanti (2008), a presença do termo informação, assim como na definição de informação o termo dados está presente, o que comprova o estreito relacionamento entre os termos dado e informação, fator que reforça a linha tênue que os relaciona, limita ou separa.

No Glossário de Termos Técnicos em Ciência da Informação (NORTE, 2010) não há entrada para dado, mas há para informação:

[...] reunião de dados em uma forma compreensível capaz de comunicação. O Conteúdo pode aparecer em qualquer formato - escrito ou impresso no papel, fornecido em bancos de dados eletrônicos, reunidos na Internet, etc. - pode aparecer até no conhecimento pessoal de uma organização. Como os termos demonstrados abaixo (especialmente engenharia de Informação, gerência de Informação, ciência de Informação), a informação é um termo que cobre muitas atividades inter-relacionadas que usam as habilidades da biblioteconomia. A Gerência de Conhecimento e

Gerência de Conteúdos são recentes manifestações da extensão do valor e o poder da informação. (PRYTHERCH, 1995). (NORTE, 2010, p. 24).

A norma ISO/IEC 2382, *Information technology — Vocabulary*, define dado como “representação reinterpretabil de informações de uma maneira formalizada adequada para comunicação, interpretação ou processamento”¹⁴⁵ e inclui uma nota de entrada “Os dados podem ser processados por humanos ou por meios automáticos.”¹⁴⁶ (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015c, tradução nossa).

Enquanto informação possui duas entradas diferentes, sendo: “<processamento de informações> conhecimento sobre objetos, como fatos, eventos, coisas, processos ou ideias, incluindo conceitos, que dentro de um determinado contexto tem um significado particular”¹⁴⁷, e, “<teoria da informação> conhecimento que reduz ou remove a incerteza sobre a ocorrência de um evento específico de um determinado conjunto de eventos possíveis”¹⁴⁸. (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015c, tradução nossa).

De acordo com o ODLIS: Online Dictionary for Library and Information Science (REITZ, 2013, tradução nossa), dados:

O plural da palavra latina *datum*, que significa "o que é dado", muitas vezes usado como um substantivo coletivo singular. Fatos, figuras ou instruções apresentadas em uma forma que pode ser compreendida, interpretada e comunicada por um ser humano ou processada por um computador. Compare com informação e conhecimento. Veja também: banco de dados, base de dados, conjunto de dados e metadados.¹⁴⁹

Enquanto informação:

Dados apresentados de forma facilmente compreensível aos quais foi atribuído significado dentro do contexto de seu uso. Num sentido mais dinâmico, a mensagem transmitida pelo uso de um meio de comunicação ou expressão. Se uma mensagem específica é informativa ou não, depende em parte da percepção subjetiva da pessoa que a recebe.

¹⁴⁵ “reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing”

¹⁴⁶ “Data can be processed by humans or by automatic means.”

¹⁴⁷ “<information processing> knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning”

¹⁴⁸ “<information theory> knowledge which reduces or removes uncertainty about the occurrence of a specific event from a given set of possible events”

¹⁴⁹ “The plural of the Latin word *datum*, meaning ‘what is given,’ often used as a singular collective noun. Facts, figures, or instructions presented in a form that can be comprehended, interpreted, and communicated by a human being or processed by a computer. Compare with information and knowledge. **See also:** data bank, database, data set, and metadata.”

Mais concretamente, todos os fatos, conclusões, ideias e trabalhos criativos do intelecto e da imaginação humanos que foram comunicados, formal ou informalmente, de qualquer forma. Em seu discurso de posse de 1801, Thomas Jefferson listou a "difusão de informações" como um dos princípios fundamentais da forma republicana de governo estabelecida pela Constituição dos Estados Unidos. Compare com o conhecimento. Veja também: desinformação e misinformação.¹⁵⁰ (REITZ, 2013, tradução nossa).

De igual maneira, considerando a presença dos termos dado e informação na Ciência da Computação, buscaram-se dicionários do domínio em busca de definições dos termos. De acordo com o Dictionary of Computer and Internet Terms (DOWNING; COVINGTON; COVINGTON, 2009, p. 124, tradução nossa) "data information A palavra era originalmente o plural de *datum*, que significa 'um único fato', mas agora é usado como um singular coletivo."¹⁵¹

Realizaram-se, ainda, buscas exploratórias em bases de dados, a fim de identificar as discussões acerca dos conceitos de dado e informação, para que possam trazer elementos para a análise necessária a esta tese.

Retoma-se, assim, a pirâmide do conhecimento, do inglês *Data-Information Knowledge-Wisdom* (DIKW), cuja autoria é atrelada a Ackoff (1989), que consiste em uma representação hierárquica, em forma de pirâmide, na qual os elementos – dado, informação, conhecimento e sabedoria – são empilhados, e cada camada acrescenta atributos aos presentes na camada inferior.

Para Ackoff (1989), há uma hierarquia de tipos de conteúdos da mente humana, em cujo topo está a sabedoria e, abaixo dela, respectivamente, a compreensão, o conhecimento, a informação e, na parte inferior, os dados. Embora o autor mencione conteúdos da mente humana, os processos são descritos na perspectiva dos sistemas de informação. Para o autor,

Dados são símbolos que representam propriedades de objetos, eventos e seus ambientes. São produtos da observação. Observar é sentenciar. A tecnologia de sensoriamento, instrumentação, é, obviamente, altamente desenvolvida. A informação, como observado,

¹⁵⁰ "Data presented in readily comprehensible form to which meaning has been attributed within the context of its use. In a more dynamic sense, the message conveyed by the use of a medium of communication or expression. Whether a specific message is informative or not depends in part on the subjective perception of the person receiving it.

More concretely, all the facts, conclusions, ideas, and creative works of the human intellect and imagination that have been communicated, formally or informally, in any form. In his inaugural address of 1801, Thomas Jefferson listed the "diffusion of information" as one of the fundamental principles of the republican form of government established under the Constitution of the United States. Compare with knowledge. See also: disinformation and misinformation."

¹⁵¹ "data information. The word was originally the plural of *datum*, which means "a single fact," but it is now used as a collective singular."

é extraída dos dados por análise em muitos aspectos dos quais os computadores são adeptos.

Dados, como minérios metálicos, não têm valor até que sejam processados em uma forma utilizável (ou seja, relevante). Portanto, a diferença entre dados e informações é funcional, não estrutural, mas os dados costumam ser reduzidos quando são transformados em informação.¹⁵² (ACKOFF, 1989, p. 3, tradução nossa).

Para o autor, os sistemas de informação armazenam, processam e recuperam dados dos quais são inferidas informações. A discussão de Ackoff (1989) realizada antes dos avanços das tecnologias computacionais, sobretudo da internet, já apontava os dados como registros factuais, cuja estruturação, análise e processamento os transformam em informação.

Cumprido destacar que o estudo de Ackoff (1989) – amplamente utilizado, sobretudo nas pesquisas de Gestão do Conhecimento – embora atrelado ao surgimento da pirâmide do conhecimento, não apresenta uma representação imagética do que seria a hierarquia por ele proposta. O que o autor faz é estabelecer que na base da hierarquia estão os dados e, sobre eles, respectivamente, informação, conhecimento, compreensão e sabedoria, sendo que o elemento superior da hierarquia inclui as categorias que estão abaixo dele.

Rowley (2007) examinou as definições dos termos que compõem a DIKW em livros-texto das disciplinas de sistemas de informação e gestão do conhecimento. Este estudo permitiu identificar que a literatura consultada é unânime na presença dos elementos dado, informação, conhecimento e sabedoria, presentes na DIKW; contudo, há a presença de iluminação/esclarecimento e, em algumas obras o elemento compreensão está contido na transformação de um elemento para outro e não como um elemento, conforme apresentado por Ackoff (1989).

A partir das obras analisadas, Rowley (2007, p. 171, tradução nossa) afirma que “[...] os dados carecem de significado ou valor, são desorganizados e não processados. Eles estabelecem as bases para definir a informação em termos de

¹⁵² “Data are symbols that represent properties of objects, events and their environments. They are products of observation. To observe is to sense. The technology of sensing, instrumentation, is, of course, highly developed. Information, as noted, is extracted from data by analysis in many aspects of which computers are adept. Data, like metallic ores, are of no value until they are processed into a useable (i.e. relevant) form. Therefore, the difference between data and information is functional, not structural, but data are usually reduced when they are transformed into information.”

dados.”¹⁵³, o que reforça a dificuldade de definir dados e de estabelecer distanciamento da definição de informação.

A autora menciona ainda a forte relação entre dados e informação, tanto na literatura científica da área de sistemas de informação quanto na gestão do conhecimento, para as quais a “[...] a informação é definida em termos de dados e é vista como dados organizados ou estruturados. Esse processamento confere relevância aos dados para uma finalidade ou contexto específico e, portanto, os torna significativos, valiosos, úteis e relevantes.”¹⁵⁴ (ROWLEY, 2007, p. 172, tradução nossa).

A discussão proposta por Rowley (2007) apresenta, com base na literatura estudada, que há consenso de que dado, informação e conhecimento são definidos com alguma relação, o que parece corroborar com a hierarquia de Ackoff (1989). Por outro lado, aponta que não há consenso quanto ao processo de transformação de dado em informação e de informação em conhecimento, o que parece culminar na falta de clareza sobre a existência de três conceitos distintos, embora haja indicativos de que a informação é vista como dados organizados e estruturados. Assim, a autora reflete que todo o dado, estando armazenado na mente humana ou em sistemas de informação, possui alguma estrutura, de forma que, se é a estrutura que diferencia dado de informação, o que é armazenado na mente humana é informação e não dado.

Essa reflexão feita por Rowley (2007) mais uma vez reforça que há uma linha tênue que separa dado de informação. Contudo, vale ponderar que, embora haja uma estrutura no momento da coleta, criação e/ou armazenamento de dados, essa estrutura também necessita de padrões semânticos para que o dado tenha significado e seja, de fato, informação para qualquer consumidor que tenha acesso a ele, seja esse consumidor humano ou aplicação/agente computacional, isso parece dar aos dados contexto e significado, visando favorecer a compreensão dos dados por máquinas.

No estudo de Ribeiro e Santos (2020), a representação hierárquica de Ackoff (1989) em pirâmide é questionada e os autores apresentam uma nova representação para a estrutura DIKW. Para além disso, no referido estudo, há uma discussão sobre

¹⁵³ “[...] data lacks meaning or value, is unorganized and unprocessed. They lay the foundations for defining information in terms of data.”

¹⁵⁴ “[...] information is defined in terms of data, and is seen to be organized or structured data. This processing lends the data relevance for a specific purpose or context, and thereby makes it meaningful, valuable, useful and relevant.”

os elementos que compõem a DIKW, sendo importante para esta tese a discussão acerca de dado e informação. Neste sentido, Ribeiro e Santos (2020), a partir de uma revisão de literatura, sistematizam os conceitos e as características de dado e informação em quadros. No que tange aos dados, a sistematização é apresentada no Quadro 4.

Quadro 4 – Conceitos e características dos dados

Conceitos	Fatos	Objetivos e discretos	Aukett (2019); Jennex e Bartczak (2013)
		Brutos	Alavi e Leidner (2001); Bierly, Kessler e Christensen (2000); Sajja e Akerkar (2010); Yao, Jin e Zhang (2014)
		Observados ou registrados	Allen (2016)
	Representações	De objetos	Ponchirolli e Fialho (2005); Sato e Huang (2015)
		Símbolos	Ackoff (1989); Aukett (2019); Sato e Huang (2015); Vandergriff (2008); Wan e Alagar (2014)
		Sinais	Baškarada e Koronios (2013); Moresi (2000)
Produtos da observação		Ackoff (1989); Gandhi (2004); Nurulin et al. (2019); Sajja e Akerkar (2010)	
Características	Não são úteis		Ackoff (1989); Car et al. (2019); Sajja e Akerkar (2010)
	Não têm significado		Ackoff (1989); Allen (2016); Baškarada e Koronios (2013); Wan e Alagar (2014)
	Não têm valor		Allen (2016); Aukett (2019); Sato e Huang (2015)
	Possuem valor (big data)		Fiorini et al. (2018); Gupta et al. (2018); Storey e Song (2017); Tian (2017); Wamba et al., (2017); Ylikoki e Porras (2019)
	Não têm interpretação		Sato e Huang (2015)
	Não têm contexto		Gandhi (2004)
	Estão fora do agente		Baškarada e Koronios (2013)
	Quantificáveis		Allen (2016); Ponchirolli e Fialho (2005)
Tipos	Não estruturados		Dammann (2019); Guolinag et al. (2008); Martins, Simões e Sá (2014); Vandergriff (2008)
	Semiestruturados		Guolinag et al. (2008); Martins, Simões e Sá (2014)
	Estruturados		Guolinag et al. (2008); Martins, Simões e Sá (2014); Vandergriff (2008)

Fonte: Ribeiro e Santos (2020, p. 71).

A partir dos conceitos e características dos dados, Ribeiro e Santos (2020) defendem que os dados possuem quatro tempos (t):

[...] t-1: disponível no ambiente, mas ainda não foi notado; t0: não estruturado, bruto, cru (já foi notado, mas ainda não recebeu tratamento); t1: semiestruturado (começou a receber tratamento ou recebeu um tratamento irregular, incompleto); e t2: estruturado (tratado). Dados semiestruturados (t1) ou mesmo estruturados (t2) não são informação, pois apenas foram organizados de alguma forma, ficando prontos ou quase prontos para receberem um significado e um contexto, que é quando a sua utilidade se concretiza. Só quando isso ocorre é que passam para a categoria da informação [...].

Os autores afirmam ainda que os dados estão fora do agente e possuem valor, embora menor que o valor da informação. Apreende-se com isso que os dados necessitam de interferência para que sejam localizados e tratados a fim de gerar valor.

No que tange à informação, considerada como coisa por Ribeiro e Santos (2020) a partir de Buckland (1991), a sistematização é apresentada no Quadro 5.

Quadro 5 – Conceitos e características das informações

Conceitos	Dados	Significados	Ackoff (1989); Aukett (2019); Baškarada e Koronios (2013); Bierly, Kessler e Christensen (2000); Hoppe et al. (2011); Sato e Huang (2015); Yao, Jin e Zhang (2014)
		Úteis	Dammann e Smart (2019); Yao, Jin e Zhang (2014)
		Processados	Moresi (2000); Ponchirolli e Fialho (2005); Sajja e Akerkar (2010)
		Contextualizados	Alavi e Leidner (2001); Dammann (2019); Dammann e Smart (2019); Gandhi (2004); Jennex e Bartczak (2013); Nurulin et al. (2019); Ponchirolli e Fialho (2005)
		Interpretados	Alavi e Leidner (2001); Takeuchi e Nonaka (2008)
	Mensagem / fluxo de mensagens	Davenport e Prusak (1998); Takeuchi e Nonaka (2008)	
Características	Estão fora do agente	Nurulin et al. (2019)	
	Possuem significado	Zhang e Benjamin (2007)	
	Possuem valor	Aukett (2019); Sato e Huang (2015)	
	Quantificáveis	Allen (2016)	

Fonte: Ribeiro e Santos (2020, p. 73).

A partir desta sistematização Ribeiro e Santos (2020, p. 72) definem informação como:

[...] dados que receberam significação, utilidade, processamento, contexto ou interpretação. Pode também ser entendida como mensagem ou fluxo de mensagens, os quais igualmente remetem a tipo de dado e ao atributo da informação apontado por Zhang e Benjamin (2007) de possuir significados, propósitos e habilidades para transformar. Além disso, assim como os dados, está fora do agente, é quantificável e possui valor.

Os autores defendem ainda que a informação pode ser convertida em dados, a partir da captura e armazenamento em qualquer suporte. Após sistematizarem os demais conceitos relativos à DIKW, Ribeiro e Santos (2020, p. 79) apresentam um quadro comparativo entre as características dos elementos DIKW, dos quais destacam-se no Quadro 6, apenas dado e informação.

Para Ribeiro e Santos (2020, p. 81) “dado e informação não são propriedades do agente”¹⁵⁵ e

[...] o processo cognitivo se inicia a partir do dado semiestruturado, o qual já começou a receber do agente algum tipo de estruturação, baseada em algum critério específico. A informação é tida como o dado que recebeu alguma significação e contextualização pelo agente.

Quadro 6 – Quadro comparativo entre as características de dado e informação

	Dado	Informação
Quantificável	+	+
Possui valor	+	++
Possui significado		+
Possui interpretação		+
Possui contexto		+
Útil		+

Fonte: Adaptado de Ribeiro e Santos (2020, p. 79).

A partir do estudo de Ribeiro e Santos (2020), os dados passam a ter valor em t1 – dado semiestruturado (começou a receber tratamento ou recebeu um tratamento irregular, incompleto) –, quando os dados estão disponíveis no ambiente, são percebidos e começam a receber tratamento/estruturação. Isso leva a crer que os dados necessitam de estrutura mínima para que possam ter valor, e quando adequadamente estruturados e acrescidos de significado e contexto é que transitam para a categoria de informação. Pode-se considerar, então, que os dados estruturados, com significado e contexto, são informação.

Nesse aspecto, os Quadros 4 e 5 auxiliam na compreensão da visão dos autores materializada no Quadro 6. No Quadro 4 os autores estudados são divididos nas características de dado, apontando: não são úteis, não têm significado, não têm valor, possuem valor (big data), não tem interpretação, não tem contexto, estão fora do agente, quantificáveis; ao passo que no Quadro 5, os autores destacam como características da informação: estão fora do agente, possuem significado, possuem valor e são quantificáveis. Com isso, Ribeiro e Santos (2020) expressam no Quadro 6 que, considerando as obras analisadas, dados e informação são considerados quantificáveis, ao passo que, no que tange a possuir valor, significado, interpretação, contexto e utilidade, a informação apresenta graus superiores aos dados.

¹⁵⁵ Entendidos pelos autores como humanos e máquinas.

Date (2004) destaca que a distinção entre os termos dado e informação é importante, embora os trate como sinônimos em diversas partes da obra, fazendo distinções entre os termos quando entende necessário. Porém, destaca que, na literatura, a distinção é feita usando “[...] dados para se referir ao que realmente é armazenado no banco de dados, e informações para se referir ao significado desses dados para determinado usuário. A distinção é claramente importante [...]” (DATE, 2004, p. 5).

Para Amaral (2016), “Dados são fatos coletados e normalmente armazenados. Informação é o dado analisado e com algum significado.”

Sezter (1999) afirma:

Definimos dado como uma seqüência de símbolos quantificados ou quantificáveis. [...] Como são símbolos quantificáveis, dados podem obviamente ser armazenados em um computador e processados por ele. Em nossa definição, um dado é necessariamente uma entidade matemática e, desta forma, puramente sintática. Isto significa que os dados podem ser totalmente descritos através de representações formais, estruturais.

Ao passo que a “Informação é uma abstração informal (isto é, não pode ser formalizada através de uma teoria lógica ou matemática), que representa algo significativo para alguém através de textos, imagens, sons ou animação.” (SEZTER, 1999). Para o autor, este não é um conceito e sim uma caracterização, visto que há termos, como “algo”, “significativo” e “alguém” que não estão bem definidos. Sezter entende que a informação requer compreensão, significado e conhecimento prévio, pode ser associada a outra informação e ao conhecimento prévio, o que é dependente de quem está em contato com a informação. Assim, considera que apenas o receptor pode determinar se o dado ou a mensagem que ele recebe é realmente uma informação.

Não é possível processar informação diretamente em um computador. Para isso é necessário reduzi-la a dados. [...] A representação da informação pode eventualmente ser feita por meio de dados. Nesse caso, pode ser armazenada em um computador. Mas, atenção, o que é armazenado na máquina não é a informação, mas a sua representação em forma de dados. (SEZTER, 1999).

Neste ponto, Sezter (1999) apresenta uma distinção entre dado e informação, a sintática e a semântica, respectivamente. Esclarece ainda o autor que

[...] os dados que representam uma informação podem ser armazenados em um computador, mas a informação não pode ser processada quanto a seu significado, pois depende de quem a recebe. O conhecimento, contudo, não pode nem ser inserido em um

computador por meio de uma representação, pois senão foi reduzido a uma informação. (SEZTER, 1999).

Sezter (1999) entende que o conhecimento está atrelado às experiências vivenciadas por alguém (pragmática), ao passo que a informação consiste em um registro que pode ser compreendido (semântica), o que requer conhecimento prévio, enquanto o dado consiste em uma representação formal e estruturada (sintática).

Para Wurman (1991, p. 42) “A palavra ‘informação’ sempre foi [...] empregada para definir diversos conceitos. Os dicionários registram que a palavra tem sua raiz no latim *informare*, que significa ‘a ação de formar matéria, tal como pedra, madeira, couro, etc.” (WURMAN, 1991, p. 42).

Enquanto para Sayão e Sales (2020, p. 32) “A palavra ‘dado’ é o plural do latim *datum*, que significa dádiva, oferta ou algo reconhecido e usado como base para cálculos, é o vínculo primordial com os fenômenos de amplo espectro em que estamos imersos.”

Davenport (1998) entende que a distinção entre dado e informação é imprecisa, e que há quem se refira aos dados como informação, ao passo que, àquela época, o conhecimento passou a ser utilizado para tratar da informação. Diante disso, o autor apresenta uma ilustração que esclarece os elementos para dados, informação e conhecimento (Quadro 7), nesta tese restritos a dados e informação, considerando serem as categorias temáticas do estudo.

Quadro 7 – Dados e Informação

Dados	Informação
<p>Simple observações sobre o estado do mundo</p> <ul style="list-style-type: none"> · Facilmente estruturado · Facilmente obtido por máquinas · Frequentemente quantificado · Facilmente transferível 	<p>Dados dotados de relevância e propósito</p> <ul style="list-style-type: none"> · Requer unidade de análise · Exige consenso em relação ao significado · Exige necessariamente a mediação humana

Fonte: Adaptado de Davenport (1998, p. 18).

Para Wurman (1991), os dados brutos podem não ser informação à medida que não são utilizados para informar e, com isso, não são compreensíveis. Destaca-se que essa afirmação se aplica ao consumo e ao processamento de dados brutos

por humanos e máquinas, cujas características de consumo são diversas, contudo, são claramente beneficiados com o tratamento dos dados. Apenas assim os dados receberão o contexto necessário para consumo e processamento adequados.

A compreensão da diferença entre dados brutos e aqueles que podem ajudar na compreensão e aumentar o conhecimento, entre informação como coisa e informação como significado, tornará você um processador de informação mais competente. Maior competência lhe dará mais confiança e mais controle, permitindo-lhe relaxar. Sentindo-se mais relaxado e menos culpado, você chegará à compreensão. (WURMAN, 1991, p. 45).

Iyamu (2019, p. 2, tradução nossa) afirma que “O refinamento ou transformação de dados em entidades significativas é informação.”¹⁵⁶

Santos e Sant’Ana (2002) a partir da análise de diferentes autores tratam “[...] o termo *dado* como um elemento básico, formado por signo ou conjunto finito de signos que não contém, intrinsecamente, um componente semântico, mas somente elementos sintáticos.”

Ao passo que a informação é entendida como

[...] um conjunto finito de dados dotado de semântica e que tem a sua significação ligada ao contexto do agente que a interpreta ou recolhe e de fatores como tempo, forma de transmissão e suporte utilizado. O valor desse conjunto poderá diferir da soma dos valores dos dados que o compõem, dependendo do processo de contextualização no agente que o recebe. (SANTOS; SANT’ANA, 2002).

Santos e Sant’Ana (2002), relacionam dados aos aspectos sintáticos e informação aos aspectos semânticos, visão que corrobora com Sezter (1999). Os autores afirmam ainda que é possível a:

[...] existência da informação, independentemente da existência de um agente humano, como emissor ou receptor, permitindo o uso do conceito de informação também para o conjunto de dados, sendo transmitido de um agente humano para outro agente humano, via algum canal (conversa), transmitido de um agente humano para um agente não humano dotado de alguma capacidade de tratamento desta informação, ou a partir deste último para um agente humano (interação) e, finalmente, a possibilidade de transmissão entre agentes não humanos (conexão). (SANTOS; SANT’ANA, 2002).

Carvalho e Lorena (2017) afirmam que o relacionamento entre dados, informação e conhecimento tem sido estudado por pesquisadores da Ciência da Informação e o representa por meio de uma hierarquia ou pirâmide, cuja base são os dados, sobrepostos pela informação e cujo topo é ocupado pelo conhecimento.

¹⁵⁶ “The refinement or transformation of data into meaningful entities is information.”

Com base nessa pirâmide, Carvalho e Lorena (2017, p. 50) entendem que os dados “[...] são fruto de observações de objetos presentes no mundo. São valores que, por si sós, não oferecem base para o entendimento de um estado ou situação. Os dados em geral fazem parte de um conjunto [...] uma coleção de objetos. Um mesmo objeto ou valor pode aparecer mais de uma vez em um conjunto de dados.”

Ao passo que,

A informação surge da análise e organização dos dados de maneira a ter algum valor adicional. Logo, as informações são usualmente extraídas a partir de relações existentes entre os dados, ou entre os dados e o domínio ou problema de onde foram extraídos, tornando-os, de alguma forma, úteis. (CARVALHO; LORENA, 2017, p. 50).

Aprende-se das afirmações que os dados são compreendidos pelos autores para além dos valores numéricos, entendidos como textos, sons, imagens, símbolos e sinais que, independentemente do formato de armazenamento, são codificados de forma numérica, e cuja contextualização e manipulação possibilitam obter novas informações.

Aparentemente, a linha que separa o dado da informação é tênue e, pode ser esclarecida do ponto de vista computacional como sendo os dados processáveis por computadores, tendo na sintática a estrutura, a estética e a forma e, na semântica a contextualização, o sentido e o significado. Ao passo que, do ponto de vista humano, a informação é o dado contextualizado e processado por meio de relacionamentos existentes com outros dados e informações presentes no tempo/espço de apreensão pelo sujeito e, ainda, no processamento com os conhecimentos prévios presentes na estrutura cognitiva do indivíduo.

Após um estudo em busca de revisitar as diferentes perspectivas acerca da natureza da informação, a partir de áreas como Teoria da Comunicação, Biblioteconomia e Ciência da Informação, Sistemas de Informação e Ciência da Computação, Rowley (1998) conclui que um conceito de informação pode variar de acordo com a área, o interesse, o contexto e a cultura, além de mudar ao longo do tempo. Desta forma, reconhece que a informação pode ser definida, mas que isso ocorre em determinados contextos, o que possibilita que múltiplas definições sejam aceitáveis e que sejam encaradas de forma positiva. Porém, alerta que, à medida que o acesso à informação e a comunicação são mediados pelas TIC, é imprescindível que haja debate acerca dos conceitos e definições, buscando evolução e convergência.

Nesse sentido, em busca de alicerçar os elementos necessários à compreensão dos dados por humanos e máquinas, é que esta tese busca aportes teórico-conceituais acerca das temáticas dado e informação, a fim de amalgamá-las às *expertises* da Ciência da Informação e, finalmente, utilizá-las para investigar e compreender a Arquitetura de Dados a fim de conceituá-la no contexto da Ciência da Informação e relacioná-la com a Arquitetura da Informação.

Posicionam-se com destaque nesta tese as discussões apresentadas por Buckland (1991), que entende que há dificuldade em definir informação de forma satisfatória e que atenda completamente a uma necessidade. Assim, o autor afirma que a informação pode ser entendida por “informação como processo”, “informação como conhecimento” e “informação como coisa”.

A informação-como-coisa é de especial interesse no estudo de sistemas de informação. [...] A intenção pode ser que os usuários se tornem informados (informação-como-processo) e que haja uma transmissão de conhecimento (informação-como-conhecimento). Mas os meios fornecidos, o que é manuseado e operado, o que é armazenado e recuperado, é a informação física (informação-como-coisa).¹⁵⁷ (BUCKLAND, 1991, p. 352, tradução nossa).

Na visão de Buckland (1991), a informação como coisa se caracteriza pelo uso indistinto do termo informação como sendo algo utilizado para informar, comunicar e transmitir conhecimento, o que se faz por meio da expressão em forma física, por meio da descrição ou representação, o que também pode ser entendido como objetos, documentos e dados.

Buckland (1991, p. 352, tradução nossa) complementa que “Essa discussão introdutória pode ser completada com referência a um quarto elemento: processamento de informações, manipulação, manipulação e derivação de novas formas ou versões de informação-como-coisa.”¹⁵⁸. Neste contexto, considera que para tornar-se informado é necessário um processamento da informação, o que nesta tese considera-se como tratamento de dados, que confere a eles estruturas sintática e semântica, bem como contexto. O Quadro 8 ilustra a distinção dos quatro aspectos da informação elencados por Buckland.

¹⁵⁷ “Information-as-thing is of special interest in the study of information systems. [...] The intention may be that users will become informed (information-as-process) and that there will be an imparting of knowledge (information-as-knowledge). But the means provided, what is handled and operated upon, what is stored and retrieved, is physical information (information-as-thing).”

¹⁵⁸ “This introductory discussion can be rounded out by reference to a fourth element: information processing, the handling, manipulating, and deriving of new forms or versions of information-as-thing.”

Quadro 8 – Quatro aspectos da Informação

	INTANGÍVEL	TANGÍVEL
ENTIDADE	2. Informação como conhecimento Conhecimento	3. Informação como coisa Dado, documento
PROCESSO	1. Informação como processo Tornando-se informado	4. Processamento da informação Processamento de dados

Fonte: Buckland (1991, p. 352, tradução nossa).

Para a discussão, ainda é necessário lembrar que Buckland (1991) parte do entendimento de que informação é algo utilizado para informar, e elenca tipos de informação: dados, textos e documentos, objetos e eventos que “[...] são coletados, armazenados, recuperados e examinados como informação, como base para se tornar informado.”¹⁵⁹ (p. 354, tradução nossa). E apresenta dados como sendo

“Dados”, como a forma plural da palavra latina “*datum*”, significa “coisas que foram dadas”. É, portanto, um termo adequado para o tipo de informação-como-coisa que foi processada de alguma forma para uso. Comumente, “dados” denota quaisquer registros armazenados em um computador.¹⁶⁰ (BUCKLAND, 1991, p. 353, tradução nossa).

É importante considerar que o uso indistinto da palavra informação, sem um complemento adjetivado, a exemplo do que faz Buckland (1991) ao utilizar a expressão “informação como coisa”, pode causar inconsistência semântica, uma vez que, por exemplo, ainda que considerado como “informação como coisa”, um dado necessita de tratamento para que faça sentido como informação e, com isso, exerça a função de informar.

Isso justifica a defesa feita pelo autor de que, se a informação é algo com potencial para informar, ela não pode ser restrita a dados estruturados, pois há coisas informativas para além de textos, números e outras tipologias entendidas como informação, e há falta de clareza de quanto e qual processamento é necessário para que dados sejam considerados informação, entendendo-a como uma qualidade para qualquer coisa com potencial de ser informativa e com a qual alguém possa aprender.

Santos e Sant’Ana (2013) entendem que o conceito de dado precisa ser compreendido “como elemento básico nos fluxos informacionais”. Com isso, pautam

¹⁵⁹ “[...] are collected, stored, retrieved, and examined as information, as a basis for becoming informed.”

¹⁶⁰ “Data,” as the plural form of the Latin word “datum,” means “things that have been given.” It is, therefore, an apt term for the sort of information-as-thing that has been processed in some way for use. Commonly “data” denotes whatever records are stored in a computer.”

o estudo na reflexão acerca dos conceitos de dado, conjunto de dados e granularidade sob o prisma da Informação e Tecnologia no domínio da Ciência da Informação.

Neste estudo, os autores entendem dados como

[...] uma unidade de conteúdo necessariamente relacionada a determinado contexto e composta pela tríade entidade, atributo e valor, de tal forma que, mesmo que não esteja explícito o detalhamento sobre contexto do conteúdo, ele deverá estar disponível de modo implícito no utilizador, permitindo, portanto, sua plena interpretação. (SANTOS; SANT'ANA, 2013).

Diante do exposto, tem-se o entendimento nesta tese de que enquanto Buckland (1991) considera informação como coisa, deve ser mantida a terminologia adequada, quer seja como dado, documento ou objeto, uma vez que eles possuem características próprias que os diferenciam entre si e que, por elas mesmas, em primeira instância, não podem ser considerados indistintamente como informação.

Essa alteração de perspectiva, na qual se propõe o uso adequado da terminologia informação e que parte do entendimento de Buckland (1991) de que informação como coisa consiste na materialização da informação, como descrição e representação, armazenadas em sistemas de informação, reposiciona o objeto de estudo da Ciência da Informação, que também utiliza o termo informação de forma indistinta, ao passo que a área se ocupa de toda a estrutura necessária ao tratamento de DIKW (ACKOFF, 1989), com atuação preponderante no tratamento dos dados.

Vale ressaltar que, nesta tese, considera-se a hierarquia piramidal proposta na DIKW inadequada¹⁶¹, uma vez que o processamento, sobretudo humano, implicado na teoria não é linear e que a estrutura cognitiva humana realiza conexões e inter-relacionamentos entre dados, informações e conhecimentos de forma contínua e randômica, envolvendo elementos novos e preexistentes, possibilitando que surjam *insights* e a apreensão de novos conhecimentos.

Neste contexto, Pinheiro (2005, p. 40) afirma que

A cadeia conceitual que caracteriza a Ciência da Informação vai desde o dado à informação e conhecimento, de acordo com a idéia de muitos de seus autores, algumas vezes incluindo saber, num crescendo de complexidade, da forma bruta e primitiva do dado à sua elaboração como informação, e sua absorção, quando relevante, na estrutura cognitiva, transformando-se em conhecimento. Esta rede de conceitos poderá ter seu processo final na cultura, aqui considerando a incorporação dessas informações relevantes entre outras

¹⁶¹ No contexto da CI, a hierarquia DIK também é rejeitada por Siqueira (2012) que apresenta uma proposta para a Arquitetura da Informação designada Hierarquia Informação-Dado-Conhecimento.

manifestações e produções e vivências do homem, individuais e coletivas.

Santos e Sant'Ana (2013) entendem que é por meio da estrutura que o uso e reuso de dados se viabiliza. Assim, “Assumindo que a estrutura de um dado é estabelecida pela identificação dos seguintes elementos <e> e <a>, que pertencem à tríade que compõe o seu esquema, a questão que surge é a diferenciação entre dados estruturados e não estruturados.”¹⁶² (SANTOS; SANT'ANA, 2013, p. 206). Para os autores, o dado estruturado possui semântica que permite a interpretação automática por máquinas, enquanto o dado não estruturado necessita de um interpretador.

Silva (2017, p. 26) afirma que “[...] separar dado de informação é um exercício inútil.” Para o autor, “[...] dado, entendido, e isto é muito importante, como uma representação mental e emocional codificada, é informação.” (p. 24-25). Silva (2017) pondera que o dado já comporta sentido, e que, em primeira instância, o produtor do sentido é o emissor para quem o dado claramente possui sentido e com isso é considerado informação.

Tem-se ainda em Stonier (1997, p. 49) a percepção de que o armazenamento e o processamento de informações entre humanos e entre máquinas é substancialmente diferente, inclusive considerando a mensagem que, no primeiro caso,

[...] envolve a codificação do morfema (no caso da palavra falada) e a letra (no caso da linguagem escrita, fonética). A mensagem resulta da criação de padrões de morfemas ou letras. Isso é muito diferente de codificar chaves liga/desliga e produzir mensagens encadeando pulsos de elétrons, que é a base dos IPSs [acrônimo para sistemas de processamento de informações] eletrônicos.¹⁶³

É importante esclarecer que, para Stonier (1997), a informação se manifesta como padrões de organização. Assim, qualquer sistema que possua uma organização possui informações. Além disso, o autor entende que o processamento da informação “[...] inclui a transdução, transporte, armazenamento e recuperação da informação, bem como a sua análise.”¹⁶⁴ (STONIER, 1997, p. 182), não se restringindo ao

¹⁶² Para Santos e Sant'Ana (2013) a tríade é composta por entidade <e>, atributo <a> e valor <v>.

¹⁶³ “[...] involves coding the morpheme (in the case of the spoken word) and the letter (in the case of written, phonetic language). The message results from having created patterns of morphemes or letters. This is very different from coding on/off switches and producing messages by stringing together pulses of electrons, which is the basis of electronic IPSs.”

¹⁶⁴ “[...] includes the transduction, transport, storage and retrieval of information, as well as its analysis.”

processamento realizado por computadores. Assim, defende que a “[...] informação não precisa ser interpretada para que ela exista.”¹⁶⁵ (STONIER, 1997, p. 186).

Floridi (2010) propõe um mapa (Figura 7) no qual posiciona os diferentes sentidos nos quais a informação pode ser abordada, partindo do conceito de dados. A indicação “Você está aqui” é alterada no mapa nos diversos capítulos da obra, como forma de indicar ao leitor os conceitos tratados e possibilitar reorientação.

Figura 7 – Mapa de conceitos de informação



Fonte: Floridi (2010, p. 20, tradução nossa).

Na Figura 7, Floridi (2010) busca ilustrar as diferentes formas de compreender a informação, iniciando pela definição baseada em dados. Neste sentido, menciona a adoção da *General Definition of Information* (GDI) como “dados + significado”. No contexto da GDI¹⁶⁶ a “[...] informação é feita de dados”¹⁶⁷ bem formados, que consistem em “[...] dados [que] estão corretamente reunidos, de acordo com as regras (sintaxe) que regem o sistema, código ou idioma escolhido.”¹⁶⁸, e “devem estar de acordo com os significados (semânticos) do sistema em questão, código ou idioma escolhido.”¹⁶⁹ (FLORIDI, 2010, p. 20-21, tradução nossa).

¹⁶⁵ “[...] that information does not need to be interpreted in order for it to exist.”

¹⁶⁶ Para Floridi (2010, p. 21, tradução nossa) “GDI) a é uma instância de informação, entendida como conteúdo semântico, se e somente se: GDI.1) a consiste em n dados, para $n > 1$; GDI.2) os dados estão bem formados; GDI.3) os dados bem formados são significativos.”

¹⁶⁷ “[...] information is made of data.”

¹⁶⁸ “[...] data are rightly put together, according to the rules (syntax) that govern the chosen system, code, or language being used.”

¹⁶⁹ “[...] must comply with the meanings (semantics) of the chosen system, code, or language in question.”

Para o autor, “[...] os dados que constituem a informação podem ser significativos independentemente de um informado.”¹⁷⁰ (FLORIDI, 2010, p. 22, tradução nossa). Floridi (2010) defende que é possível que haja dados e não haja informação, uma vez que eles não possam ser compreendidos. Com isso, infere-se que enquanto o dado não necessita de um receptor, por ele considerado informado, a informação consiste no contato com o dado por um receptor que possa interpretá-lo e compreendê-lo.

Essa inferência é bastante significativa nos processos que envolvem humanos, cuja estrutura cognitiva possibilita acionar o conhecimento prévio que auxilia na compreensão dos dados. Deve-se considerar ainda que, na ausência de conhecimento prévio, outros dados podem ser compreendidos, tornando-se parte do arcabouço informacional do sujeito de forma a permitir a compreensão dos dados que aciona. De igual maneira, a estrutura cognitiva humana possibilita que os conhecimentos adquiridos sejam materializados em forma de dados.

Por outro lado, considerando a afirmação de Floridi (2010), tem-se em conta que os processos que envolvem máquinas necessariamente atuam com base em dados, que necessitam de sintaxe e semântica para que possam ser compreensíveis por outras máquinas, sejam elas aplicações ou agentes computacionais, e também para que tenham potencial de auxiliar a compreensão por humanos.

Desta perspectiva, é possível afirmar que humanos geram e processam dados e informações, ao passo que máquinas geram e processam dados, que quando bem estruturados possuem potencial de se tornar informação por um interagente que o consoma por processos analógicos, digitais ou híbridos.

A Figura 7, anteriormente apresentada, ilustra ainda que os dados podem ser analógicos, digitais ou híbridos, evidentemente dependentes das tecnologias disponíveis para elaboração/coleta/compilação/interpretação, seleção, representação, armazenamento, tratamento, acesso e (re)uso e, de igual maneira, passíveis de fortalezas e fragilidades próprios dos suportes e metodologias utilizados.

Nesse contexto, considerando os dados digitais, a estruturação sintática e semântica com o uso de padrões internacionais reconhecidos é que possibilita a compreensão por máquinas, favorecendo o fluxo de dados entre ambientes informacionais digitais e sistemas de informação, bem como se tornando fonte para o

¹⁷⁰ “[...] data constituting information can be meaningful independently of an informee.”

aprendizado de máquinas. Considerando que esta mesma estrutura amplia o alcance e o potencial de acesso aos dados por outras máquinas, há também benefícios para usuários humanos, como a facilidade de descoberta, processamento, acesso e uso, quer seja pela manipulação em meio analógico, digital ou híbrido.

Nesta tese, tem-se discutido que, no contexto atual de ampla utilização das TIC, há a necessidade de compreensão de dados e informações por humanos e máquinas. Nesse sentido, tendo em vista Silva (2017), é possível afirmar que, na comunicação entre humanos, há um emissor de informação, imbuída de sentido, que é transmitida por meios tangíveis e/ou intangíveis e que, para o receptor, pode ser dado ou informação, considerando a necessidade de atribuir sentido. Desta forma, considerando Silva (2017) quando se trata da comunicação entre humanos, não há um emissor de dados.

Ao passo que, no contexto das máquinas, o que se tem é uma representação da informação, que não a substitui, mas que a descreve e que, em princípio, é constituída por dados, dependentes de uma estrutura semântica, constituída pelos vocabulários padronizados de metadados, para que estejam acompanhados de sentido. Assim, no contexto das máquinas, quando o dado é adequadamente estruturado na sua fonte de armazenamento, ele está acompanhado de uma estrutura que lhe fornece sentido, mas necessita da compreensão para que seja informação.

Com isso, parece possível afirmar que essa compreensão necessária para que o dado seja informação é factível apenas para humanos, uma vez que as máquinas permanecerão processando dados, que quando estruturados são imbuídos de sentido, tornando a comunicação entre máquinas mais fluída.

Neste sentido, para que esses dados se tornem informação, no contexto das máquinas, é necessário tratamento dos dados por meio de agentes computacionais oriundos, por exemplo, da Computação Cognitiva e da Inteligência Artificial. Esses agentes podem utilizar, por exemplo, bases de conhecimentos, regras de inferência e algoritmos de aprendizagem de máquina para simular a compreensão.

Com base nas reflexões realizadas, nesta tese, posiciona-se um conceito operacional de dados, para efeito teórico nesta tese.

Dados são compreendidos como um fenômeno social e/ou maquínico, materializados e utilizados para registrar fatos e/ou representar atributos de entidades e seus contextos, que podem ser estruturados em diferentes níveis e padrões necessários para a comunicação efetiva entre sistemas de informação, aplicações e agentes computacionais em benefício do processamento computacional e da utilização por consumidores humanos e máquinas. Tendo em vista que o dado por si só é insuficiente para a interpretação, há a necessidade de documentação que auxilie o (re)uso.

4.2 ARQUITETURA DE DADOS: ANTECEDENTES

As décadas de 2010 e 2020 têm sido marcadas pela crescente expansão da discussão acerca do valor dos dados para as organizações, referindo-se a eles como um ativo ou o “novo petróleo das organizações”.

De acordo com DAMA-DMBOK (2017, p. 17, tradução nossa) “Muitas organizações reconhecem que seus dados são um ativo empresarial vital. Dados e informações podem dar-lhes uma visão sobre seus clientes, produtos e serviços. Podem ajudá-las a inovar e alcançar objetivos estratégicos.”¹⁷¹. Por outro lado, DAMA-DMBOK também reconhece que muitas organizações ainda encontram dificuldades em gerenciar os dados como um ativo.

Vale destacar que, na história da humanidade, os dados e informações sempre estiveram presentes, e passaram a ser registrados pelas comunidades ancestrais na forma de inscrições rupestres.

O advento das tecnologias permitiu novas formas de registro de dados, e as tecnologias computacionais possibilitaram o registro, o armazenamento, o acesso e o uso por meio de sistemas de informação e ambientes informacionais digitais.

Os dados devem ser considerados como úteis quando possibilitam a análise (INMON; LINSTEDT, 2015). Inicialmente, isso requer que os dados tenham qualidade, veracidade, confiabilidade e precisão. Posteriormente, por meio da estruturação e com o apoio da tecnologia, é possível que os dados sejam gerenciados de forma a embasar análises, suportar tomada de decisões e outras formas de gerar valor.

¹⁷¹ “Many organizations recognize that their data is a vital enterprise asset. Data and information can give them insight about their customers, products, and services. It can help them innovate and reach strategic goals.”

Para Sayão e Sales (2020) “[...] nas últimas décadas, toda a sociedade experimenta um fenômeno inédito que tem como ponto de inflexão uma mudança na curva de disponibilidade de informação: da escassez à extrema abundância de dados.” Tal fato pode convergir com a expansão das tecnologias computacionais e com fenômenos como o *Big Data*.

Neste contexto, de acordo com DAMA-DMBOK (2017), os dados são intangíveis, duráveis, fáceis de transportar, difíceis de reproduzir, são dinâmicos, não se esgotam com o uso, possuem valor e difícil mensuração.

Carvalho e Lorena, (2017, p. 49) afirmam que

Os dados são a entrada e a saída de um Sistema de Computação. As instruções passadas por um programa para o hardware do computador basicamente armazenam, manipulam e recuperam dados. Para isso, esses dados precisam ser primeiramente representados no hardware. (CARVALHO; LORENA, 2017, p. 49).

Neste sentido,

Os computadores processam **dados** de acordo com um conjunto de instruções, chamado de **programa** ou **software**. Um programa desencadeia uma série de operações na **máquina**, ou **hardware**, que fazem com que a tarefa especificada por ele seja realizada. Assim, computadores, ou **sistemas de Computação**, são formados por três pilares principais: dados, hardware e software. (CARVALHO; LORENA, 2017, p. 48).

Essas afirmações posicionam os dados e o seu uso, considerando as tecnologias computacionais conforme seu desenvolvimento à época. DAMA-DMBOK (2017, p. 17, tradução nossa) complementa: “Derivar valor dos dados não acontece no vácuo ou por acidente. Requer intenção, planejamento, coordenação e comprometimento. Exige gestão e liderança.”¹⁷².

Há que se considerar, com base em Yoon, Aiken e Guimaraes (2005), que o desenvolvimento de sistemas requer modelagem de dados, muitas vezes executada apenas para o próprio sistema, sem ponderar sua utilidade para o compartilhamento e intercâmbio de dados dentro da própria organização, o que tem gerado a necessidade de atenção aos requisitos necessários para que os sistemas de informação sejam mais integrados. Na visão dos autores, essa integração requer atenção à qualidade dos dados e, por isso, necessita de uma Arquitetura de Dados “[...] para coordenar as atividades de gerenciamento de dados no desenvolvimento e

¹⁷² “Deriving value from data does not happen in a vacuum or by accident. It requires intention, planning, coordination, and commitment. It requires management and leadership.”

nas operações de sistemas multifuncionais.”¹⁷³ (YOON; AIKEN; GUIMARAES, 2005, p. 148, tradução nossa), cuja característica principal é estabelecer uma ligação entre os requisitos de dados de toda a organização e os sistemas implementados.

DAMA-DMBOK (2017) aponta ainda que os dados, quando compreendidos como aqueles armazenados em ambientes digitais, podem levar a entender como dados muitas coisas que antes não seriam assim consideradas. E que, por serem os dados representação de fatos, eles tendem a ser considerados como verdades sobre o mundo. Contudo, alertam que os dados podem ser uma forma de representação de si mesmo e de fatos, e que há formas diferentes de representar fatos e conceitos, o que permite que os dados assumam diferentes formas.

Neste sentido, DAMA-DMBOK (2017) enfatiza que, mesmo em uma organização, há diferentes formas de representar a mesma coisa e que essa possibilidade de multiplicidades gera a necessidade de criar padrões que possam trazer consistência aos dados, que reforçam a necessidade de “Arquitetura de Dados, modelagem, governança, administração, gerenciamento de Metadados e Qualidade de Dados, que ajudam as pessoas a entender e usar os dados.”¹⁷⁴ (DAMA-DMBOK, 2017, p. 19).

Por outro lado, Tupper (2011) afirma que, no contexto da Tecnologia da Informação, a arquitetura tem sido frequentemente ignorada. “Arquitetura - o método de projetar e planejar as coisas antes de serem construídas - está sendo negligenciada ou ignorada na pressa de desenvolver e entregar o produto. O foco mudou do processo para o produto.”¹⁷⁵ (TUPPER, 2011, p. xxii, tradução nossa).

Vale destacar que a necessidade de integrar diferentes sistemas de informação e ambientes informacionais digitais existentes na organização e fora dela tem se tornado constante, e chega a ser considerada por Yoon, Aiken e Guimaraes (2005) como um problema que levou a alterações no campo de desenvolvimento de sistemas, cujo foco em sistemas isolados deu lugar ao suporte multifuncional e interorganizacional, o que requer o tratamento adequado dos dados para assegurar sua qualidade, considerando: valor dos dados, representação, modelo de dados e qualidade da Arquitetura de Dados.

¹⁷³ “[...] to coordinate data management activities in cross-functional system development and operations.”

¹⁷⁴ “Data Architecture, modeling, governance, and stewardship, and Metadata and Data Quality management, all of which help people understand and use data.”

¹⁷⁵ “Architecture—the method of design and planning things before they are constructed—is being overlooked or bypassed in the haste to develop and deliver the product. The focus has shifted from the process to the product.”

Inmon e Linstedt (2015), a partir do contexto corporativo, afirmam que o uso de dados para análises requer inicialmente a definição de fontes nas quais os dados estão disponíveis, podendo se tratar de fontes analógicas, que necessitarão de tratamento para se tornarem digitais e, no que tange às digitais, apontam como desafio a resolução lógica, na qual diferentes fontes digitais de dados podem ser necessárias para coletar dados, que precisarão ser reunidos e processados. Inmon e Linstedt (2015, p. 28, tradução nossa) apontam alguns desses desafios:

Resolver estruturas de chave nas quais uma chave em uma parte da corporação é diferente de uma chave semelhante em outra parte da corporação.

Resolver definições em que os dados definidos de uma forma na corporação são definidos de outra forma em uma parte diferente da corporação.

Resolver cálculos em que um cálculo feito de uma forma na corporação é feito usando uma fórmula diferente em outra parte da corporação

Resolver estruturas de dados nas quais os dados estruturados de uma maneira na corporação são estruturados de maneira diferente em outra parte da corporação.

E a lista continua.¹⁷⁶

Os autores destacam que essas questões podem estar tão arraigadas que tornam as possibilidades de solução pouco satisfatórias. Assim, há a necessidade de integrar os dados em uma estrutura coesa.

Essas discussões coadunam com Inmon, Linstedt e Levins (2019, p. 293, tradução nossa) quando afirmam que “A forma como os dados são usados, a forma como os dados são moldados e a forma como os dados são armazenados progrediu ao ponto de que atualmente existe uma área de estudo que pode ser chamada de arquitetura de dados.”¹⁷⁷ Para os autores, em função da complexidade dos dados, a Arquitetura de Dados possui muitas facetas, dentre as quais destacam: a manifestação física, a ligação lógica, o formato interno e a estrutura de armazenamento de dados.

¹⁷⁶ “Resolving key structures in which a key in one part of the corporation is different from a similar key in another part of the corporation.

Resolving definitions in which data defined one way in the corporation is defined another way in a different part of the corporation.

Resolving calculations in which a calculation made one way in the corporation is made using a different formula in another part of the corporation

Resolving data structures in which data structured one way in the corporation is structured differently in another part of the corporation.

And the list goes on.”

¹⁷⁷ “The way that data are used, the way data are shaped, and the way that data are stored has progressed to the point that there is actually now an area of study that can be called data architecture.”

Yoon, Aiken e Guimaraes (2005, p. 151, tradução nossa) definem Arquitetura de Dados como “O conjunto subjacente de regras e descrições de relacionamentos que governam como os principais tipos de dados suportam os processos de negócios definidos na arquitetura de negócios.”¹⁷⁸. Esta definição foi incluída ao *Dictionary of Information Science and Technology* (KHOSROW-POUR, 2007, p. 153).

De acordo com DAMA-DMBOK (2017, p. 100, tradução nossa), Arquitetura de Dados é definida como:

Identificar as necessidades de dados da empresa (independentemente da estrutura) e projetar e manter os planos mestres para atender a essas necessidades. Usando *blueprints* mestres para orientar a integração de dados, controlar ativos de dados e alinhar investimentos em dados com a estratégia de negócios.¹⁷⁹

Ainda, para DAMA-DMBOK (2017, p. 100, tradução nossa), a Arquitetura de Dados tem como objetivos:

1. Identificar requisitos de armazenamento e processamento de dados.
2. Projetar estruturas e planos para atender aos requisitos de dados atuais e de longo prazo da empresa.
3. Preparar estrategicamente as organizações para evoluir rapidamente seus produtos, serviços e dados para aproveitar as oportunidades de negócios inerentes às tecnologias emergentes.¹⁸⁰

Desta forma, de acordo com DAMA-DMBOK (2017, p. 45, tradução nossa), “A Arquitetura de Dados define o plano para gerenciar ativos de dados, alinhando-se à estratégia organizacional para estabelecer requisitos de dados estratégicos e projetos para atender a esses requisitos.”¹⁸¹

Para Laun, Mazzuchi e Sarkani (2022, p. 119, tradução nossa) “Uma arquitetura de dados normalmente fornece uma estrutura ampla que conecta ativos de dados exclusivos a estratégias e requisitos organizacionais de dados.”¹⁸²

¹⁷⁸ “The underlying set of rules and descriptions of relationships that govern how the major kinds of data support the business processes defined in the business architecture.”

¹⁷⁹ “Identifying the data needs of the enterprise (regardless of structure), and designing and maintaining the master blueprints to meet those needs. Using master blueprints to guide data integration, control data assets, and align data investments with business strategy.”

¹⁸⁰ “1. Identify data storage and processing requirements. 2. Design structures and plans to meet the current and long-term data requirements of the enterprise. 3. Strategically prepare organizations to quickly evolve their products, services, and data to take advantage of business opportunities inherent in emerging technologies.”

¹⁸¹ “Data Architecture defines the blueprint for managing data assets by aligning with organizational strategy to establish strategic data requirements and designs to meet these requirements.”

¹⁸² “A data architecture typically provides a broad structure that connects unique data assets to organizational data strategies and requirements.”

Enquanto para Dow, Hackbarth e Wong (2013, p. 1345, tradução nossa) “As arquiteturas de dados representam a maneira como os dados são armazenados e organizados logicamente em um SI [sistema de informação].”¹⁸³ E complementam “As arquiteturas de dados identificam como os dados são armazenados e organizados logicamente.”¹⁸⁴ (DOW; HACKBARTH; WONG, 2013, p. 1348, tradução nossa).

Na prática “As arquiteturas de dados são o coração da funcionalidade de negócios. Dada a arquitetura de dados adequada, todas as funções possíveis podem ser concluídas dentro da empresa com facilidade e rapidez.”¹⁸⁵ (TUPPER, 2011, p. 57, tradução nossa).

Isso decorre da estruturação dos dados que, uma vez parte da Arquitetura de Dados, pode ser utilizada para diferentes finalidades, por diversos sistemas de informação e ambientes informacionais, para análises e visualizações de maneira unívoca, beneficiando ainda sustentabilidade para novos desenvolvimentos, com base em técnicas e tecnologias existentes em uma estrutura projetada.

Nesse sentido, Tupper (2011, p. 57, tradução nossa) alerta que “[...] é necessário que o gerenciamento organizacional trate os dados como o recurso e o ativo críticos que eles são.”¹⁸⁶

Diante da ausência de atributos relacionados à Arquitetura de Dados, Yoon, Aiken e Guimaraes (2005, p. 148-149, tradução nossa) propõem nove atributos contendo as características desejáveis para a Arquitetura de Dados, que descrevem como:

- a) *Architectural completeness*/Completeness arquitetônica: “A arquitetura é abrangente o suficiente para ser utilizada por qualquer área funcional da organização que deseje utilizá-la.”¹⁸⁷
- b) *Architectural correctness*/Correção arquitetônica: “As informações que descrevem a arquitetura são representadas corretamente com a

¹⁸³ “Data architectures represent the way that data are logically stored and organized in an IS.”

¹⁸⁴ “Data architectures identify how data are logically stored and organized.”

¹⁸⁵ “Data architectures are the heart of business functionality. Given the proper data architecture, all possible functions can be completed within the enterprise easily and expeditiously.”

¹⁸⁶ “[...] it requires that organizational management address data as the critical resource and asset that it is.”

¹⁸⁷ “The architecture is comprehensive enough to be used by any functional area of the organization wishing to utilize it.”

- metodologia apropriada. Ou seja, a organização pode usar a metodologia para manter definições de dados uniformes em toda a organização.”¹⁸⁸
- c) *Management utility*/Utilitário de gestão: “A arquitetura de dados é utilizada pela organização no planejamento estratégico e no desenvolvimento de sistemas como indicativo de sua utilidade.”¹⁸⁹
- d) *Data management quality*/Qualidade da gestão de dados: “A organização como um todo é orientada por dados. Os modelos de dados são desenvolvidos e gerenciados a partir de uma perspectiva de toda a organização, guiada pela arquitetura de dados organizacional. Os dados são gerenciados com controle distribuído de uma unidade centralizada.”¹⁹⁰
- e) *Data sharing ability*/Capacidade de compartilhamento de dados: “A arquitetura de dados serve como base para negociar e implementar acordos de troca de dados dentro da organização, antecipando, definindo e gerenciando os requisitos de compartilhamento de dados dentro da organização e entre seus parceiros de negócios usando definições padrão de metadados em toda a organização.”¹⁹¹
- f) *Functional data quality*/Qualidade de dados funcionais: “Os dados são projetados para dar suporte aos requisitos da área funcional de negócios, onde os elementos de dados para sistemas individuais são derivados de requisitos de metadados organizacionais e implementados usando sistemas organizacionais projetados para oferecer suporte à representação da informação.”¹⁹²
- g) *Data operation quality*/Qualidade da operação de dados: “A engenharia de qualidade de dados é estabelecida como uma área funcional, aplicando de

¹⁸⁸ “The information describing the architecture is correctly represented with the appropriate methodology. That is, the organization can use the methodology to maintain uniform data definitions throughout the organization.”

¹⁸⁹ “The data architecture is widely used by the organization in strategic planning and systems development as an indication of its utility.”

¹⁹⁰ “The organization as a whole is data driven. Data models are developed and managed from an organization-wide perspective, guided by the organizational data architecture. Data are managed with distributed control from a centralized unit.”

¹⁹¹ “The data architecture serves as the basis for negotiating and implementing intra-organizational data exchange agreements by anticipating, defining and managing data sharing requirements within the organization and among its business partners using organization-wide standard metadata definitions.”

¹⁹² “Data are engineered in support of business functional area requirements where data elements for individual systems are derived from organizational metadata requirements and implemented using organizational systems designed to support information representation.”

forma ativa e consistente métodos de engenharia de qualidade de dados aos elementos de dados.”¹⁹³

h) *Evolvability*/Evolução adaptativa: “A arquitetura de dados organizacional é mantida de forma flexível e evolutiva para permitir o cumprimento dos requisitos futuros do usuário.”¹⁹⁴

i) *Organizational self-awareness*/Autoconhecimento organizacional: “Capacidade da organização de investigar o uso da arquitetura e determinar os tipos de valor que ela fornece aos usuários finais. O feedback ajuda os arquitetos de dados a refinar a arquitetura para torná-la mais útil organizacionalmente.”¹⁹⁵

De acordo com Tupper (2011, p. 10, tradução nossa), “Uma arquitetura é a resposta às coleções integradas de modelos e visões dentro da área problemática que está sendo examinada.”¹⁹⁶

Tupper (2011) baseia-se no *The Open Group Architecture Framework* (TOGAF), um *framework* de arquitetura empresarial que consiste no projeto planejamento, implementação e governança de uma arquitetura empresarial ou corporativa. Neste *framework*, a Arquitetura de Dados “[...] descreve como os armazenamentos de dados corporativos são organizados e acessados.”¹⁹⁷ (TUPPER, 2011, p. 29, tradução nossa).

Inmon e Linstedt (2015) afirmam que os dados estão presentes desde o primeiro programa de computador e estão diretamente atrelados, com especial relevância às suas formas de uso, armazenamento e estruturação, o que culminou no campo de Arquitetura de Dados. Para Inmon e Linstedt (2015, p. 199, tradução nossa), há quatro aspectos relevantes na Arquitetura de Dados:

1. A manifestação física dos dados
2. A ligação lógica de dados
3. O formato interno dos dados
4. A estrutura completa de dados.¹⁹⁸

¹⁹³ “Data quality engineering is established as a functional area actively and consistently applying data quality engineering methods to data elements.”

¹⁹⁴ “The organizational data architecture is maintained in a flexible, evolving fashion to enable the fulfillment of future user requirements.”

¹⁹⁵ “Organization ability to investigate architecture use, and determine the types of value that it provides to end users. Feedback helps data architects refine the architecture to make it more useful organizationally.”

¹⁹⁶ “An architecture is the response to the integrated collections of models and views within the problem area being examined.”

¹⁹⁷ “[...] describes how the enterprise data stores are organized and accessed.”

¹⁹⁸ “1. The physical manifestation of data
2. The logical linkage of data

Considerando a arquitetura de dados como parte da arquitetura corporativa, no contexto do desenvolvimento de sistemas, Tupper (2011, p. 60-61, tradução nossa) afirma que “[...] a arquitetura de dados corporativos (ou uma cópia ativa dela) será a fonte e o repositório de todos os modelos de desenvolvimento. Isso garante desenvolvimento consistente, minimização da desintegração e simultaneidade da arquitetura de dados corporativos.”¹⁹⁹

Por outro lado, o autor pondera que pode haver sistemas legado, dos quais não haja registro de projetos e modelos, o que requer um processo de engenharia reversa, na qual “[...] um grupo de gerenciamento de repositório de modelos identifica os processos e entidades envolvidos e, em seguida, prepara um relatório para análise de requisitos de negócios.”²⁰⁰ (TUPPER, 2011, p. 199, tradução nossa). Desta forma, há o que o autor designa de engenharia reversa de modelos físicos para modelos lógicos de sistemas legados.

Os dados presentes em um sistema de informação ou ambiente informacional digital podem ser utilizados interna e externamente, o que amplia a relevância da Arquitetura de Dados, de forma que todos os elementos imprescindíveis ao acesso e (re)uso, sobretudo por máquinas, possam estar adequadamente estruturados, considerando a mensagem, o emissor, o canal e o receptor, agora como agentes maquínicos.

Para tanto, é necessário que os dados e os canais de comunicação estejam preparados, o que inclui a estrutura de segurança que possibilitará, por exemplo, definir os níveis e administradores de permissão de acesso, bem como as formas de acesso. Além disso, ainda que os dados sejam estruturados, pode ser necessário estabelecer um processo que reduza o ruído dessa comunicação.

No que se refere à Teoria Matemática da Comunicação, proposta por Claude E. Shannon em 1948, o que ocorre é a troca de sinais entre emissor e receptor, para que determinada mensagem seja transmitida utilizando canais de comunicação. Embora não tenha se preocupado inicialmente com a semântica dos dados, esse processo, nesta tese, é utilizado visando favorecer o processo de comunicação entre

3. *The internal format of data*

4. *The file structure of data*”

¹⁹⁹ “[...] *the enterprise data architecture (or an active copy of it) will be the source and repository of all development models. This ensures consistent development, minimization of disintegration, and enterprise data architecture concurrency.*”

²⁰⁰ “[...] *a model repository management group identifies the processes and entities involved and then prepares a report for business requirements analysis.*”

humanos, máquinas e humanos-máquinas. Para tanto, é necessário que haja compreensão sintática e semântica entre fornecedor e consumidor, bem como dos padrões, protocolos e serviços disponíveis como canal, para que a mensagem seja transmitida.

Com base na Teoria Matemática da Comunicação, pode ser possível alicerçar parte da estrutura da Arquitetura de Dados, visando reduzir a incerteza e gerar compreensão, para que os dados sejam adequadamente localizados, acessados, coletados e utilizados.

De acordo com Oliveira (2014, p. 152), “A web semântica tem sido estudada e utilizada para o desenvolvimento de ambientes digitais dotados de interoperabilidade semântica a partir de padrões que orientam a estruturação de dados.” Desta forma, algumas de suas técnicas são imprescindíveis para a estrutura dos dados no contexto da Arquitetura de Dados e para que possam ser utilizados facilmente em Ecologias Complexas de Dados.

Tupper (2011, p. 64-65) alerta que as características dos dados são preponderantes para o compartilhamento. Assim, as definições a respeito do compartilhamento de dados auxiliam na produtividade. Neste sentido, aponta requisitos para o compartilhamento de dados, resumidamente:

- a) o compartilhamento de dados deve ser uma política e um padrão;
 - b) todas as propriedades acerca dos dados e sua segurança devem ser definidas, especificando formas de criar, ler, atualizar e excluir dados, bem como aspectos relacionados às alterações e formas de acesso;
 - c) é necessário disponibilizar um glossário ou dicionário de termos de compartilhamento de dados completo e abrangente;
 - d) devem ser adotados padrões para nomenclatura de entidades e atributos, evitando a duplicação, visando favorecer a compreensão dos dados;
 - e) devem ser definidas lógica de validação e regras de tradução para domínios compartilhados, cujos acordos devem ser disponibilizados;
 - f) o modelo de dados compartilhado deve ser simples;
 - g) “Os dados de restrição de domínio (valor válido, intervalos válidos, tradução, existência e dados derivados de algoritmos) devem ser
-

separados dos dados de negócios sem relacionamentos de chave com os dados dos aplicativos.”²⁰¹;

- h) as hierarquias devem ser expressas, visando assegurar que todos os dados estejam disponíveis para compartilhamento;
- i) “O formato lógico de dados deve existir em um formato padrão”²⁰² documentado;
- j) devem ser estabelecidas regras de negócios relativas aos dados;
- k) o prazo e os requisitos para o compartilhamento de dados deve ser uniforme.

Uma das necessidades apontadas pela DAMA-DMBOOK (2017) é a manutenção da modelagem dos dados em um repositório digital, utilizado para gerenciar os dados e metadados em toda a estrutura envolvida.

Tupper (2011, p. 66, tradução nossa) designa esta necessidade como dicionário de dados, que consiste em “[...] um compêndio de todas as definições de dados no nível mais baixo. Ou seja, consiste em nomes de atributos de dados e nas definições e características associadas a eles.”²⁰³. O autor destaca que o dicionário de dados, independente da sua forma de implementação, deve ter seus procedimentos e políticas amplamente divulgados, para que seja utilizado em toda a empresa, favorecendo a compreensão dos dados. Ainda neste contexto, Tupper (2011) aborda a necessidade de estabelecer modelos de dados, para que os processos sejam facilmente compreendidos e, para tanto, aponta a necessidade de manter os modelos de dados em ambientes de armazenamento, que designa como banco de armazenamento de modelo de dados.

Neste sentido, na área de Ciência da Informação, Brandt (2020) propõe uma metodologia de Arquitetura da Informação para processos de trabalho, que inclui o desenvolvimento de uma matriz de metadados e repositório de metadados, e aborda a gestão de metadados, composta por identificação, registro e descrição.

Neste estudo, sobretudo no que tange a compreender os antecedentes da Arquitetura de Dados, vale apresentar a *Common Approach to Federal Enterprise Architecture*, um documento que estabelece orientações para uma abordagem comum

²⁰¹ “Domain constraint data (valid value, valid ranges, translation, existence, and algorithmically derived data) must be separated from business data with no keyed relationships to the applications data.”

²⁰² “The logical data format must exist in a standard form [...]”.

²⁰³ “[...] a compendium of all data definitions at the lowest level. That is, it consists of data attribute names and the definitions and characteristics associated with them.”

de *Enterprise Architecture* (EA) para o Poder Executivo do Governo Federal dos Estados Unidos da América (UNITED STATES, 2012). Embora tratar Arquitetura Corporativa não seja objetivo desta tese, cumpre esclarecer, com base no documento supramencionado que:

Arquitetura Corporativa significa uma base de ativos de informação estratégica, que define a missão; as informações necessárias para realizar a missão, as tecnologias necessárias para realizar a missão e os processos de transição para a implementação de novas tecnologias em resposta às mudanças nas necessidades da missão; e inclui uma arquitetura de linha de base, uma arquitetura de destino e um plano de sequenciamento.²⁰⁴ (UNITED STATES, 2012, p. 45, tradução nossa)

O documento *Common Approach to Federal Enterprise Architecture* estabelece que seja mantida a arquitetura corporativa em todas as Agências Federais²⁰⁵ norte-americanas. “Isso inclui princípios para usar a EA para ajudar as agências a eliminar o desperdício e a duplicação, aumentar os serviços compartilhados, fechar as lacunas de desempenho e promover o engajamento entre governo, indústria e cidadãos.”²⁰⁶ (UNITED STATES, 2012, p. 3, tradução nossa).

Assim, a *Federal Enterprise Architecture* (FEA) estabelece princípios e padrões de EA, para que as referidas arquiteturas, informações e tecnologias sejam consistentes entre as agências governamentais, seus consumidores e parceiros externos, além de fornecer pontos de integração entre as diferentes áreas de governança.

A abordagem comum da FEA fornece princípios e padrões sobre como as arquiteturas de negócios, informações e tecnologia devem ser desenvolvidas em todo o Governo Federal para que possam ser usadas consistentemente em vários níveis de escopo dentro e entre agências, bem como com partes interessadas externas. A abordagem comum fornece pontos de integração com outras áreas de governança, incluindo planejamento estratégico, planejamento de capital, gerenciamento de

²⁰⁴ “Enterprise Architecture means a strategic information asset base, which defines the mission; the information necessary to perform the mission, the technologies necessary to perform the mission, and the transitional processes for implementing new technologies in response to changing mission needs; and includes a baseline architecture, a target architecture, and a sequencing plan.”

²⁰⁵ “[...] o termo “Agência Federal” inclui Departamentos, Agências, Comissões, Escritórios, Conselhos e outros tipos de organizações do Poder Executivo do Governo Federal dos EUA.” (UNITED STATES, 2012, p. 5, tradução nossa).

²⁰⁶ “This includes principles for using EA to help agencies eliminate waste and duplication, increase shared services, close performance gaps, and promote engagement among government, industry, and citizens.”

programas, gerenciamento de capital humano e segurança cibernética (UNITED STATES, 2012).

Desta forma, o documento está alicerçado em resultados primários, constituído por: serviço de entrega, integração funcional, otimização de recursos e referência oficial; nível de escopo: internacional, nacional, federal, setor, agência, segmento, sistema e aplicação; elementos básicos da arquitetura corporativa federal: governança, princípios, método, ferramentas, padrões, uso, relatório e auditoria; documentação: estratégica, serviços empresariais, dados e informações, aplicativos, infraestrutura de hospedagem e segurança; planos e visualizações: roteiro empresarial, plano de transição, visualizações atuais e visualizações futuras (UNITED STATES, 2012).

No documento *Common Approach to Federal Enterprise Architecture*, a denominação utilizada é subarquitetura de dados, na qual a partir da definição das linhas e serviços de negócios, devem ser estabelecidos os fluxos de informações necessários e como harmonizá-los e padronizá-los para favorecer o compartilhamento. Além de definidas as formas de formatar, gerar, compartilhar e armazenar os dados e os recursos humanos, padrões e segurança necessários (UNITED STATES, 2012). Para tanto, o documento prevê a elaboração de dez artefatos: modelo lógico de dados (núcleo), plano de gerenciamento de conhecimento, plano de qualidade de dados, diagrama de fluxo de dados, modelo físico de dados, matriz CRUD – sigla em inglês para *create/criar*, *read/ler*, *update/atualizar* e *delete/excluir* –, diagrama de transição de estado, diagrama de sequência de eventos, dicionário de dados, biblioteca de objetos (UNITED STATES, 2012). Notadamente, para esta tese, a subarquitetura de dados é relevante.

Dos subdomínios da arquitetura são derivados Modelos de Referência, que consistem em “[...] taxonomias que fornecem categorização padronizada para informações e modelos estratégicos, de negócios e de tecnologia. Isso oferece suporte à análise e geração de relatórios entre as EAs da agência e cada um dos domínios de documentação.”²⁰⁷ (UNITED STATES, 2012, p. 31, tradução nossa).

Os Modelos de referência estabelecidos no *Common Approach to Federal Enterprise Architecture* (UNITED STATES, 2012) são: *Performance Reference Model/Modelo de Referência de Desempenho (PRM)*, *Business Reference*

²⁰⁷ “[...] taxonomies that provide standardized categorization for strategic, business, and technology models and information. This supports analysis and reporting across agency EAs and each of the documentation domains.”

Model/Modelo de Referência de Negócios (BRM), Data Reference Model/Modelo de Referência de Dados (DRM), Application Reference Model/Modelo de Referência de Aplicativo (ARM), Infrastructure Reference Model/Modelo de Referência de Infraestrutura (IRM), Security Reference Model/Modelo de Referência de Segurança (SRM).

Dentre esses modelos, o *Data Reference Model/Modelo de Referência de Dados (DRM)* contribui para esta tese no que tange aos antecedentes da Arquitetura de Dados e, por isso, é abordado.

O DRM foi projetado para fornecer uma estrutura comum flexível para o compartilhamento eficaz de informações governamentais através das fronteiras organizacionais, aumentar a integração e oportunidades de reutilização e apoiar a interoperabilidade semântica, respeitando a segurança, a privacidade e o uso apropriado dessas informações. Ele permite que as agências gerenciem informações como ativos nacionais para melhor atender ao público americano e atender às necessidades da missão.²⁰⁸ (UNITED STATES, 2012, p. 32, tradução nossa).

Dessa forma, o DRM estabelece padrões para a descrição, categorização e compartilhamento dos dados, que se refletem em: descrição de dados, contexto de dados e compartilhamento de dados. Vale destacar que o DRM “[...] é apresentado como uma estrutura abstrata a partir da qual podem ser derivadas implementações concretas.”²⁰⁹ (UNITED STATES, 2013, p. 157, tradução nossa).

Posteriormente, em 2013, foi lançado o *Federal Enterprise Architecture Framework v2 (FEAF v2)*, que descreve as ferramentas necessárias para implementar a FEA, e tem como núcleo o *Consolidated Reference Model/Modelo de Referência Consolidado (CRM)*, e o conjunto dos seis Modelos de Referência anteriormente mencionados, cuja implementação requer desenvolvimento de artefatos relacionados ao Modelo de Referência (UNITED STATES, 2013).

No FEAF v2, o DRM é descrito de maneira mais específica e, nele, está contida a estrutura do DRM:

A taxonomia do Modelo de Referência de Dados é definida por uma hierarquia em três camadas, conforme ilustrado abaixo [Figura 8]. A classificação superior da hierarquia consiste em quatro Domínios. A camada intermediária da hierarquia contém vinte e dois elementos Assunto e a classificação mais baixa da hierarquia inclui cento e

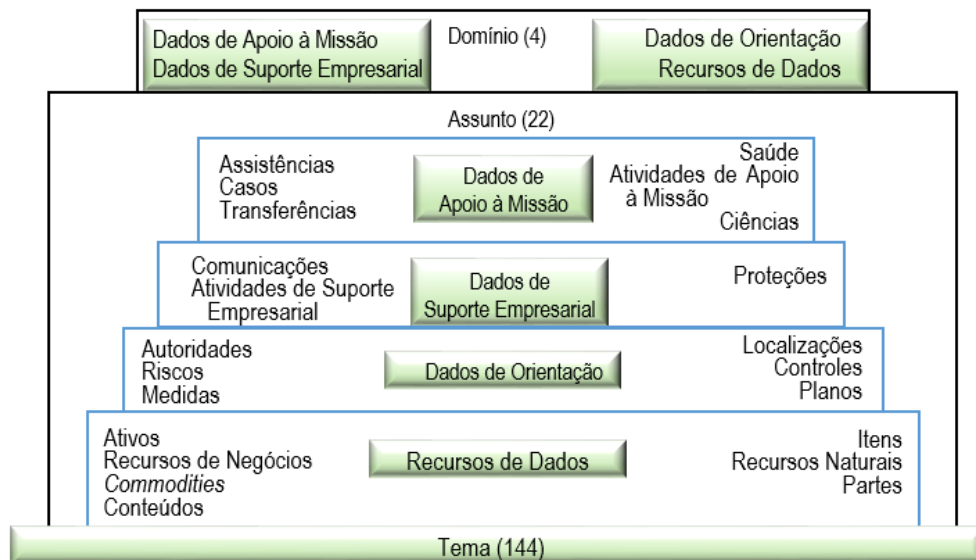
²⁰⁸ “The DRM is designed to provide a flexible common framework for effective sharing of government information across organizational boundaries, increase integration and re-use opportunities, and support semantic interoperability while respecting security, privacy, and appropriate use of that information. It enables agencies to manage information as national assets to better serve the American public and meet mission needs.”

²⁰⁹ “[...] is presented as an abstract framework from which concrete implementations may be derived.”

quarenta e quatro elementos Tema. O DRM fornece estrutura e vocabulário para que as agências formem um consenso sobre como, em nível federal, categorizar, descrever e compartilhar dados.²¹⁰ (UNITED STATES, 2013, p. 35, tradução nossa).

Figura 8 – Taxonomia DRM de alto nível

Modelo de Referência de Dados



Fonte: United States (2013, p. 35, tradução nossa).

Federal Enterprise Architecture Framework (UNITED STATES, 2013, p. 36, tradução nossa) esclarece que “[...] o elemento de tema é aberto e requer o uso de diagramas de árvore para identificar como os elementos tema na taxonomia DRM se relacionam com os elementos de assunto e domínio.”²¹¹

Conforme mencionado, o DRM se utiliza de três áreas para categorizar, descrever e compartilhar dados (Figura 9): descrição de dados, contexto de dados e compartilhamento de dados, que são explicitadas pelo FEAF (UNITED STATES, 2013):

- descrição de dados: estabelece a forma de descrição padronizada dos dados, objetivando descoberta e compartilhamento, com foco em dados estruturados e não estruturados. Para tanto, utiliza como exemplo de boas práticas: *Integration Definition for Function Modeling (IDEF)*, *The Open*

²¹⁰ “The Data Reference Model taxonomy is defined by a hierarchy in three layers, as illustrated below. The top rank of the hierarchy consists of four Domains. The middle layer of the hierarchy contains twentytwo Subject elements and the lowest rank of the hierarchy includes one hundred and forty-four Topic elements. The DRM provides a structure and vocabulary for agencies to form a consensus as to how, at a Federal level, to categorize, describe, and share data.”

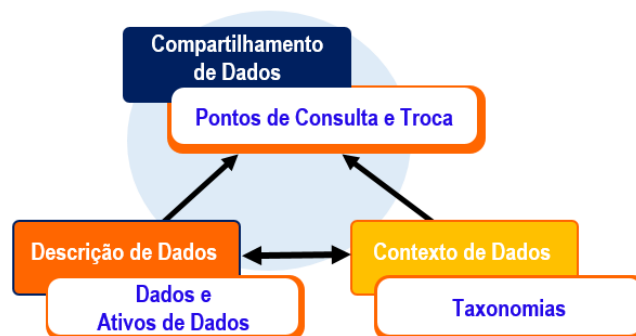
²¹¹ “[...] the topic element is open and requires use of Tree Diagrams to identify how the Topic elements in the DRM taxonomy relate to the Subject and Domain elements.”

Government Group Architecture Framework (TOGAF), Unified Modeling Language™ (UML), Department of Defense Architecture Framework v2.02 (DoDAF v2.02), ISO/IEC 11179 e Dublin Core;

- b) contexto de dados: consiste em qualquer informação que agregue significado adicional aos dados, possibilitando compreensão dos dados e de sua finalidade. Em geral, apresenta-se como um conjunto de termos organizados em listas, hierarquias e árvores. No DRM, essa taxonomia é flexível, em função de modelos de negócios e necessidades governamentais. Os exemplos citados são: catálogo de ativos de dados, e busca e descoberta de informações;
- c) compartilhamento de dados: possibilita o uso dos dados de diferentes fontes pelo(s) consumidor(es), por meio do acesso e compartilhamento dos dados, desde que habilitado pelas áreas anteriores. Como exemplo de boas práticas: *National Information Exchange Model (NIEM)*, *Data.gov*, *Linked Data* ou *Linked Open Data (LOD)* e *Information Sharing Environment Building Blocks*.

Isso posto, o documento esclarece que a taxonomia DRM possibilita que os dados sejam localizados mais facilmente, e a partir disso possam ser compartilhados. Tem-se, assim, o aumento da robustez, do potencial de localização e de compartilhamento dos dados.

Figura 9 – Áreas de padronização FEA DRM

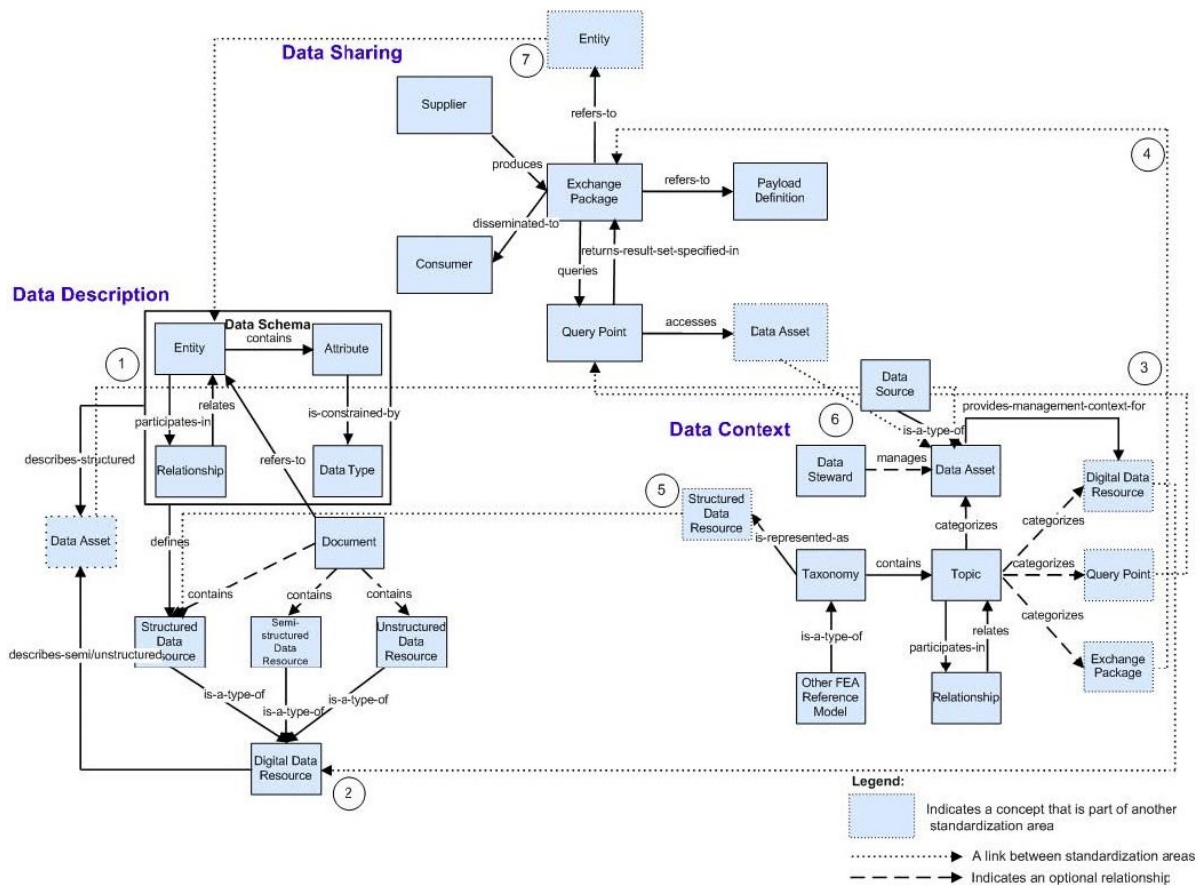


Fonte: United States (2013, p. 162, tradução nossa).

A implementação do *Data Reference Model/Modelo de Referência de Dados (DRM)* é explicitada em *Federal Enterprise Architecture Program (2005)*, no qual o modelo abstrato DRM é apresentado (Figura 10). Cumpre esclarecer que “O modelo abstrato DRM é um padrão de arquitetura para otimizar as arquiteturas de dados da

agência. É abstrato na medida em que permite múltiplas implementações técnicas [...]”²¹² (FEDERAL ENTERPRISE ARCHITECTURE PROGRAM, 2005, p.13, tradução nossa).

Figura 10 – DRM Abstract Model²¹³



Fonte: Federal Enterprise Architecture Program (2005, p. 62).

Neste sentido, o Modelo abstrato DRM possibilita que cada Agência estabeleça, considerando as especificações do referido Modelo, como será a aplicação prática no seu escopo, em função, inclusive, da infraestrutura disponível. E, antes de mais nada, aponta a necessidade de compreender o modelo a partir das três áreas de padronização mencionadas anteriormente. De acordo com Federal Enterprise Architecture Program (2005, p. 13, tradução nossa),

Na área de padronização de Descrição de Dados, o foco está em entender os dados em dois níveis de abstração: os artefatos de metadados necessários para entender os dados e como esses

²¹² “The DRM abstract model is an architectural pattern to optimize agency data architectures. It is abstract in that it allows multiple technical implementations [...]”

²¹³ Por se tratar de uma modelagem, optou-se por manter a figura no idioma original.

artefatos de metadados são agregados em um ativo de dados gerenciado. [...]

Na área de padronização do Contexto de Dados, o foco está nos mecanismos de gestão para capturar o contexto dos dados em uma organização ou [Communities of Interest] COI²¹⁴. Esses mecanismos são Taxonomias (um conjunto hierárquico de Tópicos conectados por relacionamentos) e uma descrição de Ativo de Dados (capturada em um inventário). [...]

Por último, na área de padronização de Compartilhamento de Dados, o foco está em como as informações são empacotadas e/ou expostas aos membros de uma COI. Os principais conceitos são Pacotes do Exchange como contêineres para mensagens fixas e Pontos de Consulta como descrições de pontos de acesso a dados.²¹⁵

Na Figura 10, os conceitos são expressos em caixas, e os relacionamentos, por meio das setas. É possível ainda identificar as três áreas: descrição de dados, contexto de dados e compartilhamento de dados. Todas as descrições necessárias à compreensão do *DRM Abstract Model* estão disponíveis com exemplos de aplicação em Federal Enterprise Architecture Program (2005).

A sistematização apresentada por meio do DRM possibilita vislumbrar a Arquitetura de Dados para além dos bancos de dados e sistema de informação, possibilitando uma aproximação com o tratamento dos dados, considerando a relevância da gestão de dados como ativos organizacionais.

No domínio das Ciências Agrícolas, Harfouche, Nakhle, Harfouche *et al.* (2022) propõem uma arquitetura de sistema de Inteligência Artificial (IA) centrada em humanos, constituída por arquitetura de dados, infraestrutura de tecnologia e design de arquitetura de IA. Os autores não explicitam como se constitui a arquitetura de dados. Contudo, afirmam que uma arquitetura de dados bem projetada possibilita que os conjuntos de dados tenham alta qualidade, o que beneficia pesquisadores em análises orientadas por dados. Ainda, entendem que um dos desafios é armazenar os dados atendendo aos Princípios FAIR, o que reforça a necessidade de relacionar tais princípios à Arquitetura de Dados. No que se refere aos Princípios FAIR, afirmam:

²¹⁴ *Communities of Interest (COI)*: “As Comunidades de Interesse são grupos colaborativos de usuários que requerem um vocabulário compartilhado para trocar informações em busca de metas, interesses e objetivos de negócios comuns.” (FEDERAL ENTERPRISE ARCHITECTURE PROGRAM, 2005, p. 1, tradução nossa).

²¹⁵ “In the Data Description standardization area, the focus is on understanding the data at two levels of abstraction: the metadata artifacts required to understand the data and how those metadata artifacts are aggregated into a managed Data Asset.” [...] “In the Data Context standardization area, the focus is on management mechanisms to capture the context of data in an organization or COI. Those mechanisms are Taxonomies (a hierarchical set of Topics connected by relationships) and a Data Asset description (captured in an inventory).” [...] “Lastly, in the Data Sharing standardization area, the focus is on how information is packaged for and/or exposed to members of a COI. The key concepts are Exchange Packages as containers for fixed messages and Query Points as descriptions of data access points.”

A arquitetura de dados desempenha um papel fundamental no atendimento a esses requisitos. Consiste em um conjunto de padrões que regem quais dados são coletados, se devem ser transformados (por exemplo, limpeza de dados, deduplicação, conversão de formato, estruturação, validação etc.) antes ou depois do armazenamento usando processos extrair, transformar, carregar [*Extract, Transform, Load*] (ETL) ou extrair, carregar, transformar [*Extract, Load, Transform*] (ELT) e onde (*data warehouses* ou *data lakes*) e como (matrizes, cubos, politopos ou distribuídos na memória) os dados são armazenados [...].²¹⁶ (HARFOUCHE; NAKHLE; HARFOUCHE *et al.*, 2022, p. 18, tradução nossa).

Além disso, Harfouche, Nakhle, Harfouche *et al.* (2022) indicam que, no contexto estudado, sem o uso de IA os fluxos de dados seriam caóticos, ao passo que, para o uso de IA, reconhecem haver a necessidade de infraestrutura tecnológica, no que se refere a *hardware*, *software*, rede e linguagens de programação. Tal afirmação indica e reforça o caráter multidisciplinar da Arquitetura de Dados e sua dependência de aparatos tecnológicos, sendo a incorporação da IA uma relevante proposta.

O estudo de Laun, Mazzuchi e Sarkani (2022) apresenta um modelo de dados conceitual para caracterização da resiliência em nível de sistema. Os autores afirmam que os dados coletados raramente são estruturados e/ou possuem facilidade de processamento, o que pode caracterizar o interesse atual pelos *Data Lakes*, que consistem em um ambiente que armazena dados brutos, estruturados ou não, com a finalidade de processamento, mineração e análise.

Essa afirmação reforça a necessidade de atuar na estruturação dos dados, por meio da Arquitetura de Dados, visando aprimorá-los, inclusive ao torná-los FAIR.

Dow, Hackbarth e Wong (2013) discutem o desenvolvimento de um *framework* utilizado para o projeto de um sistema de informação para a Gestão do Conhecimento Organizacional. Nele, os autores atrelam a Arquitetura de Dados à estrutura de armazenamento em banco de dados, de forma que a abordagem à AD é sensivelmente diferente da que apresentamos neste estudo. Contudo, apresenta aspectos relevantes relacionados ao armazenamento, considerando maior aproximação com o domínio da Ciência da Computação.

²¹⁶ “Data architecture plays a fundamental role in meeting these requirements. It consists of a set of standards that govern which data are collected, whether they should be transformed (e.g., data cleaning, deduplication, format conversion, structuring, validation, etc.) before or after storage using extract, transform, load (ETL) or extract, load, transform (ELT) processes, and where (data warehouses or data lakes) and how (matrices, cubes, polytopes, or distributed in-memory) the data are stored [...]”

O estudo de Ingwersen, Hawkins e Transue *et al.* (2015) discute as dificuldades de interoperabilidade e integração de dados e aponta que há casos em que os dados estão indisponíveis publicamente, e outros em que os dados estão em um formato que dificulta a compreensão e o uso por ferramentas computacionais, o que pode impedir que os dados sejam utilizados. Os autores apontam ainda que há ferramentas computacionais que possibilitam a conversão de registros, mas que as divergências na estrutura tendem a tornar a migração um processo moroso.

Desta forma, propõe o uso de *Resource Description Framework* (RDF) para dar suporte à interoperabilidade dos registros, tornando possível que dados de diversas fontes sejam relacionáveis sem a estrutura de um banco de dados central (INGWERSEN; HAWKINS; TRANSUE *et al.*, 2015).

Otto (2012) apresenta um estudo que aborda o projeto da arquitetura de dados mestre, por meio de um estudo de caso. O autor elucida que pouco se tem discutido a respeito de forma científica. Assim, define o termo como “Dados mestre são dados sobre as características dos principais objetos de negócios em uma empresa e são definidos de forma inequívoca, bem como identificados exclusivamente em toda a organização.”²¹⁷ (OTTO, 2012, p. 338, tradução nossa). E esclarece que dados mestre são aqueles que identificam e descrevem os objetos centrais da organização, tendo como exemplos dados de clientes, de fornecedores, de produtos e de materiais.

Desta forma, a Arquitetura de Dados Mestre envolve a arquitetura conceitual de dados mestre e a arquitetura das aplicações utilizadas para os dados mestre, constituindo-se em uma abordagem conceitual e outra física.

O caso estudado por Otto (2012) possibilitou alavancar aspectos conceituais, além de confirmar fatos do mundo real e, com isso, leva a discussão adiante em relação a dois aspectos principais: a questão do projeto arquitetônico que não deve ser reduzido apenas aos aspectos técnicos, mas deve equilibrar as considerações organizacionais e técnicas; e, a complexidade, sobretudo no que tange à quantidade e heterogeneidade das classes de dados mestre.

Tais estudos apontam que investigações têm sido realizadas em diferentes domínios acerca da Arquitetura de Dados, com as mais diversas abordagens, tendo em vista as características da área e sua aplicação no domínio. Nesse sentido, evidencia-se a escassez de estudos que tratem da Arquitetura de Dados no domínio

²¹⁷ "Master data is data about the characteristics of key business objects in a company and is unambiguously defined as well as uniquely identified across the organization."

da Ciência da Informação, o que amplia a relevância do objetivo geral proposto nesta tese.

4.3 ARQUITETURA DE DADOS: APORTE PARA A CONSTRUÇÃO DO CONCEITO

Considerando que o objetivo geral desta tese que consiste em investigar e compreender a Arquitetura de Dados no contexto da Ciência da Informação, a fim de discuti-la, conceituá-la e aproximá-la da Arquitetura da Informação, revisitar a literatura científica clássica foi preponderante. De igual maneira, foi imprescindível compreender e estruturar o polo teórico de forma a sustentar os elementos necessários à elaboração de um conceito, o que se apresenta, nesta seção, com o estudo da Teoria do Conceito de Ingetraut Dahlberg (1978a, 1978b).

Do ponto de vista metodológico, cuja estruturação se faz com o delineamento quadripolar, essa construção ocorre por meio da sustentação teórica própria do polo teórico do estudo e, de igual maneira, desdobra-se como atividade do polo técnico na medida que recorre aos procedimentos técnicos executados a partir da teoria do conceito, e culmina na elaboração de um conceito, apresentado na seção 4.4 que se constitui em um elemento do polo morfológico da tese.

A teoria do conceito apresentada por Dalberg (1978a, p. 142, tradução nossa) considera que

[...] todo conceito tem um referente (seja este um conjunto de objetos, um único objeto, uma atividade, um fato, um tópico, etc.) sobre o qual podem ser feitas afirmações verificáveis que determinam as propriedades e relações do referente em questão. A totalidade de todas as afirmações verificáveis e necessárias sobre um referente pode ser resumida e/ou sintetizada por um termo que representará então um conceito em qualquer processo de comunicação. Um conceito é, assim, considerado como uma unidade de conhecimento, e as afirmações sobre seu referente são consideradas os elementos de conhecimento, também conhecidos como características, do conceito dado.²¹⁸

Para Dalberg (1978b, p. 102),

Podemos agora definir a formação dos conceitos como a reunião e compilação de enunciados verdadeiros a respeito de determinado

²¹⁸ “[...] every concept has a referent (be this a set of objects, a single object, an activity, a fact, a topic, etc.) about which verifiable statements determining the properties and relationships of the referent in question can be made. The totality of all the verifiable and necessary statements on a referent may be summarized and/or synthesized by a term which will then represent a concept in any communication process. A concept is thus regarded as a knowledge unit, and the statements about its referent are found to be the knowledge elements, also known as the characteristics, of the given concept.”

objeto. Para fixar o resultado dessa compilação necessitamos de um instrumento. Este é constituído pela palavra ou por qualquer signo que possa traduzir e fixar essa compilação. É possível definir, então, o conceito como a compilação de enunciados verdadeiros sobre determinado objeto, fixada por um símbolo linguístico. Esse símbolo pode ser verbal ou não-verbal, ou seja, pode ser formado de sinais ou conjunto de sinais independentes das palavras.

Nesta tese, objetiva-se conceituar a Arquitetura de Dados no contexto da Ciência da Informação, o que se fez por meio de signos linguísticos. Para tanto, faz-se necessário compreender o objeto, bem como suas propriedades e relações, o que torna possível ainda identificar como a Arquitetura de Dados se relaciona com a Arquitetura da Informação.

No que tange ao conceito, para Dalhberg (1978b, p. 102)

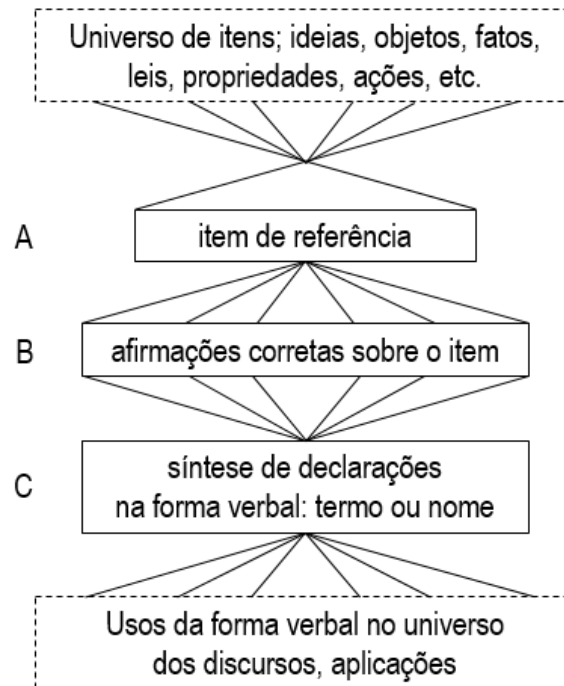
[...] todo enunciado sobre objetos contém um elemento do respectivo conceito. Estes elementos identificam-se com as chamadas características dos conceitos. Traduzem os atributos das coisas designadas. Mais uma vez convém repetir que é formulando enunciados sobre os atributos necessários ou possíveis dos objetos que se obtém as características dos respectivos conceitos.

Vale lembrar que Dahlberg (1978a) considera conceito como uma unidade de conhecimento. Desta forma, sua construção pressupõe que haja um universo de itens, dos quais um item é reconhecido como referência, sendo designado como referente (A), que pode se constituir, por exemplo, em um objeto, propriedade, ação e/ou dimensão, ou ainda um fenômeno. Sobre o referente são elencadas afirmações (B) que, quando verificadas, são consideradas aceitas e expressas em declarações verbais, como um termo ou nome (C), a partir do qual ocorre a comunicação verbal e escrita, que consiste na aplicação do conceito em declarações e discursos (DAHLBERG, 1978a). A construção de um conceito, a partir da teoria de Dahlberg (1978a) é representada por um modelo (Figura 11).

Para Dalhberg (1978a, p. 143, tradução nossa)

Este modelo de construção de conceitos [Figura 11] demonstra como pode ser gerado um conceito que ainda não existe como tal. Os componentes necessários de tal conceito são, então, seu referente (A), julgamentos sobre o referente (B) e a forma verbal (C) utilizada para representá-lo.²¹⁹

²¹⁹ "This model for concept construction demonstrates how a concept may be generated which does not yet exist as such. The necessary components of such a concept are, then, its referent (A), judgments about the referent (B) and the verbal form (C) used to represent it."

Figura 11 – Modelo para construção de conceito

Fonte: Dalhberg (1978a, p. 143, tradução nossa).

Desta forma, Dalhberg (1978a, p. 143, tradução nossa) afirma que “Um conceito é uma unidade de conhecimento, compreendendo declarações verificáveis sobre um item de referência selecionado, representado de forma verbal.”²²⁰, cujos componentes são igualmente definidos. Assim,

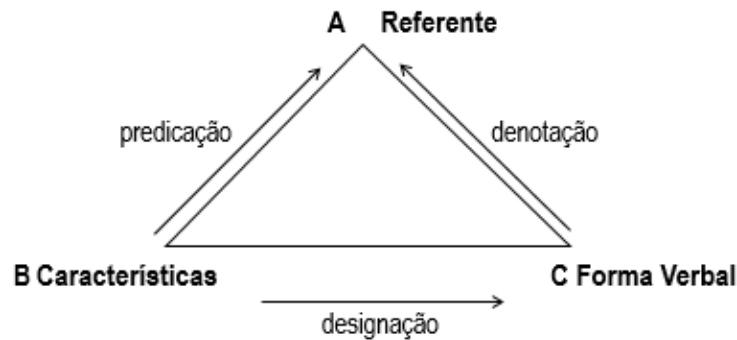
[...] Uma declaração verificável é o componente de um conceito que declara um atributo de seu item de referência. [...] Um item de referência é o componente de um conceito ao qual suas declarações verificáveis e sua forma verbal estão diretamente relacionadas, portanto, seu "referente". [...] Uma forma verbal (termo/nome) de um conceito é o componente que convenientemente resume ou sintetiza e representa um conceito com o propósito de designar um conceito na comunicação.²²¹ (DALHBERG, 1978a, p. 144, tradução nossa).

A partir dessas afirmações, a autora apresenta uma representação triangular (Figura 12), na qual a fonte de criação do conceito está no topo (A - referente), as características/significado (B) à esquerda e a forma verbal (C) à direita.

²²⁰ “A concept is a knowledge unit, comprising verifiable statements about a selected item of reference, represented in a verbal form.”

²²¹ “[...] A verifiable statement is the component of a concept which states an attribute of its item of reference. [...] An item of reference is the component of a concept to which its verifiable statements and its verbal form are directly related, thus its "referent". [...] A verbal form (term/name) of a concept is the component which conveniently summarizes or synthesizes and represents a concept for the purpose of designating a concept in communication.”

Figura 12 – O triângulo conceitual



Fonte: Dalhberg (1978a, p. 144, tradução nossa).

Para Dalhberg (1978a, p. 144, tradução nossa), a representação triangular evidencia “[...] que tipo de atividades/relações existem entre os componentes mencionados, nomeadamente predicação (B, A), designação (B, C) e denotação (C, A).”²²². Vale destacar que predicação, designação e denotação são componentes intermediários e relacionais entre as arestas do referido triângulo conceitual.

No que tange às características constituintes do conceito, de acordo com Dalhberg (1978a), podem ser divididas em características essenciais, acidentais e individualizadoras, sendo, respectivamente, aplicáveis a todos os referentes do mesmo tipo, a alguns referentes do mesmo tipo e a um único referente.

Assim, as características essenciais são compartilhadas por todos os referentes do mesmo tipo, ao passo que as acidentais estão presentes apenas em alguns desses referentes e as individualizadoras são únicas e, com isso, individualizam o referente, por se aplicarem apenas a ele.

Com base no exposto, e considerando as discussões realizadas nas seções anteriores, na seção 4.4 é apresentado o conceito de Arquitetura de Dados no contexto da Ciência da Informação.

²²² “[...] what kind of activities/relations exist between the components mentioned, namely predication (B, A), designation (B, C), and denotation (C, A).”

4.4 CONCEITO DE ARQUITETURA DE DADOS NO CONTEXTO DA CIÊNCIA DA INFORMAÇÃO

Nesta seção é realizada uma construção conceitual acerca da Arquitetura de Dados no contexto da Ciência da Informação, como um elemento do polo morfológico do delineamento quadripolar.

A Arquitetura de Dados é um campo já consolidado na Ciência da Computação, Gestão Empresarial e Sistemas de Informação e que, empiricamente, entendemos próximos do domínio da Ciência da Informação, cujo objeto de estudo de acordo com Capurro (2003) é a informação.

Contudo, diante das discussões desta tese, sobretudo na seção 4.1, entende-se que a informação é um fenômeno essencialmente humano (SILVA, 2017), cuja compreensão por máquinas se faz por meio da materialização em dados, que podem ser assim acessados e processados por humanos e aplicações ou agentes computacionais.

Desta forma, é possível compreender que os dados, entendidos como a materialização da informação visando o registro de fatos e/ou a representação de atributos de entidades e seus contextos, aproxima-se da chamada “informação registrada”, ou “informação como coisa” (BUCKLAND, 1991), mas que, visando tratar ambos os conceitos por suas características e, em busca de desemaranhar o entendimento designa-se nesta tese como dados, cujo conceito operacional é apresentado na seção 4.1.

Faz-se assim uma aproximação dos dados como um fenômeno que é também objeto de estudo da Ciência da Informação, reafirmando a compreensão de Capurro (2003) de que o objeto de estudo da CI, a informação, é indissociável dos dados, o que faz da Ciência da Informação uma área fronteira ao estudo dos dados.

Assim, como o domínio da Ciência da Informação, a Arquitetura da Informação enquanto disciplina necessita de diferentes domínios e *expertises* para o desenvolvimento científico e aplicado, conforme discutido na seção 3 desta tese. Desta forma, é possível afirmar que diferentes domínios, campos e disciplinas, considerados inter e/ou multidisciplinares, buscam aparatos teóricos de outros domínios, campos e disciplinas considerados “afins” para alicerçar o desenvolvimento.

Os aportes de Arquitetura de Dados e seus elementos constitutivos discutidos nas seções 4, 4.2 e 4.3 possibilitam compreender a Arquitetura de Dados (AD) no

contexto da Ciência da Informação e estabelecer os seus elementos conceituais essenciais, objetivo específico (c) do presente estudo, e, imprescindível para atender ao objetivo geral, no que tange a “investigar e compreender a Arquitetura de Dados a fim de conceituá-la no contexto da Ciência da Informação e relacioná-la com a Arquitetura da Informação.

Para Dahlberg, (1978b, p. 102) o “[...] o conceito é constituído de elementos que se articulam numa unidade estruturada.”, e a autora define “[...] conceito como a compilação de enunciados verdadeiros sobre determinado objeto, fixada por um símbolo linguístico.”

Após a aplicação da Teoria do Conceito de Dahlberg (1978a, 1978b), compreende-se a Arquitetura de Dados como um conceito individual, do qual decorrem os seguintes enunciados:

- a) tem como objeto os dados;
- b) atua na estruturação dos dados;
- c) necessita de estruturas sintática e semântica;
- d) necessita de esquemas e padrões de metadados;
- e) se relaciona com ciclo(s) de vida dos dados, que faz(em) parte de um projeto e cujas fases são geridas por um detentor;
- f) necessita de requisitos para localização, coleta, tratamento, armazenamento, recuperação e descarte;
- g) necessita de processamento para localização, coleta, tratamento, armazenamento, recuperação e descarte;
- h) considera princípios éticos, legais e técnicos;
- i) utiliza protocolos de comunicação para o consumo e fornecimento de dados.

Isso posto, com base na Teoria do Conceito (DALHBERG, 1978a, p. 145), é possível compreender que “[...] todos os nossos conceitos são abstrações da realidade no sentido de que são produtos e instrumentos da capacidade do homem de pensar e falar sobre a realidade na medida permitida por seu conhecimento da realidade.”

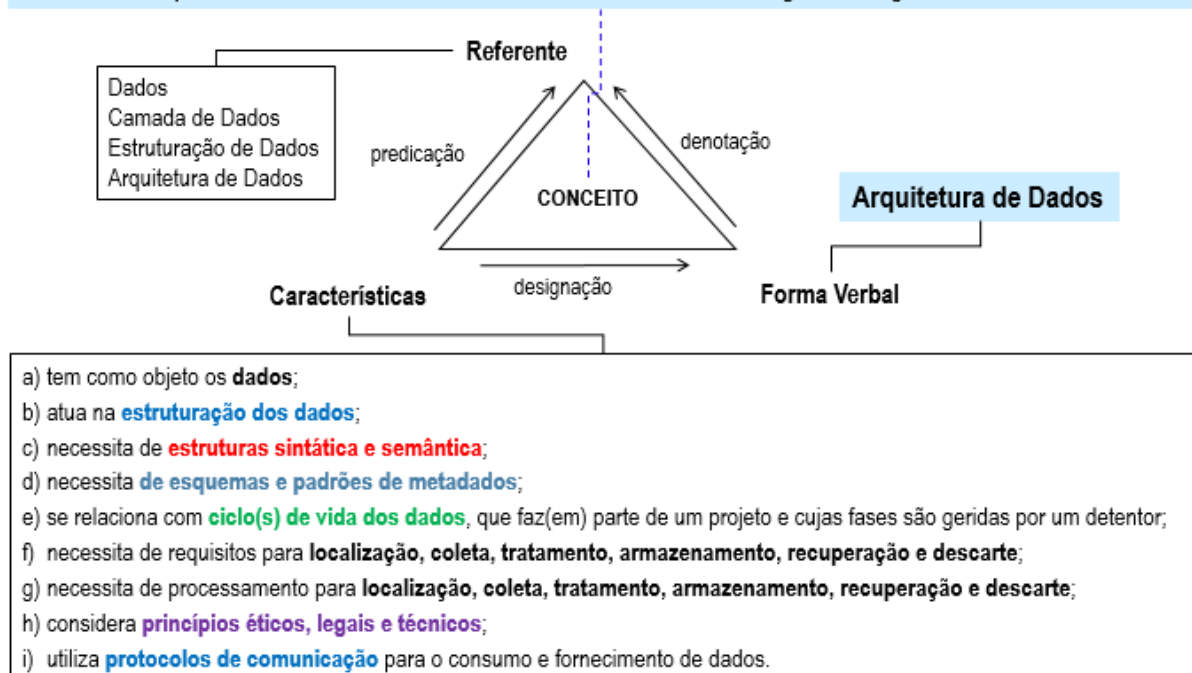
Dessa forma, a compilação dos enunciados culmina no conceito de Arquitetura de Dados no contexto da Ciência da Informação, compreendido como:

A Arquitetura de Dados atua no projeto da infraestrutura de dados e do ciclo de vida dos dados, de forma holística e coesa, considerando que os dados necessitam de padrões de metadados, estruturas sintática e semântica e atendam a princípios éticos, legais e técnicos para que sejam consumidos e fornecidos para diferentes finalidades utilizando técnicas e tecnologias emergentes.

O processo de construção do conceito de Arquitetura de Dados no contexto da Ciência da Informação, utilizando-se a Teoria do Conceito (DALHERG, 1978a, 1978b), é ilustrado na Figura 13.

Figura 13 – Construção do Conceito de Arquitetura de Dados no contexto da Ciência da Informação

A **Arquitetura de Dados** atua no projeto da **infraestrutura de dados** e do **ciclo de vida dos dados**, de forma holística e coesa, considerando que os dados necessitam de padrões de **metadados**, **estruturas sintática e semântica** e atendam a **princípios éticos, legais e técnicos** para que sejam **consumidos e fornecidos** para diferentes finalidades utilizando técnicas e tecnologias emergentes.



Fonte: Autoria própria.

No contexto da Arquitetura de Dados, é importante ressaltar que os dados são projetados para serem coletados/consumidos por quaisquer ambientes informacionais digitais e/ou sistemas de informação que necessitem e possuam os requisitos necessários para consumo.

Essa afirmação encontra base em Tupper (2011), que entende que os dados normalizados maximizam a possibilidade de compartilhamento, sobretudo por existirem independentemente de seu uso e por não serem orientados a aplicações.

Nesse sentido, a presente seção se constitui por seções terciárias visando discutir: Elementos da Arquitetura de Dados no contexto da Ciência da Informação – composta por: Ciclo de Vida dos Dados; Metadados; Estruturas Sintática e Semântica dos Dados; Princípios Éticos, Legais e Técnicos aplicados aos Dados –, Arquitetura de Dados no contexto da Ciência da Informação: Tríade e Fluxo; Entregáveis da Arquitetura de Dados e Arquiteto de Dados, no contexto da Ciência da Informação, entendendo-os como relacionados ao conceito de Arquitetura de Dados.

4.4.1 Elementos da Arquitetura de Dados no contexto da Ciência da Informação

Após a elaboração do conceito de Arquitetura de Dados, torna-se necessário discutir cada um dos elementos estruturantes do referido conceito no contexto da Ciência da Informação, visando apresentar as definições teóricas adequadas para dar sustentação ao conceito criado. Assim, as próximas subseções dialogam com o polo teórico da presente tese.

Nesta seção, são abordados os elementos que compõem o conceito de Arquitetura de Dados, visando contribuir com a compreensão do conceito. Vale destacar que não há pretensão de esgotar a análise, possibilitando que outras discussões sejam feitas, de forma complementar.

Cumprir destacar que a Arquitetura de Dados no contexto da Ciência da Informação requer o emprego de técnicas ou tecnologias emergentes, visando apontar caminhos para que os elementos possam ser considerados sem estarem atrelados a técnicas ou tecnologias específicas, considerando que isso limitaria e tornaria o desenvolvimento do projeto temporal e dependente.

4.4.1.1 *Ciclo de Vida dos Dados*

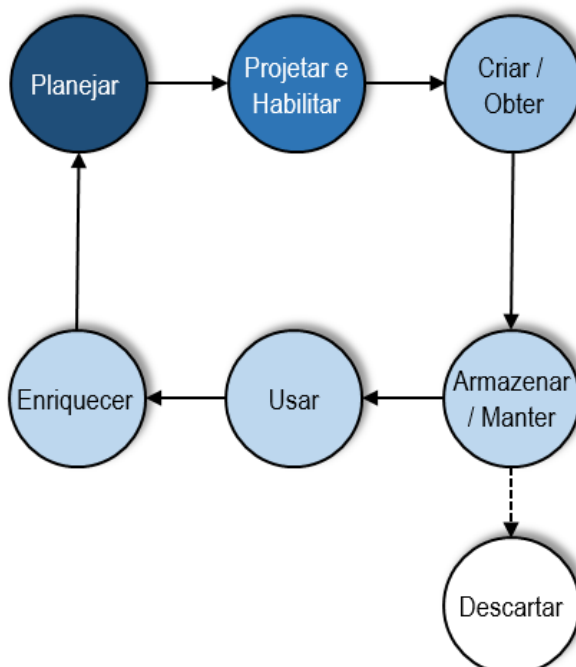
Para a abordagem do processo de acesso e uso de dados é relevante compreender como ocorrem as fases de planejamento, coleta e disponibilização dos dados, que envolvem as formas com que são criados ou coletados, tratados,

armazenados, disponibilizados para consulta e, quando necessário, descartados, etapas essas que compõem o ciclo de vida dos dados.

De acordo com DAMA-DMBOK (2017), os dados, como um ativo organizacional, possuem um ciclo de vida, que precisa ser compreendido e gerenciado, de acordo com as especificidades do ciclo de vida. Além disso, considera que os dados possuem ainda uma linhagem ou cadeia de dados, que consiste na sua trajetória desde a origem até o uso, tendo em conta as transformações que podem ocorrer.

Neste sentido, as principais atividades do ciclo de vida dos dados (Figura 14), de acordo com DAMA-DMBOK (2017) consistem em planejar, projetar e habilitar, criar/obter, armazenar/manter, usar, enriquecer e descartar. Nesse processo, os dados podem ser tratados, transformados e aprimorados. Assim, à medida que os dados são utilizados, novos dados podem ser criados. Desta forma, a gestão do ciclo de vida dos dados deve incluir aspectos de criação e uso, qualidade dos dados e metadados, segurança e análise de riscos.

Figura 14 – Principais atividades do Ciclo de Vida dos Dados



Fonte: DAMA-DMBOK (2017, p. 29).

Na perspectiva de Levitin e Redman (1993), o ciclo de vida dos dados (CVD) auxilia na descrição das atividades relacionadas aos dados, desde a criação até o descarte, e seus efeitos na qualidade dos dados. Para os autores, é possível

classificar o processo, atentando-se para o armazenamento dos dados, sendo do tipo “aquisição de dados”, quando o armazenamento dos dados é o objetivo do processo, do tipo “uso” quando os dados armazenados são acessados e processados visando obter um resultado e, do tipo “combinado”, quando ambos os objetivos, aquisição e uso, são concomitantes.

Assim, de acordo com Levitin e Redman (1993), o ciclo de aquisição (criação dos dados) é composto por quatro etapas: definição de visualização, na qual são especificadas as entidades, atributos, domínio e valores admissíveis; implementação da visualização, etapa na qual as definições realizadas anteriormente devem ser implementadas, atentando-se às características do sistema de informação e/ou ambiente informacional no que tange ao armazenamento, podendo gerar a necessidade de normalizar e documentar os dados e as linguagens utilizadas; obtenção de valores, que se utilizará de técnicas para adquirir valores específicos para cada atributo relacionado a entidade definida e cuja coleta será orientada por um conjunto de regras; e atualização de registros, que consiste na inclusão, modificação ou descarte dos registros armazenados.

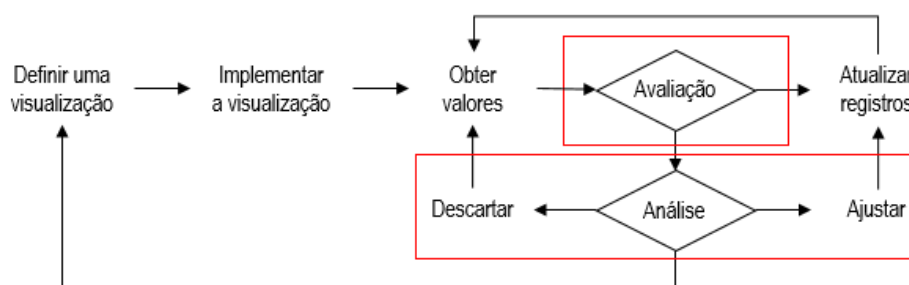
Por outro lado, para Levitin e Redman (1993) o ciclo do uso (consumo de dados) é composto por cinco etapas: definição de subvisualização, considerando que, em geral, há a necessidade de um conjunto de dados menor que a quantidade disponível, a primeira etapa consiste na definição do subconjunto; recuperação, consiste na coleta dos dados para armazenamento e uso futuro; manipulação, entendida como o processamento dos dados, que pode ser realizado por classificação, análise, manipulação e síntese; apresentação dos resultados aos usuários, que deve considerar a natureza dos resultados, o suporte de armazenamento e o usuário (humano, aplicação e/ou aplicativo computacional); uso, etapa na qual os resultados apresentados são utilizados, o que determina a qualidade dos dados, entendida como adequação ao uso.

Ainda, na visão de Levitin e Redman (1993), é necessário que haja, no ciclo de vida dos dados, uma etapa de feedback e outra para o descarte dos dados, sendo elas incluídas aos ciclos de aquisição e uso e, em outros dois modelos, considerados pelos autores como aprimorados ou enriquecidos.

Desta forma, o ciclo de aquisição, composto por definir uma visualização, implementar a visualização, obter valores e atualizar registros, quando enriquecido, incorpora outras quatro etapas (Figura 15): avaliação, na qual os dados obtidos são

avaliados, sobretudo no que tange à precisão, atualidade, integridade e consistência, e que define se os dados obtidos serão armazenados ou necessitarão de tratamento; análise, realizada quando a qualidade dos dados não é satisfatória, o que pode levar à necessidade de ajustar ou descartar e nova aquisição de dados. Considerando a relevância da qualidade dos dados, é importante que esta etapa seja documentada, para favorecer obtenções futuras; ajuste, os dados cuja qualidade for insatisfatória poderão ser tratados, seja pela correção ou inclusão de valores; e o descarte, que deve ser realizado quando os dados coletados forem insatisfatórios e não puderem ser ajustados.

Figura 15 – Um modelo do ciclo de aquisição enriquecido



Fonte: Levitin e Redman (1993, p. 220, tradução nossa).

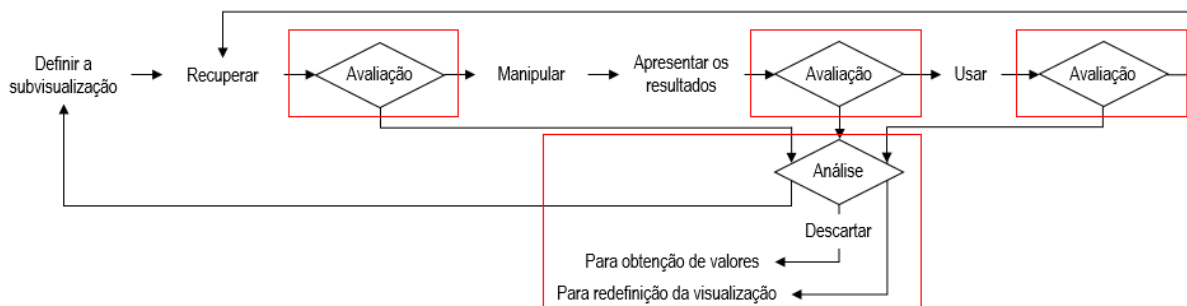
Nota: As etapas incluídas no enriquecimento do ciclo de aquisição estão destacadas em vermelho.

De igual maneira, o ciclo de uso proposto por Levitin e Redman (1993), composto por definir subvisualização, recuperar, apresentar os resultados e usar, quando enriquecido incorpora outras três etapas, todas designadas avaliação (Figura 16): cada uma das etapas de avaliação possui uma finalidade: a primeira avalia se os dados recuperados serão manipulados ou necessitam de análise visando o descarte ou uma nova obtenção de dados; a segunda avalia os dados considerando a manipulação e a apresentação, possibilitando, da mesma forma que a primeira avaliação, a análise que culmina no descarte ou em uma nova obtenção de dados; e a terceira avalia os dados após seu uso, gerando, de igual maneira, a possibilidade de análise e definição de subvisualização ou ainda de recuperação.

Os modelos apresentados por Levitin e Redman (1993) consideram duas fases relevantes do ciclo de vida dos dados, a aquisição e o uso, tratando-as de forma separadas, por considerar que os dados gerados por uma organização podem ser utilizados por várias outras. Os autores apontam ainda que as diferentes etapas

elencadas em seus modelos contribuem para estabelecer em que ponto a qualidade dos dados está insatisfatória, o que possibilita a melhoria.

Figura 16 – Modelo do ciclo de uso enriquecido

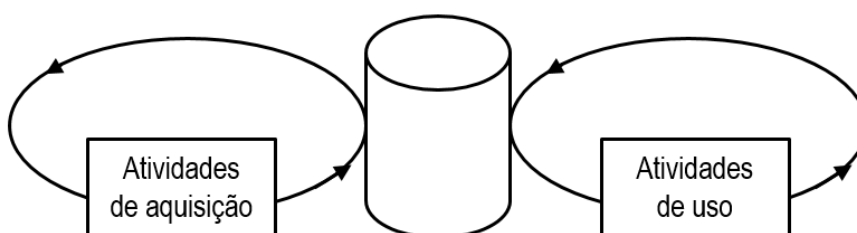


Fonte: Levitin e Redman (1993, p. 220, tradução nossa).

Nota: As etapas incluídas no enriquecimento do ciclo de uso estão destacadas em vermelho.

Assim, os ciclos de aquisição e uso são combinados (Figura 17) de forma a ilustrar que os processos relacionados aos dados podem ocorrer e se repetem em diferentes etapas dos referidos ciclos. Ainda, Levitin e Redman (1993) afirmam que os aspectos relacionados à qualidade dos dados devem ser tratados, sempre que possível, antes do armazenamento, para que não haja a necessidade de fazê-lo na base de dados.

Figura 17 – O ciclo combinado de aquisição e uso



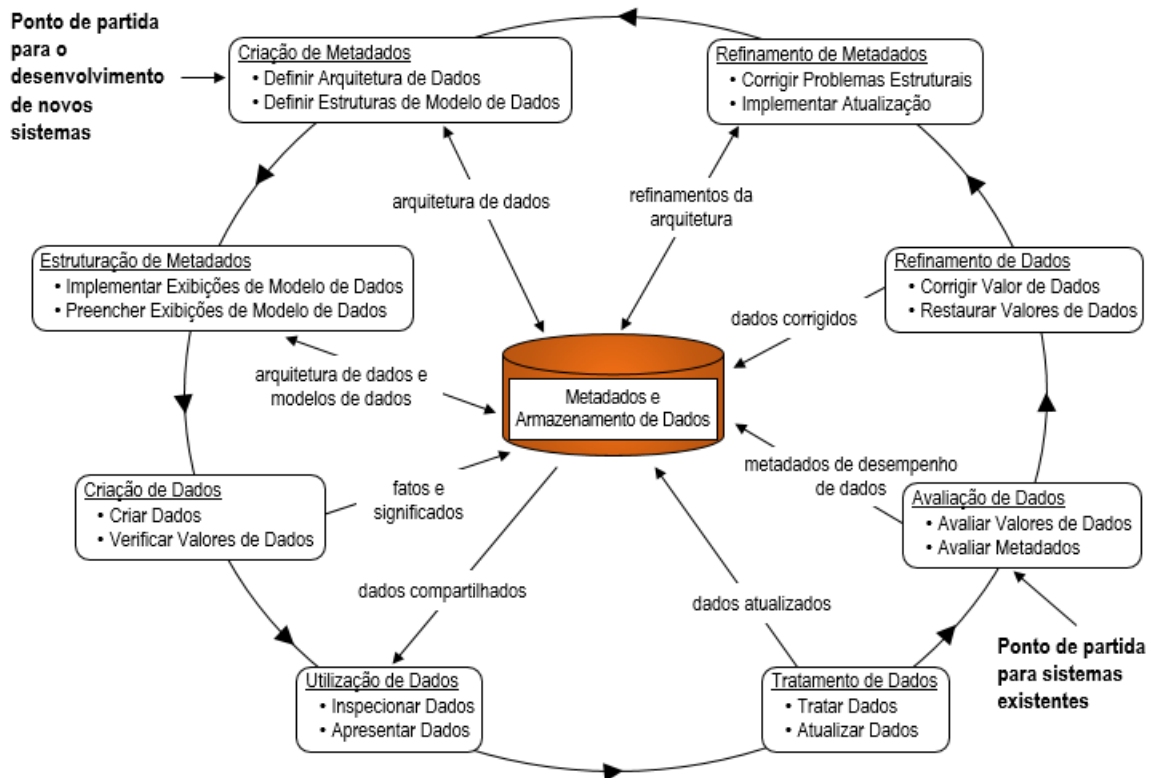
Fonte: Levitin e Redman (1993, p. 218, tradução nossa).

Na perspectiva dos modelos apresentados por Levitin e Redman (1993), podemos considerar que, no que tange à Arquitetura de Dados, conforme proposta nesta tese, os dois processos, de aquisição e uso dos dados, podem ser otimizados pelo uso de estruturas padronizadas de dados e metadados, que fornecem a semântica necessária à compreensão dos dados por máquinas, quer seja pelo uso de aplicações ou agentes computacionais, tornando as etapas de avaliação, análise e tratamento mais precisas. Além disso, diretrizes como as estabelecidas pelos

princípios FAIR, que também utiliza padrões de metadados, contribuem para que o processo de localização, acesso, interoperabilidade e uso dos dados seja favorecido.

Yoon, Aiken e Guimaraes (2005) propõem um ciclo de vida dos dados (Figura 18) visando estabelecer relações entre as diferentes fases do ciclo de vida, considerando a qualidade de dados, de forma que as organizações possam gerenciar dados com uso em sistemas individuais e integrados.

Figura 18 – Oito fases do modelo de Ciclo de Vida dos Dados



Fonte: Yoon, Aiken e Guimaraes (2005, p. 147, tradução nossa).

Nesse modelo de ciclo de vida dos dados (Figura 18), proposto por Yoon, Aiken e Guimaraes (2005), são apresentadas oito fases que contemplam fontes e usos de metadados: criação de metadados, estruturação de metadados, criação de dados, utilização de dados, manipulação de dados, avaliação de dados e refinamento de dados e refinamento de metadados. Além disso, há duas possibilidades de aplicação, sendo uma para o desenvolvimento de um novo sistema, assinalada como “Ponto de partida para o desenvolvimento novos sistemas” e outra para os sistemas já existentes, cujos dados são designados pelos autores como legado, chamado de “Ponto de partida para sistemas existentes”.

No modelo de Yoon, Aiken e Guimaraes (2005) (Figura 18) seguindo a necessidade – Ponto de partida para o desenvolvimento de novos sistemas ou Ponto de partida para sistemas existentes – as fases devem ser seguidas em sentido anti-horário. Assim, quando o ciclo é utilizado na criação de um novo sistema, ele é iniciado pela criação de metadados, ao passo que, quando o sistema, e por consequência os dados, já existem, o ciclo se inicia pela avaliação dos dados. Nesse modelo, independentemente do ponto de partida, há passagem pela Arquitetura de Dados, que no modelo está relacionada aos metadados, presente na criação, estruturação e refinamento de metadados.

No que tange aos sistemas já existentes, Yoon, Aiken e Guimaraes (2005) afirmam haver sempre uma Arquitetura de Dados, que pode não estar formalmente especificada, o que gera dificuldades no gerenciamento. Assim, reforçam a relevância da definição de requisitos, definições e estruturas relacionadas à Arquitetura de Dados. Com isso, a etapa de criação de metadados presente no ciclo de vida está diretamente relacionada à definição da Arquitetura de Dados e das estruturas de modelo de dados, seguidos da estruturação de metadados, na qual serão implementadas visualizações do modelo de dados e haverá o povoamento do referido modelo. Essas duas etapas se relacionam, uma vez que a criação dos metadados é necessária para que os modelos de dados relacionados aos diferentes sistemas de informação sejam povoados.

Ainda relacionado a essas etapas, está o refinamento de metadados, que auxilia na correção da estrutura proposta, tornando-a mais adequada. Embora haja relacionamento direto com a criação e estruturação dos metadados, quando de um novo sistema de informação, o refinamento de metadados é possível apenas quando os dados forem avaliados. Ao passo que, no caso de sistemas de informação já existentes, o ciclo de vida se inicia pela avaliação dos dados, recomendada para determinar a adequação dos dados para uso atual e/ou futuro, e é seguido do refinamento de metadados que se utiliza da avaliação para ajustar a estrutura existente, possibilitando assim a utilização da Arquitetura de Dados, para a criação de metadados e de estruturas de modelos de dados.

No domínio corporativo, para Inmon e Linstedt (2015) a maioria dos dados tem um ciclo de vida previsível, no qual os dados mais recentes tendem a ser acessados com mais frequência. Há, contudo, algumas exceções apontadas, relacionadas ao tipo de organização, como por exemplo o ramo de seguros e as

pesquisas científicas, nas quais o uso pode independender do tempo em que os dados estejam disponíveis no ciclo de vida.

O Ciclo de Vida dos Dados, de acordo com Inmon e Linstedt (2015), é dividido em duas etapas (Figura 19) que se aplicam a dados brutos detalhados. Neste Ciclo de Vida dos Dados (INMON; LINSTEDT, 2015), na primeira etapa, representada na Figura 19 por “ciclo de vida dos dados (1)” os dados entram no sistema e posteriormente passam por uma etapa de captura/edição, no qual recebem uma análise inicial e são tratados para que sejam mantidos no CVD ou, em alguns casos, descartados. Uma vez mantidos no CVD, os dados passam por um processo de organização, que consiste na indexação dos dados ou em processos de filtragem e fusão de dados, nos quais são moldados para que façam parte do CVD, na etapa de armazenamento.

Figura 19 – Ciclo de Vida dos Dados: 1 e 2



Fonte: Inmon e Linstedt (2015, p. 34, tradução nossa).

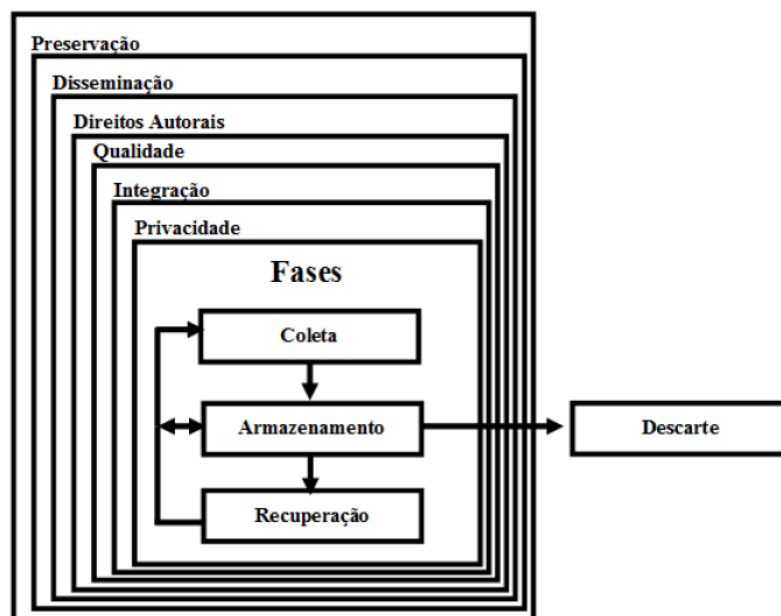
Enquanto na segunda etapa (INMON; LINSTEDT, 2015), representada na Figura 19 por “ciclo de vida dos dados (2)”, os dados armazenados podem ser submetidos a um processo de integração no qual são combinados com outros dados. Posteriormente, os dados entram em um novo processo, que os autores designam como “ciclo de utilidade” e, após seu uso, podem ser arquivados ou descartados.

No contexto da Ciência da Informação brasileira, Sant’Ana (2013) realizou um estudo no qual, a partir de modelos existentes de Ciclo de Vida dos Dados, propôs um modelo para a CI. O autor, por meio de uma pesquisa bibliográfica, analisou e discutiu Ciclos de Vida de Dados, de dados de pesquisa, de curadoria de dados, de metadados, de *Linked Open Data*, e de dados e publicações, dos quais extraiu elementos pertinentes à proposta de um Ciclo de Vida dos Dados para a Ciência da Informação, tendo como elemento central os dados.

Com isso, Sant’Ana (2013) inicialmente definiu fases e objetivos que compõem o Ciclo de Vida dos Dados na Ciência da Informação (Figura 20), sendo as fases relacionadas às etapas pelas quais os dados passam no CVD, ao passo que os objetivos são atividades necessárias e que podem ocorrer e permear diferentes fases

do CVD. Assim, o autor estabeleceu como fases, a coleta, o armazenamento, a recuperação e o descarte e, como objetivos a preservação, a disseminação, os direitos autorais, a qualidade, a integração e a privacidade.

Figura 20 – Ciclo de Vida dos Dados para Ciência da Informação – (CVD–CI)



Fonte: Sant'Ana (2013, p. 147).

Sant'Ana (2013) explica que:

No processo descrito no CVD-CI [...] a fase de coleta permite que seja iniciada a fase de armazenamento que por sua vez propicia a execução da fase de recuperação e ainda pode gerar novos dados retomando ações da fase de coleta. [...] Alcançados os objetivos previamente estabelecidos e ou alcançados os limites de tempo previstos para os dados, passa-se a fase de descarte, que dependendo da situação irá gerar dados, seja por transferência para outras bases seja para efeito de preservação de informações selecionadas da base a ser descartada, originando assim o início de atividades, ora da fase coleta, ora da fase de armazenamento.

De acordo com Sant'Ana (2013), na fase de coleta são realizadas definições iniciais como fonte de coleta, formas de obtenção e aplicação de filtros. Também se deve considerar a existência de metadados descritivos que possibilitem a avaliação, seleção e organização. Enquanto o armazenamento inclui atividades que auxiliam na persistência dos dados, como o processamento, a transformação, a modificação e a migração. Por fim, a recuperação operacionaliza o acesso aos dados, possibilitando consultas e visualizações, além de permitir a aplicação de estruturas, filtros, representação, refinamento, dentre outros.

Ainda, o autor posiciona uma fase de descarte, que consiste na eliminação dos dados, o que requer planejamento e ações para definir o escopo da limpeza, determinação de eliminação ou transferência para outro ambiente de armazenamento, lembrando que a eliminação tornará os dados inacessíveis; ou no desativamento dos dados, que são mantidos na mesma base de dados ou em outro ambiente de armazenamento, porém sem acesso pela fase de recuperação.

Sant'Ana (2019) destaca a relevância do papel dos detentores do ciclo de vida dos dados, uma vez que os algoritmos e o tratamento dos dados podem criar uma camada de opacidade sobre as operações e, com isso, dificultar a percepção dos usuários quanto às transduções realizadas no tratamento dos dados. Por isso alerta que:

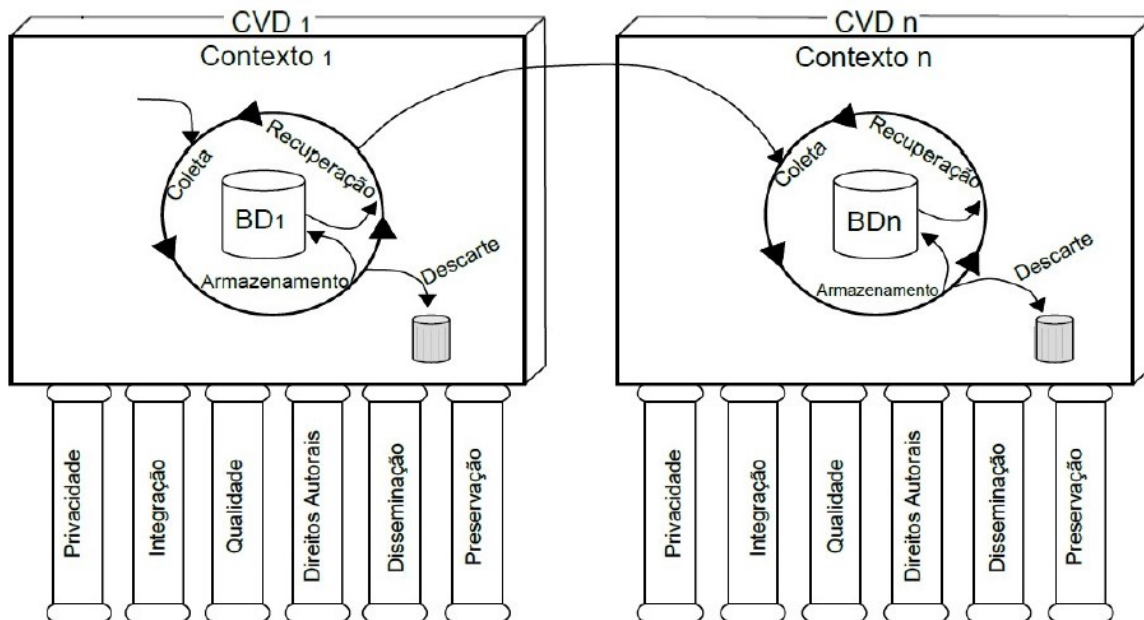
[...] apesar da aparente participação democrática dos diversos atores, são os detentores de dados [...] que controlam e direcionam a coleta, armazenamento e recuperação dos dados (SUN *et al*, 2015), sendo que tal opacidade resulta em profunda insciência dos usuários em relação à transdução dos conteúdos. (SANT'ANA, 2019, p. 124).

Essa insciência se refere aos dados que são coletados, seu tratamento, acesso e uso, bem como inferências que podem ser geradas sobre eles. Com isso, é necessário que haja conscientização de que há diferentes processos de coleta de dados, que podem estar implícitos ou explícitos no momento da interação com determinado ambiente informacional digital. Além disso, uma vez coletados, os dados recebem tratamento, armazenamento, descarte e podem ser disponibilizados para outros ciclos de vida dos dados, de acordo com os preceitos definidos pelo detentor, sobre os quais, muitas vezes, o referenciado não tem acesso.

Embora não seja parte do escopo desta discussão, é importante enfatizar o papel das legislações que buscam estabelecer limites para a coleta, armazenamento e recuperação dos dados, a exemplo da Lei Geral de Proteção de Dados Pessoais (LGPD) que “[...] dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural.”, em vigor no Brasil, de forma parcial desde 2018, e completa a partir de 2020 (BRASIL, 2018). No contexto da pesquisa, podem ser considerados ainda aspectos éticos e hegemonia de comunidades minoritárias (VIDOTTI; TORINO; CONEGLIAN, 2021).

O Ciclo de Vida dos Dados apresentado por Sant’Ana (2013) foi remodelado em Sant’Ana (2016, p. 123), cuja representação visual (Figura 21) exhibe o processo cíclico dos CVD, bem como a forma de relacionamento entre diferentes CVD.

Figura 21 – Ciclo de Vida dos Dados para Ciência da Informação – (CVD–CI)



Fonte: Sant’Ana (2016, p. 123).

Assim, Sant’Ana (2016, p. 119) esclarece que a primeira etapa de um ciclo de vida dos dados consiste na obtenção dos dados por meio da coleta, em que “são necessários esforços para que se possa estabelecer um plano de ação, análise da viabilidade bem como a execução da coleta dos dados”. O autor elucida a necessidade de competências relacionadas a esta etapa, com destaque às tecnologias digitais, Ciência da Informação e Ciência da Computação. A coleta pode ser realizada por meio da inserção de dados em um sistema de informação, ou da coleta desses dados, de bancos de dados ou outros CVD, bem como da coleta de interações de usuários com ambientes informacionais digitais.

Uma vez coletados, os dados podem ser utilizados e descartados ou ainda, mantidos para uso futuro, considerando que o custo de armazenamento tem se tornado baixo (SANT’ANA, 2016). Nesse sentido, o autor aponta que, quando realizado, o descarte pode ocorrer logo após a coleta ou posteriormente ao armazenamento. A etapa de descarte pode ocorrer em qualquer fase do CVD, uma vez que os dados não sejam mais necessários e puderem ser excluídos ou desativados (SANT’ANA, 2016). Segundo o autor, esta fase pode ser planejada por

profissionais da Ciência da Informação, contudo, a instrumentalização requer conhecimentos de Ciência da Computação.

Caso haja a definição pela manutenção dos dados coletados, Sant'Ana (2019) apresenta a segunda etapa do ciclo de vida dos dados, que consiste na persistência dos dados, objetivando seu armazenamento, o que requer competências das áreas de Ciência da Computação e Ciência da Informação.

Finalmente, os dados que foram coletados e armazenados, mantendo-se no ciclo de vida dos dados, devem ter condições necessárias para que sejam localizados, acessados e interpretados, na etapa de recuperação (SANT'ANA, 2016). Nesta etapa, o autor destaca as competências da área de Ciência da Informação e operacionalização pela Ciência da Computação.

Em todas as fases do CVD Sant'Ana (2016) elenca uma série de questionamentos que auxiliam no planejamento da coleta, do armazenamento, da recuperação e do descarte dos dados, que podem auxiliar na definição dos processos realizados em cada uma das etapas.

Vale destacar que, de acordo com a Figura 21, cada CVD está relacionado a um banco de dados, no qual os dados são armazenados e que, segundo as características estabelecidas pelo detentor do CVD, podem ser recuperados, quer seja para consultas ou ainda para alimentar outro CVD.

Assim, tanto na Figura 20 quanto na Figura 21, há fatores – privacidade, integração, qualidade, direitos autorais, disseminação e preservação – que, segundo Sant'Ana (2013, 2016), permeiam as quatro fases. Contudo, na Figura 21, a representação visual apresenta esses fatores como estruturantes e necessários à sustentação do CVD.

O CVD proposto por Sant'Ana (2013, 2016) também considera os usuários, e tem como atores:

- a) detentor: pessoa ou organização que estruturou o ciclo de vida dos dados e que, portanto, possui todo o conhecimento das suas fases. Responsável pela manutenção do ambiente e quem toma decisões acerca das formas de coleta, armazenamento, recuperação e descarte; e que pode realizar a atividade ou delegá-la a um intermediário, que realizará todas ou algumas fases do processo;
- b) intermediário: pessoa ou organização que desenvolveu a aplicação e possui conhecimentos acerca da tecnologia. Além desse, considera-se intermediário a pessoa ou organização que atua sobre o CVD por solicitação do detentor;

- c) referenciado: pessoa, organização ou atividade vinculada ou mencionada nos dados a serem disponibilizados e, por isso, consideradas referenciadas;
- d) usuário: sujeito cadastrado no sistema, que interage com o CVD, e que fornece seus dados consciente ou inconscientemente para armazenamento. (TORINO; VIDOTTI; SANT'ANA, 2019, p. 4-5).

O percurso realizado nessa seção visa destacar a necessidade de compreensão dos dados em seu Ciclo de Vida, que é composto por etapas como planejamento, definição de estratégias de visualização/coleta, tratamento, armazenamento, disponibilização, recuperação, acesso, uso, análise, avaliação e descarte, e envolve diferentes atores, como detentor, intermediário, referenciado e usuário. Dessa forma, elucida-se que o Ciclo de Vida dos Dados não se caracteriza apenas pela existência dos dados, mas por um processo cíclico (ciclo) presente nas diferentes etapas, desde a criação/coleta até a exclusão ou desativamento (de vida) dos dados em um banco de dados.

Com isso, apreende-se que o Ciclo de Vida dos Dados requer que os dados façam parte de uma estrutura formal de armazenamento, gerida por um detentor que orchestra as atividades realizadas, desde a captura (quer seja pela criação ou coleta) até a disponibilização para a recuperação, e que deve considerar fatores que permeiam as etapas do Ciclo de Vida dos Dados, como privacidade, integração, qualidade, direitos autorais, disseminação e preservação (SANT'ANA, 2013, 2016), além de aspectos éticos e legais, que podem influenciar o ciclo de vida.

Nesse sentido, o processamento adequado assegura qualidade aos dados, segurança aos envolvidos e a possibilidade de acesso e (re)uso aos dados existentes no CVD de forma colaborativa.

Por outro lado, é necessário atentar-se à relevância do tratamento adequado aos dados, seja antes ou depois de armazenados, para que possam ser recuperados, acessados e (re)utilizados para consumo ou composição de um novo CVD. Isso requer, ainda, atenção a princípios, como FAIR (*Findable, Accessible, Interoperable, Reusable*) e CARE (*Collective Benefit, Authority to Control, Responsibility, Ethics*) – no caso de povos originários ou minorias –, estruturas e padrões, protocolos.

Além disso, é importante ressaltar que, os dados que são coletados por um CVD a partir da recuperação de um outro CVD (Figura 21), devem considerar aspectos de proveniência, uma vez que, caso sejam tratados, podem, de alguma forma, divergir daqueles contidos no CVD originário. Neste contexto, a facilidade no processo de

acesso e (re)uso dos dados pode, de alguma forma, ampliar a disponibilização de dados em diferentes fontes de dados, porém com aspectos ou elementos divergentes entre eles, considerando justamente as definições do CVD que o armazena.

Considerando o Ciclo de Vida dos Dados enquanto elemento da Arquitetura de Dados no contexto da Ciência da Informação, é preponderante considerar que os dados, sejam eles criados ou coletados, entrarão em um Ciclo de Vida dos Dados, composto pelas etapas de coleta, armazenamento, disponibilização e descarte, e essas etapas são vitais para assegurar qualidade e credibilidade aos dados, mantendo históricos de ajustes que possam ocorrer nas diferentes fases e, ainda, quando de dados coletados, incluindo ao histórico a proveniência.

A discussão realizada acerca do CVD utiliza-se de alguns autores como referencial teórico, dos quais podem ser observadas diferentes boas práticas que podem, a critério do detentor do Ciclo de Vida dos Dados, ser incorporadas ao tratamento dos dados. Contudo, deve-se considerar que a discussão realizada nesta tese objetiva trazer elementos que sustentem o CVD enquanto elemento do conceito de Arquitetura de Dados no contexto da Ciência da Informação e não esgotar a discussão acerca de CVD.

Assim, é preponderante que no desenvolvimento do projeto de Arquitetura de Dados sejam tomadas as decisões concernentes ao CVD, no que tange à definição da necessidade de dados, a fim de identificar a existência ou fonte de coleta de dados e, a partir disso, estabelecer os elementos cabíveis. Quando da criação de dados, é necessário atender aos requisitos de criação de dados, ao passo que, quando da coleta dos dados, é imprescindível a análise preliminar do dado na fonte da qual será coletado, para averiguar os requisitos de coleta e os protocolos de comunicação de dados, estabelecer o tratamento adequado para que o dado possa ser coletado e compor a Arquitetura de Dados projetada.

Isso se relaciona com as “Principais Atividades do Ciclo de Vida dos Dados”, de DAMA-DMBOK (2017), apresentada na Figura 14, uma vez que é necessário planejamento, quer seja para a criação ou coleta de dados, cujos requisitos devem ser projetados e habilitados, para que os dados sejam criados ou obtidos/coletados, armazenados e mantidos, descartados ou desabilitados – se e quando necessário –, utilizados e enriquecidos. Todas essas atividades ocorrem de forma cíclica e contínua.

Nesse contexto, quando da coleta de dados de outras fontes, é preponderante considerar que outros elementos da Arquitetura de Dados no contexto da Ciência da

Informação devem ser considerados, a exemplo dos metadados e padrões utilizados, das estruturas sintáticas e semânticas e dos princípios éticos e legais que envolvem os dados. Com isso, os princípios técnicos podem ser empregados para que os dados localizados sejam acessados, processados e armazenados, o que se caracteriza como reuso de dados. Evidencia-se que, sempre que possível, devem ser adotadas técnicas e tecnologias emergentes que possibilitem automatizar o processo de reuso de dados, a fim de assegurar eficiência ao processo.

Vale considerar que, há etapas que não estão exibidas no Ciclo de Vida dos Dados e que possuem características próprias. Delas, no Fluxo de Dados na Arquitetura de Dados (Figura 32), são destacadas: o tratamento e a condição de acesso aos dados.

No que tange ao processo de aquisição, Levitin e Redman (1993) ilustram na Figura 15, relativa ao ciclo de aquisição de dados – que consiste na definição da visualização, implementação da visualização e obtenção de valores –, a necessidade de avaliar os dados obtidos, para definir se podem ser ajustados e atualizados, gerando novos valores já adequados aos requisitos para o armazenamento ou, quando não houver possibilidade de ajuste ou o custo de ajuste não se justificar, o descarte. Neste contexto, a avaliação e análise dos dados podem culminar em ajustes na requisição de dados e, com isso gerar novas definições de visualização.

Ainda, no que se refere ao ciclo de uso de dados, é importante que as requisições e os dados recuperados sejam avaliados em todas as etapas, desde a criação/coleta até o uso, o que possibilita, por meio das análises, gerar novas requisições e apoiar o processo de descarte dos dados coletados, visando inclusive novas coletas embasadas na análise – o que pode ser visualizado na Figura 16 de Levitin e Redman (1993). No que tange ao ciclo de uso de dados, ele se aplica a todo o CVD e não apenas às coletas de novos dados, e pode contribuir para assegurar a manutenção da qualidade dos dados armazenados.

Novamente, destaca-se que tanto a aquisição quanto o uso dos dados devem ser alicerçados em estruturas sintáticas e semânticas, padrões de dados e metadados e demais requisitos relacionados ao projeto de Arquitetura de Dados, que devem ser constantemente atualizados e, sempre que necessário, avaliados e ajustados. Nesse contexto, Yoon, Aiken e Guimaraes (2005) apresentam oito fases que podem ser adotadas tanto em CVD novo quanto em já existentes, uma vez que contempla desde

os requisitos de criação até a avaliação e o refinamento dos dados, o que pode auxiliar ainda na elaboração dos entregáveis da Arquitetura de Dados.

De igual maneira, no projeto de Arquitetura de Dados, devem ser considerados os atores envolvidos no CVD (TORINO; VIDOTTI; SANT'ANA, 2019), bem como os fatores relativos à privacidade, integração, qualidade, direitos autorais, disseminação e preservação (SANT'ANA, 2013, 2016) em todas as etapas que constituem o processamento dos dados, desde a criação/coleta, até a disseminação e, ainda, no descarte.

4.4.1.3 Metadados

A partir do domínio da Ciência da Informação, metadados, cujo desenvolvimento, atrelado ao uso das tecnologias de informática, “[...] refletem nitidamente a integração entre a representação e as tecnologias de informática.” (ALVES, 2010, p. 25), são apresentados e discutidos nessa seção, como elemento da Arquitetura de Dados.

Para Méndez Rodríguez (2002, p. 47, tradução nossa), metadados são utilizados para representar e organizar a informação contida em um objeto, visando melhorar o acesso, constituindo-se em “[...] estruturas de organização da informação, legíveis por máquinas, que visam tornar os dados úteis, de diferentes formas, de acordo com as necessidades específicas de cada serviço de informação digital e de acordo com a aplicação que lhes é concedida.”²²³

Alves (2010, p. 47) conceitua metadados como:

[...] atributos que representam uma entidade (objeto do mundo real) em um sistema de informação. Em outras palavras, são elementos descritivos ou atributos referenciais codificados que representam características próprias ou atribuídas às entidades [...] com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação.

Para Riley (2017, p. 2, tradução nossa),

Os metadados são fundamentais para a funcionalidade dos sistemas que contêm conteúdo, permitindo que os usuários encontrem itens de interesse,

²²³ “[...] estructuras de organización de la información, legibles por máquina, cuya finalidad es hacer útiles los datos, de distintas formas, según las necesidades concretas de cada servicio de información digital y según la aplicación que se les otorgue.”

registrem informações essenciais sobre eles e compartilhem essas informações com outras pessoas.²²⁴

Gilliland (2016) faz um apanhado da compreensão de metadados, desde seu uso enquanto padrão descritivo transparente até a relevância em sistemas computacionais, para processos de busca e intercâmbio de dados, incluindo o uso em ambientes web – com destaque para a afirmação “[...] a estrutura sempre foi importante na organização e representação da informação, mesmo antes da informatização.”²²⁵ (GILLILAND, 2016, tradução nossa).

Segundo a autora, há mais nos metadados além da descrição e recuperação de objetos

[...] os metadados não apenas identificam e descrevem um objeto de informação; também documentam como esse objeto se comporta, sua função e uso, seu relacionamento com outros objetos de informação e como ele deve ser e tem sido gerenciado ao longo do tempo.²²⁶ (GILLILAND, 2016, tradução nossa).

A contribuição de Gilliland é frequentemente utilizada na CI, tendo como referência a primeira edição, publicada em 2008. Destaca-se que, desde então, muitos avanços ocorreram no contexto dos metadados, das tecnologias empregadas e da web, o que reforça a necessidade de discussões acerca do tema.

Tais discussões têm sido realizadas em diferentes domínios, com preponderância no domínio web, sobretudo por profissionais que atuam no desenvolvimento conteúdos web e na otimização de conteúdos para mecanismos de busca (SEO) e para mecanismos de busca acadêmicos (ASEO), dada a relevância dos metadados para o posicionamento e ranqueamento web, que, por vezes, gera monetização direta e/ou indireta.

Nesse contexto, para Gill (2016, tradução nossa) “[...] metadados podem ser definidos como uma descrição estruturada dos atributos essenciais de um objeto.”²²⁷ O autor afirma que “Uma propriedade dos metadados é que eles são – ou deveriam

²²⁴ “Metadata is key to the functionality of the systems holding the content, enabling users to find items of interest, record essential information about them, and share that information with others.”

²²⁵ “[...] structure has always been important in information organization and representation, even before computerization.”

²²⁶ “[...] metadata not only identifies and describes an information object; it also documents how that object behaves, its function and use, its relationship to other information objects, and how it should be and has been managed over time.”

²²⁷ “[...] metadata can be defined as a structured description of the essential attributes of an object.”

ser – estruturados para modelar os atributos mais importantes do tipo de objeto que descrevem.”²²⁸ (GILL, 2016, tradução nossa). E continua:

A natureza estruturada dos metadados é importante. Ao modelar com precisão os atributos essenciais da classe de objetos que está sendo descrita, os metadados agregados podem servir como um catálogo - uma destilação dos atributos de uma coleção específica - tornando-se uma ferramenta útil para usar e gerenciar essa coleção [...].²²⁹ (GILL, 2016, tradução nossa).

Vale destacar que há uma ampla variedade de padrões de metadados disponíveis, cuja adoção e uso está intimamente relacionada ao objetivo e domínio de uso. Dessa forma, cada padrão possui suas especificidades, diretrizes e extensão, o que requer que haja conhecimento no momento do uso. Destaca-se ainda que a adoção de padrões de metadados possibilita o intercâmbio de dados e registros, favorecendo a comunicação fluída entre sistemas de informação e/ou ambientes informacionais digitais.

A crescente de ferramentas colaborativas e o ambiente web que possibilita a disponibilização de conteúdos e objetos por qualquer indivíduo, ampliou o uso dos metadados que inicialmente era realizado prioritariamente por profissionais especializados, de forma que esses indivíduos colaboram na ampla disseminação dos conteúdos. Contudo, a generalização do uso pode apresentar como problema a falta de padronização, inconsistências e falta de confiabilidade, o que pode prejudicar a interoperabilidade dos metadados.

Considerando as funcionalidades, Gilliland (2016, tradução nossa) apresenta cinco tipos de metadados:

- a) administrativos: “Metadados usados no gerenciamento e administração de coleções e recursos de informação”²³⁰;
- b) descritivos: “Metadados usados para identificar, autenticar e descrever coleções e recursos de informações confiáveis relacionados”²³¹;
- c) preservação: “Metadados relacionados à gestão de preservação de coleções e recursos de informação”²³²;

²²⁸ “One property of metadata is that it is—or should be—structured to model the most important attributes of the type of object that it describes.”

²²⁹ “The structured nature of metadata is important. By accurately modeling the essential attributes of the class of objects being described, metadata in aggregate can serve as a catalog – a distillation of the attributes of the particular collection – thereby becoming a useful tool for using and managing that collection [...].”

²³⁰ “Metadata used in managing and administering collections and information resources”

²³¹ “Metadata used to identify, authenticate, and describe collections and related trusted information resources”

²³² “Metadata related to the preservation management of collections and information resources”

- d) técnicos: “Metadados relacionados a como um sistema funciona ou os metadados se comportam”²³³;
- e) uso: “Metadados relacionados ao nível e tipo de uso de coleções e recursos de informação”²³⁴.

Aos quais, Riley (2017, tradução nossa) acrescenta:

- a) metadados estruturais: “[...] descrevem as relações de partes de recursos entre si”²³⁵;
- b) linguagem de marcação: “Integra metadados e sinalizadores para outros recursos estruturais ou semânticos no conteúdo”²³⁶.

Riley (2017, tradução nossa) apresenta ainda os usos primários para cada tipo de metadados:

- a) descritivos: descoberta, exibição e interoperabilidade;
- b) preservação e técnicos: interoperabilidade, gerenciamento de objetos digitais e preservação;
- c) uso: interoperabilidade e gerenciamento de objetos digitais;
- d) estruturais: navegação;
- e) linguagem de marcação: navegação e interoperabilidade.

Os metadados podem ser criados manual ou automaticamente, sendo, no último caso, gerados pelos sistemas de informação e/ou dispositivos utilizados na criação, acesso e uso dos objetos ou, ainda, quando da interação do sujeito com o ambiente informacional digital e/ou sistema de informação, dos quais os dados podem ser capturados e utilizados para diferentes finalidades.

Os metadados são considerados como um dos princípios da Gestão de Dados, sobretudo em função da sua relevância para o gerenciamento de dados, o que torna necessário o gerenciamento dos metadados.

De acordo com DAMA-DMBOK (2017, p. 27, tradução nossa) “Os metadados descrevem quais dados uma organização possui, o que eles representam, como são classificados, de onde vieram, como se movem dentro da organização, como evoluem

²³³ “Metadata related to how a system functions or metadata behaves”

²³⁴ “Metadata related to the level and type of use of collections and information resources”

²³⁵ “[...] describes the relationships of parts of resources to one another”

²³⁶ “Integrates metadata and flags for other structural or semantic features within content”

com o uso, quem pode ou não usá-los e se são de alta qualidade.²³⁷. Dessa forma os dados e seu ciclo de vida serão melhor compreendidos no contexto organizacional, considerando que sua criação e manutenção, em geral, ocorrem de forma díspar e descentralizada.

É preciso Metadados para gerenciar dados: Gerenciar qualquer ativo requer ter dados sobre esse ativo (número de funcionários, códigos contábeis, etc.). Os dados usados para gerenciar e usar dados são chamados de Metadados. Como os dados não podem ser mantidos ou tocados, entender o que é e como usá-los requer definição e conhecimento na forma de Metadados. Os metadados se originam de uma variedade de processos relacionados à criação, processamento e uso de dados, incluindo arquitetura, modelagem, administração, governança, gerenciamento de qualidade de dados, desenvolvimento de sistemas, operações de TI e negócios e análises.²³⁸ (DAMA-DMBOK, 2017, p. 21, tradução nossa).

Para Yoon, Aiken e Guimaraes (2005, p. 151, tradução nossa), metadados consistem em

Informação que se refere a uma ou mais outras informações que podem existir como formas físicas separadas. Em suma, dados sobre dados. No mundo da tecnologia da informação, o termo é frequentemente usado para indicar dados que se referem a recursos digitais disponíveis em uma rede.²³⁹

E estrutura de metadados consiste em: “Uma série de diretrizes que descrevem os componentes potenciais necessários para definir, descrever, publicar e manter metadados.”²⁴⁰ (YOON; AIKEN; GUIMARAES, 2005, p. 151, tradução nossa).

Na norma ISO/IEC 1179-1, metadados são definidos como “dados [...] que definem e descrevem outros dados”²⁴¹ (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015a, p. 3, tradução nossa).

No contexto do ciclo de vida dos dados, Sant’Ana (2013) afirma que:

²³⁷ “Metadata describes what data an organization has, what it represents, how it is classified, where it came from, how it moves within the organization, how it evolves through use, who can and cannot use it, and whether it is of high quality.”

²³⁸ “It takes Metadata to manage data: Managing any asset requires having data about that asset (number of employees, accounting codes, etc.). The data used to manage and use data is called Metadata. Because data cannot be held or touched, to understand what it is and how to use it requires definition and knowledge in the form of Metadata. Metadata originates from a range of processes related to data creation, processing, and use, including architecture, modeling, stewardship, governance, Data Quality management, systems development, IT and business operations, and analytics.”

²³⁹ “Information that refers to one or more other pieces of information that can exist as separate physical forms. In short, data about data. In the information technology world the term is often used to indicate data that refers to digital resources available across a network.”

²⁴⁰ “A series of guidelines describing potential components required to define, describe, publish and maintain metadata.”

²⁴¹ “data [...] that defines and describes other data”

Não seria possível tratar o ciclo de vida dos dados sem levar em consideração a importância da contextualização dos dados armazenadas, sejam eles administrativos, descritivos, técnicos, estruturais ou de preservação, a insuficiente carga semântica, própria dos dados, torna o uso de metadados um fator chave de sucesso.

Tal afirmação reforça a importância do contexto e da semântica para os dados em todas as etapas do ciclo de vida dos dados. Esse olhar do autor a partir do domínio da Ciência da Informação aproxima os dados da representação da informação, podendo ser assim compreendida por possibilitar, por meio de técnicas da representação tornar os dados compreensíveis para humanos e máquinas, de forma a auxiliar na compreensão necessária para que os dados sejam apreendidos como informação.

Destaca-se que Wurman (1996) afirma que não existe um curso para que as pessoas aprendam “compreensão”. Acredita-se que essa afirmação do autor se refira especificamente à compreensão por humanos, que está atrelada a aspectos cognitivos e subjetivos. Contudo, a compreensão por máquinas utiliza técnicas de programação, que podem se valer, por exemplo, de estruturas – defende-se nesta tese o uso de estruturas padronizadas internacionalmente para que a compreensão não seja dependente de aspectos locais –, bem como de bases de treinamento e bibliotecas, para as quais as estruturas e o contexto são igualmente relevantes.

No campo da estrutura dos dados, a compreensão se dá por características sintáticas e semânticas, uso de vocabulários, padrões e protocolos.

Pode-se considerar que a compreensão entre humanos, realizada por meio da comunicação – transmissor, receptor, canal, mensagem – pode ter ruídos, inconsistências e falhas. No cenário computacional, considerando a compreensão entre máquinas, tais ocorrências podem ser minimizadas com o uso de estruturas e padrões de dados (mensagem) e protocolos de comunicação (canal), para que a comunicação entre emissor e receptor seja mais efetiva.

As comunidades de informação estão cientes de que quanto mais altamente estruturado for um objeto de informação, mais essa estrutura pode ser explorada para busca, manipulação e inter-relação com outros objetos de informação. Capturar, documentar e aplicar essa estrutura, no entanto, só pode ocorrer se houver suporte para tipos específicos de metadados.²⁴² (GILLILAND, 2016).

²⁴² *“Information communities are aware that the more highly structured an information object is, the more that structure can be exploited for searching, manipulating, and interrelating with other information objects. Capturing, documenting, and enforcing that structure, however, can only occur if supported by specific types of metadata.”*

Dessa forma, faz-se necessário investir na estruturação de dados para que técnicas computacionais diversas possam ser utilizadas no consumo, processamento e análise dos dados de forma facilitada. Além disso, é possível fazer uso de aparatos tecnológicos para o processo de estruturação dos dados, por meio de metadados padronizados internacionalmente, reduzindo o custo de tratamento manual.

No contexto dos bancos de dados, Date (2004, p. 17) afirma que “A padronização da representação dos dados é particularmente desejável como auxílio ao intercâmbio de dados, ou a migração de dados entre sistemas [...]”, e complementa: “Da mesma forma, os padrões de nomenclatura e documentação de dados também são muito desejáveis como auxílio ao compartilhamento e à compreensão dos dados.”

Nesta tese, é importante discutir a norma ISO/IEC 11179, denominada *Information technology – Metadata registries* (MDR)/Tecnologia da informação – Registros de metadados (MDR), que está dividida em 6 partes – *Framework/Estrutura*, *Classification/Classificação* (retirada), *Registry metamodel and basic attributes/Metamodelo de registro e atributos básicos*, *Formulation of data definitions/Formulação de definições de dados*, *Naming principles/Princípios de nomenclatura*, *Registration/Registro*; às quais foi incluída a parte 7, denominada *Metamodel for data set registration/ Metamodelo para registro de conjunto de dados* – e busca promover:

- descrição padrão dos dados;
 - entendimento comum de dados entre elementos organizacionais e entre organizações;
 - reutilização e padronização de dados ao longo do tempo, espaço e aplicativos;
 - harmonização e padronização de dados dentro de uma organização e entre organizações;
 - gestão dos componentes das descrições dos dados;
 - reutilização dos componentes das descrições dos dados.²⁴³
- (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015a, p. v, tradução nossa).

No escopo da parte 1 da referida norma, dedicada à estrutura (ISO/IEC 11179-1), está estabelecido que essa parte (ISO/IEC 11179-1) fornece elementos necessários à compreensão de cada uma das 6 partes da norma e se constitui na base conceitual de metadados e registros de metadados.

²⁴³ “– *standard description of data; – common understanding of data across organizational elements and between organizations; – re-use and standardization of data over time, space, and applications; – harmonization and standardization of data within an organization and across organizations; – management of the components of descriptions of data; – re-use of the components of descriptions of data.*”

Considerando a relevância da compreensão no escopo da referida norma, nesta tese, utiliza-se a referida norma como referencial para alguns elementos da Arquitetura de Dados no contexto da Ciência da Informação, em especial no que tange aos metadados e às estruturas sintática e semântica.

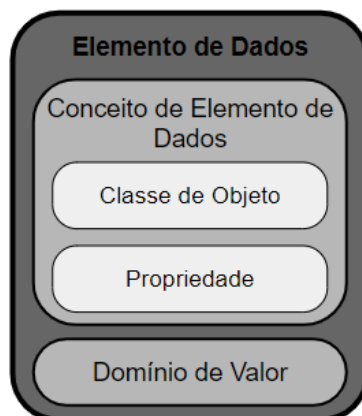
Vale destacar que a própria ISO/IEC 11179-1 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015a) explicita que o foco da norma é metadados enquanto descrição de dados. Assim, reforça que não aborda metadados de forma genérica, a exemplo da comum utilização do termo em dados descritivos. Assim, no contexto da norma, “Um registro de metadados (MDR) é um banco de dados de metadados.” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015a, p. v, tradução nossa). A ISO/IEC 11179 se aplica à estrutura de descrição geral para dados independentemente de seu tipo, vínculo e finalidade.

No que tange aos metadados, a norma menciona que são dados que definem e descrevem outros dados, de forma mais específica,

Isso significa que metadados são dados e dados se tornam metadados quando são usados dessa maneira. Isso acontece em circunstâncias particulares, para propósitos específicos e com certas perspectivas, pois nenhum dado é sempre metadado. O conjunto de circunstâncias, propósitos ou perspectivas para os quais alguns dados são usados como metadados é chamado de contexto. Assim, metadados são dados sobre dados em algum contexto.²⁴⁴ (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015a, p. 9, tradução nossa).

Assim, na ISO/IEC 11179-1 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015a), é estabelecido que um elemento de dados é constituído por conceito de elemento de dados, que se subdivide em classe de objeto e propriedade, e domínio de valor (Figura 22).

²⁴⁴ “This means that metadata are data, and data become metadata when they are used in this way. This happens under particular circumstances, for particular purposes, and with certain perspectives, as no data are always metadata. The set of circumstances, purposes, or perspectives for which some data are used as metadata is called the context. So, metadata are data about data in some context.”

Figura 22 – Modelo fundamental e informal de um Elemento de Dados

Fonte: International Organization for Standardization (2015a, p. 11, tradução nossa).

Considerando que a ISO/IEC 11179 está no domínio da tecnologia, apresenta-se a explicação da Figura 22 a partir a própria norma:

A totalidade de objetos para os quais desejamos coletar e armazenar dados é a extensão de uma classe de objeto. Classes de objetos são conceitos e correspondem às noções incorporadas em classes em modelos orientados a objetos e entidades em modelos entidade-relacionamento. [...] Propriedades são o que os humanos usam para distinguir ou descrever classes de objetos. São características, não necessariamente essenciais, da classe de objeto e formam sua intenção. Eles também são conceitos e correspondem às noções incorporadas em atributos (sem tipos de dados associados) em modelos orientados a objetos ou entidade-relacionamento. [...]. Os domínios de valor são conjuntos de valores permitidos para elementos de dados.²⁴⁵ (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015a, p. 11-122, tradução nossa).

Com base na ISO/IEC 11179-1 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015a), e a partir da compreensão da Figura 22, apreende-se que, a partir do domínio da Ciência da Informação, o modelo se aplica a Metadados, estruturas utilizadas para a representação de entidades a partir de seus atributos e valores. Desta forma, “Elemento de Dados” pode ser compreendido como “Metadados”, “Classe de Objeto” como “Entidade”, “Propriedade” como “Atributo” e “Domínio de Valor” como “Valor”.

²⁴⁵ “The totality of objects for which we wish to collect and store data is the extension of an object class. Object classes are concepts, and they correspond to the notions embodied in classes in object-oriented models and entities in entity-relationship models. [...] Properties are what humans use to distinguish or describe object classes. They are characteristics, not necessarily essential ones, of the object class and form its intention. They are also concepts, and they correspond to the notions embodied in attributes (without associated datatypes) in object-oriented or entity-relationship models. [...]. Value domains are sets of permissible values for data elements.”

Assim, metadados, amplamente definidos como “dados sobre dados” podem ser compreendidos a partir da representação gráfica da Figura 23, como: metadados são dados (valor) sobre dados (entidade e atributo).

Figura 23 – Metadados



Fonte: Adaptado de International Organization for Standardization (2015a, p. 11).

Cumprе esclarecer, a partir do domínio da Ciência da Informação, que metadados são estruturas de representação, atualmente utilizadas para armazenamento, recuperação e acesso a dados em sistemas computacionais, que expressa uma modelagem conceitual, a partir da qual se realiza uma implementação física. Assim, as entidades, que podem ser entendidas como objetos, possuem características próprias denominadas atributos, sobre os quais devem ser utilizados valores para registrar dados representacionais acerca de um objeto.

Para tanto, são utilizadas estruturas de representação padronizadas, denominadas padrões de metadados, para que os atributos sejam codificados e possam conferir a eles a semântica necessária para que sejam compreensíveis facilmente processáveis por humanos e máquinas, independentemente do tipo de dados, da finalidade de armazenamento, do idioma utilizado, do país, dentre outros. Destaca-se que o atributo em si possui uma sintaxe, por meio da qual a estrutura do padrão de metadados é registrada e, da mesma forma, o valor, entendido como a representação de uma característica de um objeto real, deve possuir sintaxe para a manutenção da qualidade dos registros.

Vale destacar que, os metadados possuem duas abordagens: computacional – armazenamento, recuperação, acesso e uso –, e, representação da informação – análise, descrição, indexação e características sintáticas – e que ambas são objeto da Ciência da Informação.

A norma ISO/IEC 11179-2 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2020, p. xv, tradução nossa) elucida que

O processamento de dados e o intercâmbio eletrônico de dados dependem fortemente de dados precisos, confiáveis, controláveis e verificáveis registrados em bancos de dados. Um pré-requisito para o uso e interpretação corretos e adequados dos dados é que tanto os usuários quanto os proprietários dos dados tenham um entendimento comum do significado e da representação dos dados. Para facilitar esse entendimento comum, várias características ou atributos dos dados devem ser definidos. Essas características dos dados são conhecidas como “metadados”, ou seja, “dados que descrevem dados”.²⁴⁶

Isto posto, a ISO/IEC 11179-3+A1 apresenta um modelo conceitual (Figura 24) que fornece a estrutura de um registro de metadados, para que sejam especificados e implementados. Tendo em vista que a referida norma tem aplicação no campo de tecnologia da informação, é de especial interesse para “[...] desenvolvedores de informações, gerentes de informações, administradores de dados, desenvolvedores de padrões, desenvolvedores de aplicativos, modeladores de negócios e outros responsáveis por tornar os dados compreensíveis e compartilháveis.”²⁴⁷ (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2020, p. xv, tradução nossa), aos quais incluímos Arquitetos de Dados, uma vez que a estrutura dos dados e metadados é parte do projeto de Arquitetura de Dados.

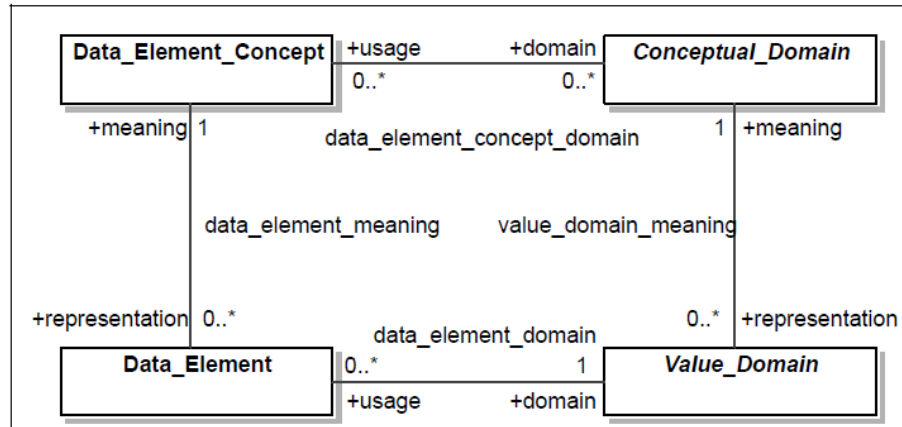
A Figura 24 pode ser dividida horizontalmente, de forma a apresentar na parte superior aspectos conceituais, e na parte inferior aspectos representacionais, bem como os relacionamentos possíveis. “Essa visão efetivamente divide o metamodelo entre um nível conceitual (ou semântico) (no topo) e um nível representacional (abaixo). O nível representacional descreve os artefatos de informação (em contraste

²⁴⁶ “Data processing and electronic data interchange rely heavily on accurate, reliable, controllable and verifiable data recorded in databases. A prerequisite for correct and proper use and interpretation of data is that both users and owners of data have a common understanding of the meaning and representation of the data. To facilitate this common understanding, a number of characteristics, or attributes, of the data have to be defined. These characteristics of data are known as “metadata”, that is, “data that describes data”.”

²⁴⁷ “[...] to information developers, information managers, data administrators, standards developers, application developers, business modellers and others who are responsible for making data understandable and shareable.”

com as construções semânticas do nível superior).”²⁴⁸ (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2020, p. 93, tradução nossa).

Figura 24 – High-level Data Description metamodel²⁴⁹



Fonte: International Organization for Standardization (2020, p. 93).

Cumpra esclarecer que o metamodelo apresentado na Figura 24 é um modelo conceitual de alto nível, de forma que apresenta apenas os elementos principais, que são detalhados na norma ISO/IEC 11179-3+A1 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2020).

Para auxiliar na interpretação da Figura 24, a partir da ISO/IEC 11179-3+A1, tem-se como exemplo para o nível conceitual: *Conceptual_Domain*, “países do mundo”, que pode se relacionar com diferentes *Data_Element_Concept*, como “país de nascimento”, “país de residência”, “país de falecimento”. Cujas repercussões no nível de representação utilizaria em *Value_Domain*, os valores autorizados para representar o que foi estabelecido no nível conceitual, cujo relacionamento se dá por meio de *value_domain_meaning*.

Assim, o domínio conceitual fornece significado para que *Value_Domain* possa fornecer uma representação, que se utiliza de uma estrutura conceitual externa ao modelo. Nesse caso, por exemplo, pode ser utilizada a ISO 3166 - padrão internacional para códigos de país e códigos para as suas subdivisões. Ao passo que *Data_Element* consiste na representação do *Data_Element_Concept*, seguindo os padrões e definições adotados no Ciclo de Vida dos Dados, utilizando, por exemplo,

²⁴⁸ “This view effectively splits the metamodel between a conceptual (or semantic) level (at the top) and a representational level (below). The representational level describes the information artifacts (in contrast to the semantic constructs of the upper level).”

²⁴⁹ Por se tratar da estrutura de um registro de metadados, optou-se por manter a figura no idioma original.

atributos de um padrão de metadados (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2020). Nesse caso, por exemplo, o Data_Element para o Data_Element_Concept “país de nascimento” utilizando Schema.org²⁵⁰ seria birthPlace.

Essa exemplificação materializa a relação entre os aspectos conceituais e representacionais necessários à materialização dos dados, bem como dos metadados.

A ISO/IEC 11179-4 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2004, p. 4-6, tradução nossa) estabelece os requisitos para a criação de metadados, a saber:

- a) ser indicado no singular: “O conceito expresso pela definição de dados deve ser expresso no singular. (Uma exceção é feita se o próprio conceito for plural.)”²⁵¹;
- b) afirmar o que é o conceito, não apenas o que não é: “Ao construir definições, o conceito não pode ser definido exclusivamente pela afirmação do que o conceito não é.”²⁵²;
- c) ser declarado como uma frase ou sentença descritiva (na maioria dos idiomas): “Uma frase é necessária (na maioria dos idiomas) para formar uma definição precisa que inclua as características essenciais do conceito. A simples indicação de um ou mais sinônimos é insuficiente. Simplesmente reafirmar as palavras do nome em uma ordem diferente é insuficiente. Se for necessária mais do que uma frase descritiva, use frases completas e gramaticalmente corretas.”²⁵³;
- d) conter apenas abreviaturas comumente compreendidas: “A compreensão do significado de uma abreviatura, incluindo siglas e inicialismos, geralmente fica restrita a um determinado ambiente. Em outros ambientes, a mesma abreviação pode causar interpretação errônea ou confusão.

²⁵⁰ “Schema.org é uma atividade colaborativa e comunitária com a missão de criar, manter e promover esquemas para dados estruturados na Internet, em páginas da Web, em mensagens de e-mail e muito mais.” Disponível em: <https://schema.org/>.

²⁵¹ “The concept expressed by the data definition shall be expressed in the singular. (An exception is made if the concept itself is plural.)”

²⁵² “When constructing definitions, the concept cannot be defined exclusively by stating what the concept is not.”

²⁵³ “A phrase is necessary (in most languages) to form a precise definition that includes the essential characteristics of the concept. Simply stating one or more synonym(s) is insufficient. Simply restating the words of the name in a different order is insufficient. If more than a descriptive phrase is needed, use complete, grammatically correct sentences.”

Portanto, para evitar ambiguidade, palavras completas, e não abreviações, devem ser usadas na definição. Exceções a este requisito podem ser feitas se uma abreviação for comumente entendida [...]”²⁵⁴.

- e) ser expresso sem incorporar definições de outros dados ou conceitos subjacentes: “[...] a definição de um segundo elemento de dados ou conceito relacionado não deve aparecer na definição propriamente dita do elemento de dados primário. As definições dos termos devem ser fornecidas em um glossário associado. Se a segunda definição for necessária, ela pode ser anexada por uma nota no final do texto principal da definição primária ou como uma entrada separada no dicionário.”²⁵⁵

Além dos requisitos, a ISO/IEC 11179-4 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2004, p. 4-6, tradução nossa) estabelece, ainda, como recomendações: indicar o significado essencial do conceito; ser preciso e inequívoco; ser conciso; ser claro; ser expresso sem incorporar lógica, uso funcional, informações de domínio ou informações de procedimento; evitar o raciocínio circular; usar a mesma terminologia e estrutura lógica consistente para definições relacionadas; ser apropriado para o tipo de item de metadados que está sendo definido.

Vale destacar que os requisitos e as recomendações estabelecidos pela ISO/IEC 11179-4 contribuem para assegurar a semântica dos dados e metadados, mantendo o registro adequado.

De igual maneira, a ISO/IEC 11194-5 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015b, p. v, tradução nossa) estabelece princípios e regras de nomenclatura, considerando que

O objetivo de qualquer convenção de nomenclatura é permitir o desenvolvimento de nomes para itens que tenham o máximo de clareza e transparência de significado, combinados com concisão, exigindo mínimo esforço de interpretação por parte do usuário final, sujeito às restrições do sistema sob o qual os itens são processados.²⁵⁶

²⁵⁴ “Understanding the meaning of an abbreviation, including acronyms and initialisms, is usually confined to a certain environment. In other environments the same abbreviation can cause misinterpretation or confusion. Therefore, to avoid ambiguity, full words, not abbreviations, shall be used in the definition. Exceptions to this requirement may be made if an abbreviation is commonly understood [...]”

²⁵⁵ “[...] the definition of a second data element or related concept should not appear in the definition proper of the primary data element. Definitions of terms should be provided in an associated glossary. If the second definition is necessary, it may be attached by a note at the end of the primary definition's main text or as a separate entry in the dictionary.”

²⁵⁶ “The goal of any naming convention is to allow development of names for items that have maximum clarity and transparency of meaning, combined with concision, demanding minimal effort of interpretation by the end user, subject to the constraints of the system under which the items are processed.”

Desta forma, a utilização de princípios e padrões de nomenclatura auxilia na compreensão dos termos, sobretudo quando comparado ao uso da linguagem natural. Ainda, as partes 4 e 5 da ISO/IEC 11179 reforçam a importância de documentar as diretrizes, padrões e estruturas relacionadas ao registro e à nomenclatura dos dados.

A nomenclatura objetiva ainda descrever como os nomes são formulados, considerando a consistência e a semântica, cujas definições devem constar em um documento de referência, que possa ser acionado por humanos e máquinas

No contexto do *Data Reference Model* (DRM) metadados são relevantes pois:

- Promovem uma compreensão mais clara;
- Ajudam os usuários a encontrar informações;
- Promovem uso e reuso mais eficientes da informação;
- Promovem uma melhor gestão de dados, nomeadamente em ambientes de “Big Data”; e
- Usam métodos de práticas recomendadas de descrição de dados DRM.²⁵⁷ (UNITED STATES, 2013, p. 164, tradução nossa).

No contexto do Modelo Referencial de Dados (DRM), de acordo com *Federal Enterprise Architecture Program* (2005, p. 18, tradução nossa), “Se a Descrição de Dados for feita corretamente, ela terá um impacto positivo na eficácia da missão. Se for mal feito, impede essa eficácia, às vezes com resultados desastrosos quando os dados necessários para a tomada de decisões não podem ser encontrados.”²⁵⁸ Tal afirmação pode ser generalizada para outros contextos e para quaisquer tipos de dados.

Farrel e Bengtson (2019) afirmam que, no contexto dos dados de pesquisa, a comunidade tende a demorar para adotar práticas de compartilhamento de dados. Para os autores, “O compartilhamento eficaz de dados não é uma tarefa trivial. Requer nova infraestrutura, adoção de novas práticas e consenso por parte dos investigadores dentro das disciplinas.”²⁵⁹ (FARREL; BENGTON, 2019, p. 3, tradução nossa).

Farrel e Bengtson (2019) destacam ainda que os dados coletados podem ter valor significativo se houver possibilidade de uso futuro, o que requer que os dados sejam imbuídos de contexto, ainda que o contexto original dos dados seja alterado.

²⁵⁷ “• Promotes clearer understanding; • Helps users find information; • Promotes more efficient use and reuse of information; • Promotes better data management, particularly in environments of “Big Data”; and • Uses DRM Data Description Best Practice Methods.”

²⁵⁸ “If Data Description is done right it has a positive impact on mission effectiveness. If it is done wrong it impedes that effectiveness, sometimes with disastrous results when data needed for decision making cannot be found.”

²⁵⁹ “Effective data sharing is not a trivial undertaking. It requires new infrastructure, adoption of new practices, and consensus by the investigators within disciplines”

Para os autores, a coleta de dados, ainda que possua metadados, necessita que eles sejam adequadamente estruturados, para que termos e definições utilizados sejam perenes e precisos. “Os metadados devem ser estruturados logicamente para fornecer uma estrutura para conectar os dados em uma coleta de dados ao resto do mundo de maneira significativa.”²⁶⁰ (FARREL; BENGTSON, 2019, p. 5, tradução nossa).

O estudo realizado por Farrel e Bengtson (2019), no contexto dos dados de pesquisa, aponta que há dificuldade em estabelecer a granularidade de metadados descritivos necessários para uma coleta de dados. Os autores apontam que muitos repositórios de dados de pesquisa adotam um padrão mínimo de metadados descritivos, mas alertam sobre a necessidade de discutir o uso de metadados adicionais, que possibilitem descrever os dados de maneira exaustiva, incluindo aspectos relevantes ao contexto original do experimento. Neste sentido, embora o estudo supracitado se utilize de dados de pesquisa, tal ponderação pode ser utilizada em muitos outros contextos, nos quais possa haver a necessidade de descrição exaustiva de entidades.

Tais aspectos, desde a relevância e a forma de estruturação dos dados e metadados até sua operacionalização e dificuldades de compartilhamento, podem ser trabalhados, por exemplo, com a adoção de um vocabulário padronizado de metadados, atendimento a princípios técnicos como FAIR e apoio de ferramentas de representação da informação.

Nesse sentido, destaca-se a preocupação elencada por Gilliland (2016) no que tange aos desafios a serem enfrentados pelos profissionais, na criação de objetos informacionais e sistemas de informação, aos quais incluem-se os ambientes informacionais digitais, tendo em vista os objetos, usuários e funcionalidades:

- a) identificar o(s) esquema(s) de metadados que melhor atenda(m) às necessidades do gerenciamento que será realizado;
- b) definir a especificidade de uso dos metadados, para estabelecer a granularidade;
- c) estabelecer os critérios para representação da informação – incluindo ferramentas como vocabulários controlados, tesouros, taxonomias e folksonomias.

²⁶⁰ “Metadata should be structured logically to provide a framework to connect the data within a data collection to the rest of the world in a meaningful way.”

Vale ter em mente que há complexidade no estabelecimento desses elementos e das padronizações que o uso de metadados requer. Contudo, segundo Woodley (2016), embora práticas locais sejam adequadamente realizadas, é no momento do compartilhamento entre diferentes instâncias que surgem os problemas de interoperabilidade, para os quais a adoção de metadados bem estruturados e mapeados pode ser a solução.

Com isso, a relevância dos metadados para a Arquitetura de Dados é reforçada, objetivando tornar os dados compreensíveis por humanos e máquinas, bem como favorecer a localização, o acesso, o uso e o reuso para diferentes finalidades, com barreiras mínimas.

Além disso, a interoperabilidade é imprescindível para a constituição das Ecologias Complexas de Dados, o que requer, para além da adequada estrutura dos metadados, o uso de protocolos comuns de comunicação de dados e, possivelmente, a compatibilização de estruturas, por meio da conversão de registros ou do *crosswalk* de metadados.

[...] nos referiremos ao mapeamento como a atividade intelectual de comparar e analisar dois esquemas de metadados, e *crosswalks* como o produto visual do mapeamento. Um *crosswalk* é uma tabela ou gráfico que mapeia os relacionamentos e equivalências entre dois ou mais formatos de metadados.²⁶¹ (WOODLEY, 2016, tradução nossa).

No que tange ao mapeamento de metadados, destaca-se a necessidade de compatibilização semântica, no que se refere aos atributos e aos valores utilizados. Em ambos os casos, pode haver complexidade no tratamento, no caso dos atributos, principalmente aquelas relacionadas à granularidade e, no caso dos valores, aos padrões utilizados para a representação.

Neste sentido, ressalta-se a necessidade de documentar as definições relativas aos metadados e seu ciclo de vida, a fim de que possam apoiar mapeamentos e *crosswalks* sempre que se fizer necessário. Tal afirmação encontra lugar na ISO/IEC 11179-4 “As definições devem ser escritas para facilitar a compreensão por qualquer usuário e pelos destinatários dos dados compartilhados.”²⁶² (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2004, p. 4, tradução nossa).

²⁶¹ “[...] we will refer to mapping as the intellectual activity of comparing and analyzing two metadata schemas, and to crosswalks as the visual product of mapping. A crosswalk is a table or chart that maps the relationships and equivalencies between two or more metadata formats.”

²⁶² “Definitions should be written to facilitate understanding by any user and by recipients of shared data.”

Como possibilidade de tratamento de metadados, Woodley (2016) aponta:

[...] converter ou transformar metadados de um padrão para outro; migrar metadados de um padrão legado para outro diferente; integrando dois padrões de metadados; e coleta ou agregação de metadados criados usando um padrão comunitário compartilhado ou padrões de metadados diferentes.²⁶³

Na Arquitetura de Dados, o tratamento dos dados e metadados é preponderante para assegurar a entrada dos dados no Ciclo de Vida dos Dados de forma adequada, atendendo às definições realizadas, o que favorece a consistência dos dados e metadados e o intercâmbio dos mesmos em ambiente interno e/ou externo.

Os metadados, enquanto elemento da Arquitetura de Dados, atuam como elo entre a Ciência da Informação e as tecnologias empregadas pelos Sistemas de Informação, o que encontra lugar nas afirmações de Alves (2010) e Méndez Rodríguez (2002) posicionadas no início da seção. Nesse contexto, atuam como aparato tecnológico imprescindível para a estrutura dos dados em sistemas de informação, possibilitando que os dados sejam acionáveis e contextualizados.

Vale destacar que os metadados são amplamente discutidos na Ciência da Informação. Contudo, em geral as discussões se ancoram nas suas tipologias (GILLILAND, 2016; RILEY, 2017), nos esquemas e padrões disponíveis e na representação da informação no campo de valor dos metadados. Essa abordagem é claramente importante para a Arquitetura de Dados, uma vez que são parte da compreensão e composição teórico-conceitual dos metadados, e que impactam a qualidade dos dados representacionais em ambientes informacionais, visível, por exemplo, na Recuperação da Informação, na Encontrabilidade da Informação, na criação de perfis de aplicação e na compatibilização de dados, com favorecimento a humanos e máquinas

Por outro lado, a Ciência da Informação tem pouco envolvimento na discussão dos registros de metadados enquanto atributo codificado, considerando, por exemplo a ISO/IEC 11179. Nesse sentido, os metadados, que consistem na estrutura de representação da informação, e são igualmente importantes e impactam nos processos de Recuperação da Informação, *Linked Data*, enriquecimento de dados,

²⁶³ “[...]converting or transforming metadata from one standard to another; migrating metadata from one legacy standard to a different one; integrating two metadata standards; and harvesting or aggregating metadata created using a shared community standard or different metadata standards.”

localização, acesso, processamento, uso, integração e interoperabilidade de dados, sobretudo para máquinas, cuja eficiência beneficia diretamente humanos e máquinas. E, essa discussão é imprescindível para a Arquitetura de Dados, uma vez que é parte da estruturação dos dados, por meio de esquemas, padrões e estruturas, que precisam ser compreensíveis por humanos e máquinas, para que haja localização, acesso, uso e reuso.

Dessa forma, para a Arquitetura de Dados no contexto da Ciência da Informação, os metadados devem ser considerados teórico-conceitualmente de forma integral, com atenção aos campos de atributo e valor, cuja carga semântica é destacada por Sant'Ana (2013).

Vale reforçar que os metadados estão diretamente vinculados aos outros elementos da Arquitetura de Dados, por ser parte do Ciclo de Vida dos Dados, requerer estruturas sintática e semântica, representar atributos relacionados aos princípios éticos e legais aplicados aos dados e ser parte integrante e fundamental dos princípios técnicos aplicados aos dados.

Assim, na Arquitetura de Dados, os metadados contribuem para a comunicação de dados entre máquinas e, quando estruturados e padronizados, com carga sintática e semântica, possibilitam que os dados (mensagem) sejam transmitidos por protocolos de comunicação de dados (canal) entre máquinas (emissor e receptor) com o menor ruído possível, o que se caracteriza pela compreensão de dados por máquinas.

Por isso, é importante que os metadados atendam a requisitos internacionais para que os dados, independentemente de qualquer decisão local, possam ser localizados, compreendidos, acessados e processados por qualquer consumidor – com destaque para as requisições automáticas ou automatizadas.

Assim, na Arquitetura de Dados é possível que os metadados sejam padronizados por estruturas já existentes e/ou ainda, quando da necessidade de Registros de metadados/*Metadata registries* (MDR), é fundamental o atendimento às normas relacionadas, com destaque nesta tese à ISO/IEC 11179, sobretudo para a criação de metadados, cuja semântica e documentação favorecem o fornecimento e consumo.

Nesse contexto, é recomendável ainda que a documentação relacionada aos metadados seja clara e concisa, além de fornecida em padrões abertos e, em diferentes formatos, disponível em um repositório de metadados. O referido repositório

de metadados, juntamente com requisitos para inclusão e exclusão de metadados e perfil de aplicação de metadados são entregáveis da Arquitetura de Dados que estão diretamente relacionados ao elemento metadados e que se relaciona ainda com o modelo lógico de dados e os mapas de fluxos de dados.

Os atributos de metadados, quando padronizados – quer seja por esquemas e padrões existentes ou pela criação de registros de metadados –, podem, na medida em que estão documentados, fornecer a semântica necessária para a comunicação de dados entre sistemas de informação, ainda que seja necessário realizar alguma compatibilização ou conversão de registros.

Isso reforça a necessidade de gerenciamento dos metadados por meio do repositório digital de metadados, além de se constituir em um dos elementos chave dos Princípios FAIR.

Destaca-se que a estruturação e documentação adequadas dos metadados favorecem o compartilhamento entre diferentes instâncias, interna e externamente à Arquitetura de Dados, bem como a constituição de Ecologia Complexa de Dados (ECD).

4.4.1.4 Estruturas Sintática e Semântica dos Dados

No processo de conceituação da Arquitetura de Dados, a partir da Teoria do Conceito (DAHLBERG, 1978a, 1978b), apresenta-se como elemento a necessidade de que os dados tenham estruturas sintática e semântica. Para tanto é imprescindível compreender como essas estruturas se estabelecem.

Inicialmente, temos em Barbosa (2016, p. 2) a sintaxe do ponto de vista linguístico

Sintaxe vem do termo grego *syntaxis* (ordem, disposição). Tradicionalmente, ela se refere à parte da Gramática que descreve o modo como as palavras são combinadas para compor as sentenças. Essa descrição é organizada sob a forma de regras.

Em Cunha e Cavalcante (2008, p. 340), sintática é compreendida como “[...] estudo da combinação dos signos entre si.”, e sintaxe, como:

[...] ordem das palavras na mensagem, ou em parte da mensagem, como as orações, palavras e sentenças. A sintaxe também pode ser definida como o estudo das palavras que formam os períodos e os parágrafos e dos parágrafos que formam o discurso. Em resumo, sintaxe é a estrutura gramatical da língua.

Cunha e Cavalcante (2008, p. 340), apresentam ainda a sintaxe relacionada aos metadados e sistemas de informação:

No Dublin Core, a forma e estrutura pelas quais os elementos de metadados e seus componentes são combinados para formar um registro de metadados. [...] as regras que definem a sequência dos elementos de uma linguagem de programação.

Para Khosrow-Pour (2007, p. 663, tradução nossa), sintaxe é “O arranjo gramatical das palavras nas frases. (Galitsky, 2005c)”²⁶⁴. E metadados sintáticos são “Metadados que descrevem a estrutura e/ou formato de um documento. Por exemplo, idioma do documento, tamanho do documento e tipo MIME podem ser incluídos como elementos de metadados sintáticos. (Banerjee et al., 2005)”²⁶⁵.

Coneglian (2020, p. 103), ao utilizar a sintática como um dos níveis do processamento de linguagem natural, elucida que:

No nível humano, as pessoas criam frases seguindo regras sintáticas, em que as palavras assumem determinados papéis, criando assim uma oração, que tenha sentido gramatical. Em nível de sistema, requer-se uma gramática e um analisador para realizar o processamento. Em suma, a saída de um analisador sintático é uma representação da sentença, contendo as relações de dependência entre os termos.

De acordo com Dienstbach *et al.* (2021, p. 14) semântica “[...] é a área da Linguística que se ocupa dos processos lógicos, cognitivos e discursivos responsáveis pela produção e pela compreensão dos significados de palavras, frases e enunciados que se manifestam nas situações de uso da língua.” e, investiga:

- as propriedades formais que compõem os conceitos;
- os conhecimentos dos falantes a respeito daquilo que falam e ouvem (ou escrevem e leem);
- as pistas contextuais que orientam os sentidos de palavras e frases. (DIENSTBACH *et al.*, 2021, p. 14).

Ainda no contexto da linguística, Dienstbach *et al.* (2021) aborda a semântica formal, uma vertente da semântica, para a qual o significado está relacionado à referencialidade entre os idiomas e as referências a objetos, pessoas, fatos do mundo real, considerando a sua descrição, o que requer conhecimento do falante acerca do idioma e do mundo real. Além da referencialidade, a semântica formal requer

²⁶⁴ “The grammatical arrangement of words in sentences. (Galitsky, 2005c)”

²⁶⁵ “Metadata that describes a document’s structure and/or format. For example, document language, document size, and MIME type might all be included as elements of syntactic metadata. (Banerjee et al., 2005)”

composicionalidade, para a qual o significado é também dependente da estrutura gramatical da sentença.

Desta forma, para a semântica formal, na linguística, o significado é dependente da relação entre a língua, a referência ao objeto, pessoa ou fato existente no mundo externo ao falante, bem como da estrutura sintática de uma sentença.

Essa abordagem da semântica formal é facilmente compreensível no contexto humano, no qual o processamento é cognitivo. Mas deixa claro que há a necessidade de conhecimentos referenciais, dos termos que representam objetos, pessoas e fatos que são externos ao indivíduo falante, bem como da própria língua para, inclusive, processar metáforas, ironias e outros aspectos que podem estar presentes em uma sentença verbal ou registrada.

Cunha e Cavalcante (2008, p. 330) definem semântica como:

“Estudo da relação de significação nos signos e da representação do sentido dos enunciados” (aur). ⇔ enunciado, signo. 1.2. “Ciência que estuda a significação [...] A semântica pode ocupar-se da significação ao nível da palavra, da sentença, do discurso ou do texto” (bid). 1.3. Estudo das relações entre os signos e os objetos que são por eles designados, bem como o relacionamento com os significantes e seus significados. ⇔ sentido, significação, significado, significante, signo. 2. inf parte da definição de uma linguagem de programação que se ocupa de especificar o efeito de um texto que está construído segundo as regras de sintaxe da linguagem. 3. bib/cat no Dublin Core, sentido dos elementos individuais dos metadados e seus componentes.

Para Khosrow-Pour (2007, p. 606, tradução nossa), semântica,

Em uma linguagem, é o significado de uma *string*, ao contrário da sintaxe, que descreve como os símbolos da linguagem são combinados. A maioria das linguagens de programação tem sua sintaxe definida formalmente (tradicionalmente em BNF [Backus-Naur Form]), enquanto as linguagens de especificação formal também têm sua semântica definida formalmente. (Dasso & Funes, 2005) 2: A informação deve estar de forma inteligível para quem a recebe. (Kisielnicki, 2006) [...]. No estudo da linguagem, a semântica se preocupa com o significado das palavras, expressões e frases, muitas vezes em relação à referência e à verdade. [...] O objetivo da semântica é atribuir um significado aos elementos sintáticos. Para serviços da Web, é o significado acordado de dados, funções, QoS e assim por diante trocados entre dois ou mais serviços. (Cardoso, 2006).²⁶⁶

²⁶⁶ “In a language, it is the meaning of a string, as opposed to syntax, which describes how the symbols of the language are combined. Most programming languages have their syntax defined formally (traditionally in BNF), while formal specification languages have also their semantics defined formally. (Dasso & Funes, 2005) 2: Information should be in a form which is intelligible to those receiving it. (Kisielnicki, 2006) [...]. In the study of language, semantics is concerned with the meaning of words, expressions, and sentences, often in relation to reference and truth. [...] The purpose of semantics is to assign a meaning to syntactical elements. For Web services, it is the agreed-upon meaning of data, functions, QoS, and so forth exchanged between two or more services. (Cardoso, 2006)”

Considerando que a obra de Khosrow-Pour (2007) está no domínio da tecnologia da informação, outras definições de semântica são apresentadas e, possuem relação com as discussões desta tese. É o caso de Banco de dados corporativo semanticamente modelado, apresentado como

Um banco de dados que é um reflexo da realidade das atividades nas quais uma empresa se engaja e dos recursos e pessoas envolvidas nessas atividades. A semântica está presente no modelo conceitual, mas pode não ser prontamente aparente no banco de dados implementado. (Dunn & Grabski, 2005).²⁶⁷ (KHOSROW-POUR, 2007, p. 606, tradução nossa).

Neste contexto, apreende-se que os dados podem ser estruturados semanticamente no modelo conceitual, embora seja recomendável que essa estrutura se reflita na implementação física dos dados em sistemas de informação e ambientes informacionais digitais planejados a partir da Arquitetura de Dados. O que se pode fazer por meio de tecnologia semântica, definida como

Uma tecnologia de software que permite que o significado e as associações entre as informações sejam conhecidos e processados em tempo de execução. Para que uma tecnologia semântica funcione verdadeiramente dentro de um sistema, deve haver um modelo de conhecimento de alguma parte do mundo que seja usado por um ou mais aplicativos em tempo de execução. (Kapetanos, 2005)²⁶⁸ (KHOSROW-POUR, 2007, p. 605, tradução nossa).

E, no que tange aos metadados, Khosrow-Pour (2007, p. 604, tradução nossa) define metadados semânticos como

Metadados que descrevem o conteúdo de um documento dentro do contexto de um determinado domínio de conhecimento. Por exemplo, para documentos relacionados ao domínio de segurança interna, os metadados semânticos podem incluir nomes de terroristas, afiliações de grupos e assim por diante. (Banerjee et al., 2005) 2: Dados sobre dados que descrevem o conteúdo dos dados. Um exemplo representativo de metadados semânticos é relacionar dados com classes de uma ontologia – isto é, o uso de ontologia para descrever dados. (Aleman-Meza et al., 2005)²⁶⁹.

²⁶⁷ “A database that is a relection of the reality of the activities in which an enterprise engages, and the resources and people involved in those activities. The semantics are present in the conceptual model, but might not be readily apparent in the implemented database. (Dunn & Grabski, 2005)”

²⁶⁸ “A software technology that allows the meaning of and associations between information to be known and processed at execution time. For a semantic technology to be truly at work within a system, there must be a knowledge model of some part of the world that is used by one or more applications at execution time. (Kapetanos, 2005)”

²⁶⁹ “Metadata that describes the content of a document within the context of a particular domain of knowledge. For example, for documents relating to the homeland security domain, semantic metadata may include terrorist names, group affiliations, and so forth. (Banerjee et al., 2005) 2: Data about data that describes the content of the data. A representative example of semantic metadata is relating data with classes of an ontology—that is, the use of ontology for describing data. (Aleman-Meza et al., 2005)”

É por meio da semântica dos dados que os dados armazenados em diferentes fontes serão compatibilizados para o modelo de dados utilizado, o que requer que haja documentação acionável por humanos e máquinas visando estabelecer a compreensão.

Para Coneglian (2020, p. 104), a semântica “[...] aprofunda as relações entre as palavras, além de apontar os seus significados.” Assim, “Quando pensado sob a perspectiva humana, a semântica está na interpretação do sentido dos termos, de forma individual e relacionada a outros termos.” Ao passo que “No nível de sistema, o processamento semântico irá determinar os possíveis significados de uma determinada frase ou conjunto de frases, focando principalmente na interação entre os termos dentro da sentença.”

O estudo de Coneglian (2020) aborda processamento de linguagem natural, e, por isso, destaca a complexidade do tratamento semântico, uma vez que há subjetividade na compreensão do significado dos termos e, por vezes, a necessidade de desambiguação a partir do contexto, que será dado por outros termos. Com isso, a abordagem de Coneglian (2020) está na sintaxe e na semântica empregadas no contexto da busca realizada por processamento de linguagem natural, o que justifica a preponderância dada pelo autor para a expressão dos termos.

Por outro lado, nos sistemas de informação, a compreensão do significado também se aplica aos metadados, tanto no campo de atributo quanto no campo de valor. Nesse caso, a sintaxe dará a estrutura e a semântica o significado aos dados e metadados.

A norma ISO/IEC 11179-1 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015a, p. v, tradução nossa) “[...] aborda a semântica dos dados, a representação dos dados e o registro das descrições desses dados. É por meio dessas descrições que uma compreensão precisa da semântica e uma descrição útil dos dados são encontradas.”²⁷⁰

Desta forma, considerando que a norma aborda especificamente registros de metadados, a compreensão da sua relevância nesta tese, para além do que se discute acerca dos metadados na Representação da Informação, está nas estruturas sintática

²⁷⁰ “[...] addresses the semantics of data, the representation of data and the registration of the descriptions of that data. It is through these descriptions that an accurate understanding of the semantics and a useful depiction of the data are found.”

e semântica por eles fornecida e que é de especial relevância para a Arquitetura de Dados.

“Um [registro de metadados] MDR pode conter a semântica dos dados. Uma compreensão dos dados é fundamental para seu projeto, harmonização, padronização, uso, reuso e intercâmbio.”²⁷¹ (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015a, p. v, tradução nossa).

Destaca-se, contudo, da própria norma, que

[...] é aplicável à formulação de representações de dados, conceitos, significados e relacionamentos a serem compartilhados entre pessoas e máquinas, independentemente da organização que produz os dados. Não é aplicável à representação física de dados como bits e bytes no nível da máquina.²⁷² (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015a, p. 1, tradução nossa).

Semântica é definida na norma ISO/IEC 11179-5 como “ramo da ciência linguística que lida com os significados das palavras” e, sintaxe como “[...] estrutura de expressões em uma língua, e as regras que governam a estrutura de uma língua” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015b, p. 5-6, tradução nossa).

A referida norma estabelece ainda que sintaxe, semântica e léxico são dependentes de definições locais, o que amplia a necessidade de documentação dos padrões e estruturas utilizados, de forma acionável, visando a compreensão por humanos e máquinas. “As regras sintáticas, semânticas e lexicais variam de acordo com as organizações, como corporações ou órgãos de definição de padrões para setores de negócios; cada um pode estabelecer regras para a formação de nomes dentro de seu(s) contexto(s).”²⁷³ (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015b, p. 8, tradução nossa).

A norma ISO/IEC 11179-5 esclarece que as regras semânticas auxiliam na transmissão do significado, as regras sintáticas especificam a ordenação consistente, enquanto as regras lexicais ampliam a precisão e reduzem a redundância.

Os princípios semânticos são entendidos como:

²⁷¹ “An [metadata registries] MDR may contain the semantics of data. An understanding of data is fundamental to their design, harmonization, standardization, use, re-use and interchange.”

²⁷² “[...] is applicable to the formulation of data representations, concepts, meanings and relationships to be shared among people and machines, independent of the organization that produces the data. It is not applicable to the physical representation of data as bits and bytes at the machine level.”

²⁷³ “Syntactic, semantic and lexical rules vary by organizations such as corporations or standards-setting bodies for business sectors; each can establish rules for name formation within its context(s).”

A semântica diz respeito aos significados das partes do nome e possivelmente aos separadores que os delimitam. O conjunto de regras semânticas documenta se os nomes transmitem ou não significado e, em caso afirmativo, como. As partes do nome podem ser derivadas de conjuntos de estruturas que identificam relacionamentos entre (classificar) membros.²⁷⁴ (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015b, p. 9, tradução nossa).

Ao passo que os princípios sintáticos são entendidos como:

A sintaxe especifica a disposição das partes dentro de um nome. O arranjo pode ser especificado como relativo ou absoluto, ou alguma combinação dos dois. O arranjo relativo especifica partes em termos de outras partes [...]. A disposição absoluta especifica uma ocorrência fixa da parte [...].²⁷⁵ (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015b, p. 9, tradução nossa).

Diante do exposto, destaca-se que “Para entender, relacionar e interpretar dados, é necessário o significado explícito dos dados, que é dado pelo aproveitamento efetivo das tecnologias e abordagens semânticas.” (CONEGLIAN; DIEGER; SANTAREM SEGUNDO; CAPRETZ, 2018, p. 141).

Nesse sentido, e considerando a Arquitetura de Dados que possui dentre os elementos a atuação na estruturação dos dados e a necessidade das estruturas sintática e semântica, bem como de esquemas e padrões de metadados, é imprescindível que haja compreensão dos padrões sintáticos e semânticos para o planejamento e projeto da Arquitetura de Dados e para o tratamento dos dados.

Vale destacar que, no que tange aos dados, a semântica formal é necessária para que haja compreensão clara da semântica pelas máquinas. Contudo, elevar a carga semântica dos dados é um processo que necessita possibilitar a compreensão do significado dos dados para máquinas, processo esse imprescindível quando se trata de comunicação de dados entre sistemas de informação e/ou ambientes informacionais e, cuja complexidade é ampliada na Ecologia Complexa de Dados.

Vale lembrar que Berners-Lee (2009) apresentou a seguinte definição “Metadados são informações compreensíveis por máquina sobre recursos da web ou outras coisas”²⁷⁶. Considerando que esta tese realiza o percurso de buscar desemaranhar dados e informações, torna-se necessário esclarecer que aquilo que o

²⁷⁴ “Semantics concerns the meanings of name parts and possibly separators that delimit them. The set of semantic rules documents whether or not names convey meaning, and if so, how. Name parts may be derived from structure sets that identify relationships among (classify) members.”

²⁷⁵ “Syntax specifies the arrangement of parts within a name. The arrangement may be specified as relative or absolute, or some combination of the two. Relative arrangement specifies parts in terms of other parts [...]. Absolute arrangement specifies a fixed occurrence of the part [...].”

²⁷⁶ “Metadata is machine understandable information about web resources or other things”.

autor chama informações devemos compreender como dados. Da definição de Berners-Lee (2009), é importante destacar o fato de que os metadados são tidos como “compreensíveis por máquinas”, ao que o autor acrescenta:

Estamos falando aqui de informações que os agentes de software podem usar para facilitar a nossa vida, garantir que obedecemos nossos princípios, a lei, verificar se podemos confiar no que estamos fazendo e fazer tudo funcionar de forma mais suave e rápida. Os metadados têm semântica e estrutura bem definidas.²⁷⁷

Assim, a “Interoperabilidade de dados pode ser contextualizada a partir da capacidade fornecida aos sistemas para interpretar de maneira automática e precisa o significado dos dados trocados.” (CONEGLIAN; DIEGER; SANTAREM SEGUNDO; CAPRETZ, 2018, p. 143).

Desta forma, como em qualquer processo de comunicação, é necessário que haja elementos para que determinada mensagem seja enviada por um emissor, por meio de um canal, para um receptor que, ao recebê-la, necessita compreendê-la. Assim, considerando que emissor e receptor são agentes ou aplicações computacionais, portanto máquinas, a mensagem a ser transmitida, na forma de dados, precisa utilizar de uma linguagem, que possua sintaxe definida, de forma que haja ordenação, disposição e demais elementos que a estructurem e que precisam ser ou se fazer conhecidos pelo receptor para que a comunicação se efetive.

É imprescindível que o receptor compreenda o significado da mensagem, o que pode requerer conhecimento acerca do objeto tratado na mensagem, acesso a pistas contextuais e, no contexto da semântica, referencialidade, para que possa fazer uso.

Assim, os dados devem ser estruturados por meio de metadados internacionalmente conhecidos e adotados, para que tenham carga semântica, que pode ser elevada por meio do uso de tecnologias da web semântica, a exemplo de linguagens de marcação, estabelecimento de relacionamentos e ontologias.

Soares *et al.* (2020) afirmam que, no contexto dos dados FAIR, é necessário que os metadados tenham regras definidas e claras, que expressem características sintáticas e semânticas. “A semântica dá o significado de cada elemento do esquema de metadados, definindo sua função no registro, enquanto a sintaxe determina como

²⁷⁷ “We are talking here about information which software agents can use in order to make life easier for us, ensure we obey our principles, the law, check that we can trust what we are doing, and make everything work more smoothly and rapidly. Metadata has well defined semantics and structure.”

formatar os metadados de forma interoperável.”²⁷⁸ (SOARES *et al.*, 2020, p. 7, tradução nossa).

No que tange às tecnologias da web semântica, “Em comparação com outras tecnologias, as tecnologias semânticas fornecem conhecimento prévio para o contexto dos dados, interoperabilidade, escalabilidade, integração e aceitos como padrão de expressividade de dados.” (CONEGLIAN; DIEGER; SANTAREM SEGUNDO; CAPRETZ, 2018, p. 141).

Ao que Coneglian, Roa-Martinez, Ferreira, Vidotti e Santarem Segundo (2019, p. 26) acrescentam “A [*Extensible Markup Language*] XML pode ser compreendida como a sintaxe utilizada para a disponibilização das informações, que permite agregar semântica aos documentos e possibilita que cada aplicação crie uma interpretação da marcação atribuída ao conteúdo.” Ao passo que a *Resource Description Framework* (RDF) possibilita estabelecer relacionamentos entre recursos, utilizando propriedades, nas chamadas triplas RDF, o que favorece a semântica dos dados e, com isso, sua interoperabilidade.

Destaca-se que Berners-Lee (2009), cuja primeira edição data de 1997, afirma:

No futuro, gostaríamos de avançar para um estado em que quaisquer metadados ou, eventualmente, qualquer forma de dados legíveis por máquina carregassem uma referência à especificação da semântica de todas as asserções feitas nele.

Tal estado se faz presente sempre que há declarações especificadas que auxiliam na ampliação do nível de semântica dos dados.

Isso posto, a Arquitetura de Dados pode auxiliar na melhoria do nível sintático e semântico dos dados, por considerá-los imprescindíveis para a compreensão do sentido dos dados por aplicações e/ou agentes computacionais, o que favorece a comunicação de dados, em fluxos de integração e interoperabilidade, além do processamento, uso e reuso.

Considerando que a Arquitetura de Dados requer a compreensão dos dados por máquinas, as estruturas sintáticas e semânticas são fundamentais para que haja comunicação e compreensão sem ruídos. Dessa forma, as estruturas sintáticas e

²⁷⁸ “The semantic gives the meaning of each element of the metadata schema, setting its function on the registry, while the syntax determines how to format the metadata in an interoperable way.”

semânticas estão intimamente ligadas aos metadados, com aplicação aos atributos e valores utilizados na representação.

Na Arquitetura de Dados, os elementos sintáticos estabelecem a estrutura dos metadados, ao passo que os elementos semânticos atuam no significado. Essas estruturas são refletidas na documentação dos dados e expressas no repositório digital de metadados, o que auxilia na compreensão por humanos e máquinas.

Na Arquitetura de Dados, o uso de esquemas de metadados padronizados e internacionalmente aceitos possibilita que os atributos utilizem um vocabulário que beneficia a compreensão dos dados por máquinas, por meio da sua estrutura semântica. Nesse sentido, quando da opção pela criação de esquema de metadados próprios, fazê-lo de forma normalizada, conforme estabelece, por exemplo a ISO/IEC 11179, e com documentação consistente, possibilita que esse vocabulário, apesar de próprio, seja compreendido por máquinas e, com isso, possa ser acionado e convertido para outro(s) padrão(ões) de dados mediante processamento.

Obviamente, a criação de novos esquemas de metadados pode gerar, em alguns casos, silos de dados, sobretudo em estruturas que utilizem padrões sintáticos e semânticos muito divergentes dos esquemas padronizados do domínio. Nesse contexto, vale lembrar que a programação de sistemas de informação se faz pelas linguagens de programação e os dados, em geral, são interoperados em processamentos que utilizam linguagens de marcação, sendo estas linguagens constituídas por regras próprias que estabelecem suas estruturas sintáticas e semânticas.

Com isso, de forma análoga ao que ocorre na linguagem compreensível por humanos, é necessário que a comunicação entre máquinas ocorra tendo em vista a linguagem e suas regras. Caso contrário, será necessário investimento para a conversão, que consiste em uma espécie de tradução dessas linguagens e regras para uma outra que seja conhecida, o que consiste no processo de conversão de registros.

Essa analogia pretende esclarecer que, assim como no contexto humano, não há uma linguagem e estruturas de linguagem universal no contexto das máquinas. Contudo, cada uma delas possui regras próprias, de forma que todos os sistemas que operam com as mesmas linguagens e estruturas possam ter os dados compreensíveis entre si. Essa compreensão é que possibilita fluidez entre os dados em diferentes

sistemas de informação, o que se beneficia com a Arquitetura de Dados no contexto da Ciência da Informação, conceituada nesta tese.

4.4.1.5 Princípios Éticos, Legais e Técnicos aplicados aos Dados

A crescente variedade, volume e velocidade com que os dados são criados, coletados e armazenados traz consigo inúmeros desafios que podem ser enfrentados por meio de definições realizadas no projeto de Arquitetura de Dados, visando a estruturação e o tratamento adequados nos processos de coleta, armazenamento, tratamento e disseminação.

Neste sentido, evidencia-se a relevância dos princípios éticos, legais e técnicos implicados na criação, coleta, armazenamento, fornecimento e consumo de dados, independentemente da tipologia e finalidade.

No Brasil, muito se discute na atualidade acerca dos dados pessoais, dados sensíveis, e das condições legais que estão implicadas na criação, fornecimento e consumo de tais dados. Ao passo que, sobretudo no que concerne aos dados governamentais e aos dados de pesquisa, há interesse crescente na gestão, visando armazenamento, acesso e (re)uso.

Do estudo de Damiani *et al.* (2021) que menciona a complexidade da transformação digital na área de saúde, por diferentes razões, que passam pelo processo de uso de tecnologias digitais, integração e transformação de dados, habilidade dos usuários e os desafios éticos e de privacidade, destaca-se:

Desafios éticos e de privacidade: proteção da privacidade do paciente e critérios éticos fortes e inquestionáveis, com foco constante no benefício para pacientes e cuidados, devem ser continuamente validados e rastreados, alavancando órgãos e especialistas independentes, com estrita adesão às normas regulatórias e práticas robustas enraizadas em execução diária de todos os atores.²⁷⁹ (DAMIANI *et al.*, 2021, p. 2, tradução nossa).

Embora a preocupação apresentada pelos autores esteja no domínio da saúde, ela é utilizada nesta tese como base para uma generalização possível. É imprescindível que haja atenção à privacidade dos sujeitos referenciados nos dados, com atenção estrita aos aspectos éticos e legais vigentes. Neste sentido, vale lembrar

²⁷⁹ “Ethical and privacy challenges: protection of patient’s privacy and strong, unquestionable ethical criteria, with constant focus of benefit for patients and care, must be continuously validated and tracked, leveraging independent bodies and experts, with strict adherence to regulatory norms and robust practices engrained in daily execution from all actors”

que cada jurisdição possui legislações próprias que devem ser conhecidas e observadas. Esta questão se coloca como um desafio no domínio web, no qual as fronteiras podem não estar bem estabelecidas ou ainda serem desconhecidas, requerendo atenção no tratamento dos dados.

No que tange aos aspectos técnicos, há requisitos definidos em caráter local. Contudo, é imprescindível que seja expandida a atenção para os requisitos técnicos internacionais, que norteiam as práticas de coleta, armazenamento, disseminação, compartilhamento, integração e interoperabilidade de dados, independente da ferramenta tecnológica utilizada.

Isto posto, é possível compreender que os princípios éticos e legais atuam de forma preponderante na condução do tratamento dos dados visando resguardar características relevantes dos sujeitos e contextos a eles envolvidos. Ao passo que os princípios técnicos se ocupam das condições técnicas e tecnológicas para operacionalizar as requisições dos princípios éticos e legais aplicados aos dados e, na medida do que tais princípios possibilitem, possam ampliar a condição de localização, acesso e (re)uso dos dados armazenados.

Tem-se, assim, três aspectos relacionados ao Big Data (variedade, volume e velocidade), de forma que é necessário igualmente reforçar a relevância da veracidade, o que pode se constituir em um aspecto ético. A veracidade está evidentemente atrelada à fidedignidade de armazenamento do dado criado ou coletado. Porém, Damiani *et al.* (2021) fazem um alerta sobre a presença de dados ausentes ou que estão errados, e são de fácil detecção, sobre os quais serviços automatizados podem ser aplicados para a identificação e, quando possível, correção, dentre os quais mencionam, como exemplo, “[...] ninguém tem 4 m de altura, nem vive 200 anos; não há estágio 7 na classificação do câncer, e assim por diante.”²⁸⁰ (DAMIANI *et al.*, 2021, p. 2, tradução nossa).

Há de se ponderar ainda as questões de segurança, compreendidas como as definições de proteção aos dados disponíveis nos sistemas de informação e/ou ambientes informacionais digitais no que tange ao acesso, uso, alteração e perdas, tendo em vista a necessidade de integridade, disponibilidade e confidencialidade, quando se aplicarem.

²⁸⁰ “[...] nobody is 4 m tall, or lives 200 years; there is no stage 7 in cancer classification, and so on.”

Tendo em vista que os Princípios Éticos, Legais e Técnicos aplicados aos Dados consistem em Elementos da Arquitetura de Dados no contexto da Ciência da Informação, essa discussão preliminar será aprofundada nas seções 4.4.1.5.1 a 4.4.1.5.3.

4.4.1.5.1 Princípios éticos aplicados aos dados

A crescente disponibilização de dados em diferentes sistemas de informação é incontestavelmente uma realidade há muito tempo. O advento das tecnologias possibilita que dados sejam capturados de forma automática nas mais diversas interações realizadas em sistemas de informação, ambientes informacionais digitais e aparatos tecnológicos. E essa ampla variedade de dados é capturada muitas vezes de forma imperceptível ao usuário.

A possibilidade de gerar valor a partir dos dados, por seu uso, por exemplo, em sistemas de recomendação, *marketing* de produtos e serviços, análises e previsões, tornou-os passíveis de monetização direta e indireta, o que consiste do aproveitamento comercial dos dados, sejam eles capturados nas interações, armazenados por razão específica ou coletados de outras fontes de armazenamento.

Em paralelo, a solicitação social de transparência dos dados, sobretudo aqueles vinculados a processos realizados por instituições públicas e o movimento de pesquisadores e demais profissionais no bojo da Ciência Aberta, fizeram um movimento em prol da abertura dos dados, sejam eles governamentais ou de pesquisa.

No Brasil, estes cenários levaram à necessidade do estabelecimento de marcos legais que possibilitassem reger as formas de coleta, tratamento, armazenamento, acesso, uso, compartilhamento e reuso dos dados em diferentes âmbitos. Dentre eles, estão a Lei de Acesso à Informação (LAI) (BRASIL, 2011) e a Política de Dados Abertos do Governo Federal (BRASIL, 2016a).

A Lei de Acesso à Informação (BRASIL, 2011) estabelece no art. 8 que, “É dever dos órgãos e entidades públicas promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas.”, utilizando-se dos instrumentos e meios disponíveis, com obrigatoriedade de disponibilização nos *websites* oficiais.

Já a Política de Dados Abertos do Poder Executivo Federal (BRASIL, 2016a) estabelece seus objetivos no art. 1, dentre os quais se destacam os incisos I, “promover a publicação de dados contidos em bases de dados de órgãos e entidades da administração pública federal direta, autárquica e fundacional sob a forma de dados abertos”, e IV, “facilitar o intercâmbio de dados entre órgãos e entidades da administração pública federal e as diferentes esferas da federação”.

De acordo art. 2, inciso III da Política de Dados Abertos do Poder Executivo Federal (BRASIL, 2016a)

dados abertos - dados acessíveis ao público, representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte.

Contudo, tais ordenamentos jurídicos estabeleceram a abertura dos dados sem especificar Princípios Éticos e Técnicos para o tratamento dos dados, desde sua criação/coleta até a disponibilização para acesso e reuso por terceiros.

Neste sentido, a LAI estabelece no art. 6º, inciso III que órgãos e entidades do poder público devem assegurar “III - proteção da informação sigilosa e da informação pessoal, observada a sua disponibilidade, autenticidade, integridade e eventual restrição de acesso.” (BRASIL, 2011). Além disso, dedica a Seção III à Proteção e Controle de Informações Sigilosas e a Seção V às Informações Pessoais.

A Política de Dados Abertos do Poder Executivo Federal (BRASIL, 2016a) estabelece no Art. 5º § 3º “A [Infraestrutura Nacional de Dados Abertos] INDA poderá estabelecer normas complementares relacionadas com a elaboração do Plano de Dados Abertos, bem como relacionadas a proteção de informações pessoais na publicação de bases de dados abertos nos termos deste Decreto.”

Ainda no Brasil, no âmbito da pesquisa científica já existiam preceitos éticos aplicados à pesquisa envolvendo seres humanos e animais.

No que tange à pesquisa envolvendo seres humanos, há duas resoluções do Ministério da Saúde vigentes, a Resolução MS/CNS nº 466 de 12 de dezembro de 2012 (BRASIL, 2012) e Resolução MS/CNS nº 510 de 07 de abril de 2016 (BRASIL, 2016b), além da Norma Operacional nº 001/2013, que são a base para a atuação do chamado do Sistema CEP/CONEP, formado pela Comissão Nacional de Ética em Pesquisa (CONEP), instância máxima nacional de avaliação de pesquisas envolvendo seres humanos, vinculada ao Conselho Nacional de Saúde (CNS), e pelos Comitês

de Ética em Pesquisa (CEP) que estão disponíveis em diferentes regiões do país. A atuação do Sistema CEP/CONEP utiliza a Plataforma Brasil²⁸¹ como base nacional para os registros de pesquisas envolvendo seres humanos.

Na Resolução MS/CNS nº 466 (BRASIL, 2012), a pesquisa envolvendo seres humanos é entendida como “[...] pesquisa que, individual ou coletivamente, tenha como participante o ser humano, em sua totalidade ou partes dele, e o envolva de forma direta ou indireta, incluindo o manejo de seus dados, informações ou materiais biológicos”. E a seção III é dedicada aos Aspectos Éticos da Pesquisa Envolvendo Seres Humanos, que devem atender a fundamentos éticos e científicos, subdivididos em três categorias – eticidade da pesquisa, exigências em pesquisa com seres humanos e exigências em pesquisas com seres humanos que utilizam metodologias experimentais na área biomédica.

Para a Resolução MS/CNS nº 466 (BRASIL, 2012), a eticidade da pesquisa implica em:

- a) respeito ao participante da pesquisa em sua dignidade e autonomia, reconhecendo sua vulnerabilidade, assegurando sua vontade de contribuir e permanecer, ou não, na pesquisa, por intermédio de manifestação expressa, livre e esclarecida;
- b) ponderação entre riscos e benefícios, tanto conhecidos como potenciais, individuais ou coletivos, comprometendo-se com o máximo de benefícios e o mínimo de danos e riscos;
- c) garantia de que danos previsíveis serão evitados; e
- d) relevância social da pesquisa, o que garante a igual consideração dos interesses envolvidos, não perdendo o sentido de sua destinação sócio-humanitária.

Também estabelece como exigências para a realização de pesquisas envolvendo seres humanos, em qualquer área do conhecimento:

- a) ser adequada aos princípios científicos que a justifiquem e com possibilidades concretas de responder a incertezas;
- b) estar fundamentada em fatos científicos, experimentação prévia e/ou pressupostos adequados à área específica da pesquisa;
- c) ser realizada somente quando o conhecimento que se pretende obter não possa ser obtido por outro meio;
- d) buscar sempre que prevaleçam os benefícios esperados sobre os riscos e/ou desconfortos previsíveis;
- e) utilizar os métodos adequados para responder às questões estudadas, especificando-os, seja a pesquisa qualitativa, quantitativa ou quali-quantitativa;
- f) se houver necessidade de distribuição aleatória dos participantes da pesquisa em grupos experimentais e de controle, assegurar que, *a priori*, não seja possível estabelecer as vantagens de um

²⁸¹ Disponível em: <https://plataformabrasil.saude.gov.br/login.jsf>

- procedimento sobre outro, mediante revisão de literatura, métodos observacionais ou métodos que não envolvam seres humanos;
- g) obter consentimento livre e esclarecido do participante da pesquisa e/ou seu representante legal, inclusive nos casos das pesquisas que, por sua natureza, impliquem justificadamente, em consentimento a *posteriori*;
- h) contar com os recursos humanos e materiais necessários que garantam o bem-estar do participante da pesquisa, devendo o(s) pesquisador(es) possuir(em) capacidade profissional adequada para desenvolver sua função no projeto proposto;
- i) prever procedimentos que assegurem a confidencialidade e a privacidade, a proteção da imagem e a não estigmatização dos participantes da pesquisa, garantindo a não utilização das informações em prejuízo das pessoas e/ou das comunidades, inclusive em termos de autoestima, de prestígio e/ou de aspectos econômico-financeiros;
- j) ser desenvolvida preferencialmente em indivíduos com autonomia plena. Indivíduos ou grupos vulneráveis não devem ser participantes de pesquisa quando a informação desejada possa ser obtida por meio de participantes com plena autonomia, a menos que a investigação possa trazer benefícios aos indivíduos ou grupos vulneráveis;
- k) respeitar sempre os valores culturais, sociais, morais, religiosos e éticos, como também os hábitos e costumes, quando as pesquisas envolverem comunidades;
- l) garantir que as pesquisas em comunidades, sempre que possível, traduzir-se-ão em benefícios cujos efeitos continuem a se fazer sentir após sua conclusão. Quando, no interesse da comunidade, houver benefício real em incentivar ou estimular mudanças de costumes ou comportamentos, o protocolo de pesquisa deve incluir, sempre que possível, disposições para comunicar tal benefício às pessoas e/ou comunidades;
- m) comunicar às autoridades competentes, bem como aos órgãos legitimados pelo Controle Social, os resultados e/ou achados da pesquisa, sempre que estes puderem contribuir para a melhoria das condições de vida da coletividade, preservando, porém, a imagem e assegurando que os participantes da pesquisa não sejam estigmatizados;
- n) assegurar aos participantes da pesquisa os benefícios resultantes do projeto, seja em termos de retorno social, acesso aos procedimentos, produtos ou agentes da pesquisa;
- o) assegurar aos participantes da pesquisa as condições de acompanhamento, tratamento, assistência integral e orientação, conforme o caso, enquanto necessário, inclusive nas pesquisas de rastreamento;
- p) comprovar, nas pesquisas conduzidas no exterior ou com cooperação estrangeira, os compromissos e as vantagens, para os participantes das pesquisas e para o Brasil, decorrentes de sua realização. Nestes casos deve ser identificado o pesquisador e a instituição nacional, responsáveis pela pesquisa no Brasil. Os estudos patrocinados no exterior também deverão responder às necessidades de transferência de conhecimento e tecnologia para a equipe brasileira, quando aplicável e, ainda, no caso do desenvolvimento de novas drogas, se comprovadas sua segurança e eficácia, é obrigatório seu registro no Brasil;

- q) utilizar o material e os dados obtidos na pesquisa exclusivamente para a finalidade prevista no seu protocolo, ou conforme o consentimento do participante;
- r) levar em conta, nas pesquisas realizadas em mulheres em idade fértil ou em mulheres grávidas, a avaliação de riscos e benefícios e as eventuais interferências sobre a fertilidade, a gravidez, o embrião ou o feto, o trabalho de parto, o puerpério, a lactação e o recém-nascido;
- s) considerar que as pesquisas em mulheres grávidas devem ser precedidas de pesquisas em mulheres fora do período gestacional, exceto quando a gravidez for o objeto fundamental da pesquisa;
- t) garantir, para mulheres que se declarem expressamente isentas de risco de gravidez, quer por não exercerem práticas sexuais ou por as exercerem de forma não reprodutiva, o direito de participarem de pesquisas sem o uso obrigatório de contraceptivos; e
- u) ser descontinuada somente após análise e manifestação, por parte do Sistema CEP/CONEP/CNS/MS que a aprovou, das razões dessa descontinuidade, a não ser em casos de justificada urgência em benefício de seus participantes. (BRASIL, 2012).

Ainda de acordo com Resolução MS/CNS nº 466 (BRASIL, 2012), às exigências anteriores, quando as pesquisas envolvendo seres humanos utilizarem metodologias experimentais na área biomédica, somam-se:

- a) estar fundamentadas na experimentação prévia, realizada em laboratórios, utilizando-se animais ou outros modelos experimentais e comprovação científica, quando pertinente;
- b) ter plenamente justificadas, quando for o caso, a utilização de placebo, em termos de não maleficência e de necessidade metodológica, sendo que os benefícios, riscos, dificuldades e efetividade de um novo método terapêutico devem ser testados, comparando-o com os melhores métodos profiláticos, diagnósticos e terapêuticos atuais. Isso não exclui o uso de placebo ou nenhum tratamento em estudos nos quais não existam métodos provados de profilaxia, diagnóstico ou tratamento;
- c) utilizar o material biológico e os dados obtidos na pesquisa exclusivamente para a finalidade prevista no seu protocolo, ou conforme o consentimento dado pelo participante da pesquisa; e
- d) assegurar a todos os participantes ao final do estudo, por parte do patrocinador, acesso gratuito e por tempo indeterminado, aos melhores métodos profiláticos, diagnósticos e terapêuticos que se demonstraram eficazes:
 - d.1) o acesso também será garantido no intervalo entre o término da participação individual e o final do estudo, podendo, nesse caso, esta garantia ser dada por meio de estudo de extensão, de acordo com análise devidamente justificada do médico assistente do participante.

Paralelamente, a Resolução MS/CNS nº 510 (BRASIL, 2016b) estabelece o regramento para pesquisas em Ciências Humanas e Sociais que utilizem dados obtidos junto aos participantes ou de informações que os identifiquem e, com isso, possam ocasionar riscos à vida cotidiana. O Capítulo II da Resolução é dedicado aos Princípios Éticos das Pesquisas em Ciências Humanas e Sociais e estabelece:

- I - reconhecimento da liberdade e autonomia de todos os envolvidos no processo de pesquisa, inclusive da liberdade científica e acadêmica;
- II - defesa dos direitos humanos e recusa do arbítrio e do autoritarismo nas relações que envolvem os processos de pesquisa;
- III - respeito aos valores culturais, sociais, morais e religiosos, bem como aos hábitos e costumes, dos participantes das pesquisas;
- IV - empenho na ampliação e consolidação da democracia por meio da socialização da produção de conhecimento resultante da pesquisa, inclusive em formato acessível ao grupo ou população que foi pesquisada;
- V – recusa de todas as formas de preconceito, incentivando o respeito à diversidade, à participação de indivíduos e grupos vulneráveis e discriminados e às diferenças dos processos de pesquisa;
- VI - garantia de assentimento ou consentimento dos participantes das pesquisas, esclarecidos sobre seu sentido e implicações;
- VII - garantia da confidencialidade das informações, da privacidade dos participantes e da proteção de sua identidade, inclusive do uso de sua imagem e voz;
- VIII - garantia da não utilização, por parte do pesquisador, das informações obtidas em pesquisa em prejuízo dos seus participantes;
- IX - compromisso de todos os envolvidos na pesquisa de não criar, manter ou ampliar as situações de risco ou vulnerabilidade para indivíduos e coletividades, nem acentuar o estigma, o preconceito ou a discriminação; e
- X - compromisso de propiciar assistência a eventuais danos materiais e imateriais, decorrentes da participação na pesquisa, conforme o caso sempre e enquanto necessário.

Tais aspectos, embora sejam direcionados ao contexto da pesquisa científica, podem ser generalizados para o tratamento ético dos dados armazenados em qualquer Ciclo de Vida dos Dados, visando antever e solucionar os possíveis riscos eminentes ao tratamento de dados de seres humanos.

Embora tais documentos tratem claramente de Princípios Éticos capazes de resguardar os direitos das pessoas, não estabelecem claramente como operacionalizá-los no que tange às formas de tratamento, uso e reuso dos dados nos quais são referenciadas, ou seja, os seus dados pessoais. Neste sentido, no Brasil, foi instituída a Lei Geral de Proteção de Dados Pessoais (BRASIL, 2018), que, por não se ater a questões éticas, foi abordada em seção destinada aos Princípios Legais.

Ainda no contexto da Pesquisa Científica, novamente no bojo do Movimento de Ciência Aberta e em decorrência da abertura de dados de pesquisa, houve um movimento internacional em torno dos Princípios FAIR (WILKINSON *et al.*, 2016) visando ao tratamento técnico desses dados, sobretudo no que concerne à capacidade das máquinas de maximizar a localização, o acesso, a interoperabilidade e o (re)uso de dados, que se aplica a dados, metadados e infraestruturas de

armazenamento de dados, incluindo os dados de pesquisa. Tais princípios, considerados nesta tese como Princípios Técnicos aplicados aos Dados, são tratados em seção própria.

Neste contexto, visando, de igual maneira, tratar os aspectos éticos relacionados aos dados de pesquisa, especificamente no que concerne à soberania dos dados indígenas, entendida por Stone e Calderon (2019, tradução nossa) como “[...] o direito dos povos indígenas e das nações de governar a coleta, propriedade e aplicação de seus próprios dados [...]”²⁸², sobre os quais somente os povos indígenas possuem condições de tomar decisões, considerando interesses e valores próprios, o *Global Indigenous Data Alliance* estabeleceu, em 2018, durante a *International Data Week and Research Data Alliance Plenary*, os *CARE Principles for Indigenous Data Governance*/Princípios CARE para a Governança de Dados Indígenas.

Destaca-se que CARE é um acrônimo para *Collective Benefit*/Benefício Coletivo, *Authority to Control*/Autoridade para Controlar, *Responsibility*/Responsabilidade e *Ethics*/Ética.

De acordo com Carroll *et al.* (2020b, tradução nossa), o estabelecimento dos Princípios CARE

[...] visava capacitar os povos indígenas, mudando o foco da governança de dados de consulta para relações baseadas em valores que promovem a participação indígena equitativa em processos de reutilização de dados, o que resultará em resultados mais equitativos, bem como preservando relacionamentos construídos na confiança e respeito.²⁸³

Desta forma, os Princípios CARE são claramente orientados a pessoas e, com isso, complementam os Princípios FAIR no processo de abertura de dados. Neste sentido, cumpre destacar que, como dados indígenas, são considerados pelos povos indígenas a língua, o conhecimento, os costumes, as tecnologias, os recursos naturais e os territórios, cuja expressão, em geral ocorre de forma oral e são considerados como essenciais para desenvolvimento e os direitos dos referidos povos.

Neste contexto, Carroll *et al.* (2020a) esclarece que os Princípios CARE são estabelecidos considerando o interesse dos povos indígenas em proteger seus dados e interesses e, ao mesmo tempo, possibilitar que os referidos dados possam ser

²⁸² “[...] the rights of Indigenous Peoples and nations to govern themselves and the data about them [...]”.

²⁸³ “[...] aimed to empower Indigenous Peoples, by shifting the focus of data governance from consultation to values-based relationships that promote equitable Indigenous participation in processes of data reuse, which will result in more equitable outcomes, as well as preserving relationships built on trust and respect.”

abertos e compartilhados. Assim, os princípios CARE estabelecem a governança sobre os dados dos povos indígenas, considerando os interesses científico e dos povos indígenas.

Os Princípios CARE são estruturados em 12 subprincípios a eles associados e que se relacionam entre si, cuja representação está disponível na Figura 25.

Figura 25 – Princípios CARE para Governança de Dados Indígenas



Fonte: Carroll *et al.* (2020a), traduzido por Vidotti, Torino, Coneglian (2021, p. 214).

O primeiro dos Princípios CARE, **Collective Benefit/Benefício Coletivo** estabelece que “Os ecossistemas de dados devem ser projetados e funcionar de forma a permitir que os povos indígenas se beneficiem dos dados.”²⁸⁴ (GLOBAL

²⁸⁴ “Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data.”

INDIGENOUS DATA ALLIANCE, 2019, tradução nossa). O princípio *Authority to Control*/Autoridade para Controlar consiste em:

Os direitos e interesses dos povos indígenas nos dados indígenas devem ser reconhecidos e sua autoridade para controlar esses dados deve ser autorizada. A governança de dados indígenas permite que os povos indígenas e órgãos governamentais determinem como os povos indígenas, bem como terras indígenas, territórios, recursos, conhecimentos e indicadores geográficos, são representados e identificados nos dados.²⁸⁵ (GLOBAL INDIGENOUS DATA ALLIANCE, 2019, tradução nossa).

O princípio *Responsability*/Responsabilidade estabelece que:

Aqueles que trabalham com dados indígenas têm a responsabilidade de compartilhar como esses dados são usados para apoiar a autodeterminação e benefício coletivo. A responsabilidade requer evidências significativas e abertamente disponíveis desses esforços e benefícios para os Povos Indígenas.²⁸⁶ (GLOBAL INDIGENOUS DATA ALLIANCE, 2019, tradução nossa).

E, no princípio designado *Ethics*/Ética, “Os direitos e o bem-estar dos povos indígenas devem ser a principal preocupação em todos os estágios do ciclo de vida dos dados e em todo o ecossistema de dados.”²⁸⁷ (GLOBAL INDIGENOUS DATA ALLIANCE, 2019, tradução nossa).

Visando favorecer a visualização, os Princípios e Subprincípios CARE são apresentados de forma sistematizada no Quadro 9.

Os Princípios CARE se constituíram para a governança dos dados indígenas. Contudo, além destes, nações, pessoas e comunidades são tidas por Carroll *et al.* (2020a) como atores nas sociedades globais contemporâneas. Deste modo, os autores entendem que os Princípios CARE abordam aspectos relevantes ao tratamento de dados de diferentes populações, sobre os quais haja a necessidade de elas mesmas manterem a primazia sobre os seus dados, considerando, por exemplo, privacidade, uso, reuso e gestão. Stone e Calderon (2019) já afirmavam que “Os

²⁸⁵ “Indigenous Peoples rights and interests in Indigenous data must be recognised and their authority to control such data be empowered. Indigenous data governance enables Indigenous Peoples and governing bodies to determine how Indigenous Peoples, as well as Indigenous lands, territories, resources, knowledges and geographical indicators, are represented and identified within data.”

²⁸⁶ “Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples’ self-determination and collective benefit. Accountability requires meaningful and openly available evidence of these efforts and the benefits accruing to Indigenous Peoples.”

²⁸⁷ “Indigenous Peoples’ rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem.”

princípios CARE certamente nos levam a considerar as pessoas refletidas nos dados e como nossas ações com eles podem impactá-las.”²⁸⁸.

Neste contexto, Vidotti, Torino e Coneglian (2021, p. 217) consideram que os Princípios CARE podem ser expandidos para o tratamento de outras comunidades que possuam igualmente a necessidade de exercer governança sobre seus dados, e destacam as “minorias sociais, como: quilombolas, comunidades ribeirinhas, de assentamentos sociais, de periferias e LGBTQ+.”

Quadro 9 – Princípios e Subprincípios CARE

Princípio	Subprincípio	Descrição
Benefício Coletivo	<i>C1: For inclusive development and innovation/Para desenvolvimento inclusivo e inovação</i>	Os governos e instituições devem apoiar o (re)uso de dados pelos povos indígenas e comunidades, com vistas à inovação, geração de valor e desenvolvimento local
	<i>C2: For improved governance and citizen engagement/Para melhorar a governança e o envolvimento do cidadão</i>	Os dados possibilitam o envolvimento entre governos, instituições e cidadãos, proporcionam transparência e auxiliam no planejamento, avaliação e tomada de decisões, além de oferecer informações de interesse dos povos indígenas
	<i>C3: For equitable outcomes/Para resultados equitativos</i>	Os dados indígenas estão relacionados aos seus valores e podem se estender à sociedade, de forma que todos os dados gerados devem beneficiar os povos indígenas de forma justa.
Autoridade para Controlar	<i>A1: Recognizing rights and interests/Reconhecimento de direitos e interesses</i>	Os povos indígenas devem ter reconhecidos os direitos e interesses nos seus dados e conhecimentos, o que se faz por meio do consentimento livre, prévio e manifesto na coleta, incluindo os usos, as políticas de dados e os protocolos utilizados na coleta
	<i>A2: Data for governance/Dados para governança</i>	Os povos indígenas devem exercer governança sobre seus dados, que devem estar a eles disponíveis e acessíveis
	<i>A3: Governance of data/Governança de dados</i>	Os povos indígenas podem desenvolver protocolos de governança e acesso aos seus dados, sobretudo aqueles concernentes ao conhecimento indígena

²⁸⁸ “The CARE principles certainly prompt us all to consider the people reflected in data and how our actions with it may impact on them.”

Princípio	Subprincípio	Descrição
Responsabilidade	R1: For positive relationships/Para relacionamentos positivos	O uso dos dados indígenas é possível quando pautado em relacionamentos de respeito, reciprocidade e confiança, de forma que o pesquisador seja o responsável por assegurar que os dados coletados, suas interpretações e uso garantam e respeitem a dignidade dos povos indígenas
	R2: <i>For expanding capability and capacity</i> /Para expandir a competência e a capacidade	O uso de dados indígenas requer responsabilidade recíproca, na competência em dados junto aos povos indígenas e o desenvolvimento de infraestruturas digitais que possibilitem a coleta, a gestão, a segurança e a governança de dados
	R3: <i>For indigenous languages and worldviews</i> /Para línguas e visões de mundo indígenas	Tais recursos devem gerar dados baseados nas linguagens, experiências, valores, princípios e visões de mundo dos povos indígenas
Ética	E1: <i>For minimizing harm and maximizing benefit</i> /Para minimizar danos e maximizar benefícios	Minimizar danos que podem ser oriundos de estigmas ou déficits relacionados aos povos indígenas, pautando a coleta, o tratamento e o uso em preceitos éticos, alinhados às estruturas éticas indígenas e com os direitos estabelecidos por meio da Declaração das Nações Unidas sobre os Direitos dos Povos Indígenas
	E2: <i>For justice</i> /Por justiça	Utilizar processos éticos de forma a tratar os desequilíbrios de poder e recursos, bem como a forma com que afetam os direitos indígenas e humanos
	E3: <i>For future use</i> /Para uso futuro	A governança de dados deve considerar o potencial uso futuro pautado em bases éticas, valores e princípios dos povos indígenas e, ainda, expressar nos metadados a proveniência, o propósito, os direitos de uso, incluindo as limitações, as obrigações no uso e o consentimento

Fonte: Adaptado de Global Indigenous Data Alliance (2019) e Vidotti, Torino e Coneglian (2021).

Considerando a necessidade do tratamento adequado dos dados em todas as etapas da pesquisa, bem como em todo o Ciclo de Vida dos Dados, Vidotti, Torino e Coneglian (2021) consideram que devem ser adotadas práticas designadas como #SejaJUSTOeCUIDADOSO, que considera que os dados precisam atender a

Princípios Técnicos e Éticos. As práticas de #SejaJUSTOeCUIDADOSO, segundo os autores “[...] devem estar presentes no fazer diário daqueles que atuam, direta ou indiretamente, com processos relacionados à coleta, à análise, ao tratamento, ao armazenamento e à disponibilização de dados, sobretudo, quando envolvem seres humanos.” (VIDOTTI; TORINO; CONEGLIAN, 2021, p. 219).

No que tange aos Princípios Éticos no tratamento de Dados de Pesquisa, Vidotti, Torino e Coneglian (2021, p. 219), entendem como CUIDADOSO:

- respeitar os princípios CARE, considerando a hegemonia das comunidades específicas, suas visões de mundo, soberania e governança sobre os seus dados, com:
- foco no desenvolvimento inclusivo;
- estabelecimento de relações equitativas, de confiança, reciprocidade e respeito;
- atendimento aos preceitos éticos e legais na coleta, tratamento, armazenamento e disponibilização dos dados;
- identificação dos dados;
- proteção aos direitos, interesses, valores e cultura;
- participação na governança e controle sobre os seus dados;
- melhoria na representação dos dados;
- capacitação para uso dos dados.

Diante do exposto, as afirmações de Vidotti, Torino e Coneglian (2021) no que concernem à possibilidade de aplicação dos Princípios CARE a outras comunidades e outros tipos de dados, além dos dados de pesquisa, são reafirmados nesta tese.

Assim, na Arquitetura de Dados no contexto da Ciência da Informação, os Princípios Éticos aplicados aos dados, desde a etapa de planejamento, até a coleta, tratamento, armazenamento e disponibilização, quando houver, devem ser rigorosamente atendidos. Tal recomendação visa resguardar inicialmente e de forma veemente os indivíduos referenciados nos dados, mas ao mesmo tempo, o detentor do Ciclo de Vida dos Dados.

Destaca-se que, nessa seção, foram utilizadas algumas bases teóricas para a discussão apresentada, em caráter exploratório e vinculada ao contexto brasileiro, no qual a presente tese se insere. Contudo, a discussão pode ser aprofundada e realizada em outros contextos, para envolver legislações de outras jurisdições e ainda comparar marcos legais relacionados aos Princípios Éticos aplicados aos Dados em suas diferentes formas.

O que se pretende destacar é a relevância dos Princípios Éticos aplicados aos Dados como elemento da Arquitetura de Dados com vistas a contribuir para a discussão dos referidos princípios para o tratamento de dados de seres humanos

referenciados nos dados, independente da finalidade de coleta e do tipo de dados gerados, possibilitando ao indivíduo o tratamento isonômico, respeitoso e cuidadoso em atenção aos direitos a eles estabelecidos, sua autonomia, dignidade e interesses.

4.4.1.5.2 Princípios legais aplicados aos dados

Na Arquitetura de Dados, os Princípios Legais Aplicados aos Dados, assim como os Princípios Éticos, estão fortemente atrelados à coleta, ao tratamento, ao armazenamento e à disponibilização de dados, considerando os diferentes contextos, tecnologias e finalidades.

Na seção dedicada aos Princípios Éticos aplicados aos Dados, alguns marcos regulatórios foram utilizados na apresentação da discussão, com destaque para a Lei de Acesso à Informação (BRASIL, 2011) e para a Política de Dados Abertos do Poder Executivo Federal (BRASIL, 2016a), além das Resoluções do Ministério da Saúde relacionadas aos aspectos éticos de pesquisa com seres humanos (BRASIL, 2012, 2016b).

Nessa seção, busca-se discutir os Princípios Legais Aplicados aos Dados, que estão de certa forma atrelados aos Princípios Éticos, mas que necessitam de abordagem distinta para acrescentar à discussão elementos específicos. Destaca-se que o tratamento ético dos dados contribui para os Princípios Legais, contudo, não é suficiente para tratá-los.

Novamente considerando que cada jurisdição possui um ordenamento jurídico, e que a presente tese está inserida no cenário brasileiro, são trazidos para a análise marcos legais brasileiros. Destaca-se assim, que tal discussão pode ser ampliada para outras jurisdições, de forma comparativa ou a partir de outras abordagens que enfoquem os princípios legais aplicados aos dados.

No Brasil, o tratamento de dados pessoais tem respaldo na Lei Geral de Proteção de Dados Pessoais (LGPD), instituída pela Lei nº 13.709, de 14 de agosto de 2018 (BRASIL, 2018), com alterações dadas pela Lei nº 13.853, de 8 de julho de 2019 (BRASIL, 2019). A referida LGPD estabelece no Art. 1º que

Esta Lei dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. (BRASIL, 2018).

Também estabelece que as normas da referida Lei (BRASIL, 2018) “[...] são de interesse nacional e devem ser observadas pela União, Estados, Distrito Federal e Municípios.” Outrossim, especifica a aplicação no “Art. 3º Esta Lei aplica-se a qualquer operação de tratamento realizada por pessoa natural ou por pessoa jurídica de direito público ou privado, independentemente do meio, do país de sua sede ou do país onde estejam localizados os dados [...]”, tendo como condições, nos incisos I a III, que o tratamento dos dados ocorra em território brasileiro, o tratamento objetivo oferta e/ou fornecimento de bens ou serviços, e a coleta dos dados pessoais ocorra em território brasileiro – neste caso, considerados quando o titular nele se encontre.

Cumpre esclarecer, neste ponto, que os atores destacados pela LGPD possuem designações divergentes das apresentadas por Sant’Ana (2013, 2016) no Ciclo de Vida de Dados, a saber: detentor, intermediário, referenciado e usuário. A LGPD os designa como:

[...]

V - titular: pessoa natural a quem se referem os dados pessoais que são objeto de tratamento;

VI - controlador: pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais;

VII - operador: pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador;

VIII - encarregado: pessoa indicada pelo controlador e operador para atuar como canal de comunicação entre o controlador, os titulares dos dados e a Autoridade Nacional de Proteção de Dados (ANPD);

IX - agentes de tratamento: o controlador e o operador;

[...] (BRASIL, 2018).

Visando contribuir com o relacionamento entre as designações adotadas por Sant’Ana (2013, 2016) no Ciclo de Vida dos Dados e Brasil (2018) na LGPD, considerando as atribuições apresentadas pela LGPD, foi elaborado o Quadro 10.

É importante, ainda, compreender como o tratamento é considerado na LGPD (BRASIL, 2018)

[...]

X - tratamento: toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração;

[...]

Quadro 10 – Relacionamento entre os Atores envolvidos no Tratamento dos Dados

Atribuição	LGPD Brasil (2018)	Ciclo de Vida dos Dados Sant'Ana (2013, 2016)
Pessoa natural a quem se referem os dados pessoais que são objeto de tratamento	Titular	Referenciado
Pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais	Controlador	Detentor
Pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador	Operador	Intermediário
Pessoa indicada pelo controlador e operador para atuar como canal de comunicação entre o controlador, os titulares dos dados e a Autoridade Nacional de Proteção de Dados (ANPD)	Encarregado	Detentor e Intermediário
Agentes de tratamento	Controlador e Operador	Detentor e Intermediário

Fonte: Baseado em Brasil (2018) e Torino, Vidotti e Sant'Ana (2019).

Isso posto, cumpre ressaltar que, em grande parte, a LGPD trata das formas de tratamento dos dados pessoais, bem como das formas de consentimento fornecido pelo titular e das sanções que podem ser aplicadas ao controlador, operador e encarregado no caso de descumprimento das determinações legais.

Nesse sentido, é importante ter em mente que todo dado digital é criado ou coletado no contexto de um Ciclo de Vida dos Dados (CVD), como compreendido por Sant'Ana (2013, 2016). Assim, o indivíduo (Titular/Referenciado) informa ou registra seus dados na fase de coleta de determinado CVD e, a partir disso, toda a gestão desse dado passa a ser responsabilidade de um terceiro, que, de acordo com a LGPD (BRASIL, 2018) pode ser uma pessoa física ou jurídica responsável pelas decisões referentes ao tratamento dos dados pessoais ou pelo tratamento propriamente dito (Controlador, Operador, Encarregado/Detentor, Intermediário).

Nesse contexto, a LGPD (BRASIL, 2018) determina que “Toda pessoa natural tem assegurada a titularidade de seus dados pessoais e garantidos os direitos

fundamentais de liberdade, de intimidade e de privacidade, nos termos desta Lei.” e estabelece, no Art. 18, que

O titular dos dados pessoais tem direito a obter do controlador, em relação aos dados do titular por ele tratados, a qualquer momento e mediante requisição:

I - confirmação da existência de tratamento;

II - acesso aos dados;

III - correção de dados incompletos, inexatos ou desatualizados;

IV - anonimização, bloqueio ou eliminação de dados desnecessários, excessivos ou tratados em desconformidade com o disposto nesta Lei;

V - portabilidade dos dados a outro fornecedor de serviço ou produto, mediante requisição expressa, de acordo com a regulamentação da autoridade nacional, observados os segredos comercial e industrial; (Redação dada pela Lei nº 13.853, de 2019)

VI - eliminação dos dados pessoais tratados com o consentimento do titular, exceto nas hipóteses previstas no art. 16 desta Lei;

VII - informação das entidades públicas e privadas com as quais o controlador realizou uso compartilhado de dados;

VIII - informação sobre a possibilidade de não fornecer consentimento e sobre as consequências da negativa;

IX - revogação do consentimento, nos termos do § 5º do art. 8º desta Lei.

Destaca-se que, embora o Titular/Referenciado tenha a prerrogativa inicial da tomada de decisão acerca do fornecimento dos seus dados pessoais e, posteriormente possa requerer informações acerca dos dados armazenados no CVD, a ele não é possibilitada a verificação física do espaço de armazenamento e da camada de compartilhamento dos dados no CVD, cabendo essa atividade apenas ao Controlador, Operador, Encarregado/Detentor, Intermediário, que é quem são, na prática e de fato, os “titulares” dos dados armazenados.

Nesse aspecto, ainda, é importante mencionar que todas as interações realizadas por usuários em sistemas de informação e/ou ambientes informacionais digitais podem ser coletadas e, nesse sentido, requerer o consentimento do Titular/Referenciado sobretudo da forma que a LGPD determina, especificando as finalidades, pode ser dificultoso. De igual maneira, deve-se ter em mente que esse processo de coleta não deixará de ser realizado. Embora, em geral, tais coletas não atinjam o que a LGPD entende como dados pessoais²⁸⁹, por meio delas é possível realizar cruzamentos de dados com outras bases que gerenciem dados pessoais e, com isso, possibilitar a identificação da pessoa.

²⁸⁹ “[...] informação relacionada a pessoa natural identificada ou identificável” (BRASIL, 2018).

Destaca-se ainda que a LGPD não considera como dados pessoais aqueles que forem anonimizados²⁹⁰, excetuando-se os casos em que a anonimização possa ser revertida.

Ainda no bojo da LGPD, é importante destacar o Art. 46, que determina a competência pela adoção de medidas de segurança dos dados pessoais aos responsáveis pelo tratamento dos mesmos:

Os agentes de tratamento devem adotar medidas de segurança, técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou qualquer forma de tratamento inadequado ou ilícito. (BRASIL, 2018).

E a eles – Controladores, Operadores/Detentor, Intermediário – também compete, conforme Art. 50, a governança e adoção de boas práticas relativas ao tratamento dos dados pessoais:

Os controladores e operadores, no âmbito de suas competências, pelo tratamento de dados pessoais, individualmente ou por meio de associações, poderão formular regras de boas práticas e de governança que estabeleçam as condições de organização, o regime de funcionamento, os procedimentos, incluindo reclamações e petições de titulares, as normas de segurança, os padrões técnicos, as obrigações específicas para os diversos envolvidos no tratamento, as ações educativas, os mecanismos internos de supervisão e de mitigação de riscos e outros aspectos relacionados ao tratamento de dados pessoais. (BRASIL, 2018).

Nesse sentido, considerando a Arquitetura de Dados no contexto da Ciência da Informação, é necessário que, no momento do planejamento do Ciclo de Vida dos Dados, sejam considerados aspectos relacionados ao tratamento dos dados, sobretudo dos dados pessoais e dos dados sensíveis, considerando os impactos que o tratamento pode acarretar à(s) pessoa(s) referenciada(s) nos dados, pois em caso de tratamento inadequado, os impactos negativos podem ser irreparáveis.

Essa é uma questão relevante no contexto de uso intensivo de tecnologias e aparatos tecnológicos aos quais, muitas vezes, as pessoas – consideradas pela LGPD como titulares e no Ciclo de Vida dos Dados como referenciadas – não se atentam quando do fornecimento de dados que realizam cotidianamente no uso de tecnologias. Desde as mais simples, como o acionar de um portão eletrônico vinculado a um sistema de informação, ao uso de cartões magnéticos, as chamadas telefônicas e

²⁹⁰ “[...] utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo” (BRASIL, 2018).

tantas outras atividades cotidianas, inúmeras ações deixam dados de pessoas, seus comportamentos e atividades registrados em sistemas de informação.

Apreende-se que, da forma com que os atores são considerados pela LGPD e pelo CVD, há duas perspectivas complementares para o tratamento dos dados. A primeira perspectiva, da LGPD (BRASIL, 2018), se refere às formas de tratamento dos dados entendendo o indivíduo como “titular” dos seus dados pessoais, que uma vez armazenados em uma base de dados passam a ser geridos pelo “controlador”, responsável sobre a tomada de decisões sobre os dados e que pode transferir o tratamento propriamente dito para o “operador”. Desta forma, tem-se uma abordagem na qual o indivíduo é colocado como titular dos seus dados pessoais, o que é verdadeiro e, portanto, cabe a ele determinar, em primeira instância, que dados ele quer registrar em sistemas de informação.

Porém, à medida que o uso de tecnologias e aparatos digitais é parte do cotidiano dos indivíduos, muitas vezes o registro de dados em sistemas de informação não é uma decisão, mas uma condição, como, por exemplo, ao abrir o navegador de internet e acessar qualquer site, que imediatamente registra o IP de acesso. O mesmo ocorre quando do acesso a um prédio comercial para uma consulta médica e é necessário realizar um cadastro, registrar a imagem e utilizar um cartão magnético para a entrada e saída por uma catraca de acesso. Como essas, muitas atividades cotidianas são realizadas e deixam dados armazenados em sistemas de informação.

Já a segunda perspectiva, no Ciclo de Vida dos Dados (SANT’ANA, 2013, 2016), o indivíduo é tido como “referenciado” nos dados armazenados em um sistema de informação, cujo “detentor” é o responsável pela tomada de decisão sobre os dados e a quem compete o tratamento ou a delegação de competência ao “intermediário”. Essa designação tem como abordagem o CVD considerando que os dados, uma vez armazenados, possuem um detentor, e os indivíduos são neles referenciados.

Essa abordagem, embora dê ênfase ao sistema de informação, é claramente mais precisa, uma vez que, ainda que o indivíduo tenha a titularidade sobre seus dados pessoais, quando ele os armazena em uma base de dados sobre a qual não tem controle, o que faz é uma espécie de “tutela” dos dados para o “detentor” que passa a exercer controle sobre os dados armazenados.

Outra questão relevante é que, embora a LGPD estabeleça que o “titular” tem o direito de requerer alterações ou mesmo a exclusão dos seus dados de determinada base de dados, esse direito se aplica ao momento futuro ao do requerimento, visto

que, no contexto digital, podem ter ocorrido diversas ações que podem manter esses dados inalterados, a exemplo de *backup* distribuído, preservação digital distribuída, compartilhamento previamente autorizado, sobre os quais pode não haver mais condições de interferência.

Nesse contexto, Sant'Ana (2013, 2016) explicita que em determinado CVD, os dados armazenados, uma vez disponibilizados pela fase de coleta, podem ser coletados por outro(s) CVD sobre o(s) qual(is) não se tem gerência, uma vez que cada CVD possui seu “detentor”.

Tendo em vista que questões como essas necessitam de ampla análise, e sobre a qual a presente tese não tem pretensão de se debruçar, essa discussão preliminar pretende apenas destacar a relevância da temática para que ela seja foco de outras pesquisas científicas.

Destaca-se, contudo, no que tange à Arquitetura de Dados, que as legislações vigentes na jurisdição na qual o planejamento está sendo executado, bem como as que com elas se relacionam, são imprescindíveis para o tratamento adequado dos dados.

Ainda no que tange aos Aspectos Legais relacionados aos Dados, quando se trata de dados descritivos utilizados na representação, os Aspectos Técnicos relacionados aos dados, por meio das Boas Práticas para Dados na Web (LÓSCIO; BURLE; CALEGARI, 2017) e dos Princípios FAIR (GOFAIR, 2021; FORCE21, 2021), enfatizam a necessidade do uso de licenças. Neste sentido, tais licenças se referem à expressão dos direitos, sobretudo os direitos de uso sobre os dados disponíveis por meio das representações, preferencialmente utilizando estruturas de metadados.

Tais licenças fazem parte dos Princípios Legais aplicados aos Dados, visto que estão diretamente ligados a legislações, a exemplo, no Brasil, da Lei de Propriedade Industrial (BRASIL 1996), da Lei de Proteção da Propriedade Intelectual de Programa de Computador (BRASIL, 1998a), e da Lei de Direitos Autorais (BRASIL, 1998b), sendo essa última de interesse desta tese.

No que tange à Lei de Direitos Autorais brasileira (BRASIL, 1998b), ficam estabelecidas, no Art. 7, as obras intelectuais protegidas: “São obras intelectuais protegidas as criações do espírito, expressas por qualquer meio ou fixadas em qualquer suporte, tangível ou intangível, conhecido ou que se invente no futuro [...]” seguida de algumas exemplificações disponíveis à época, como: textos de obras literárias, artísticas ou científicas; conferências, alocações, sermões; obras dramáticas

e coreográficas; composições musicais; obras audiovisuais; projetos arquitetônicos; adaptações, traduções e outras transformações de obras originais, apresentadas como criação intelectual nova; programas de computador; bases de dados, desde que se constitua como uma criação intelectual nova.

Destaca-se que, no escopo das obras protegidas, há criações intelectuais das mais diversas finalidades e sobre as quais são possíveis diferentes formas de aproveitamento. Tal fato, somado à característica de obsolescência da própria lei e dos suportes e tecnologias por ela apontados, dificulta que haja equidade nos direitos de formas de tratamento estabelecidas, sobretudo quando interpretadas na atualidade. Apesar disso, enquanto a referida Lei estiver vigente, é aplicável no contexto brasileiro.

Nesse sentido, afirma Wachowicz (2014, p. 10) que “A Revolução Tecnológica com suas novas tecnologias ocasionou a criação de novos bens intelectuais, os denominados bens informáticos. A literatura contemporânea acerca do Direito Intelectual relativamente aos bens informáticos é escassa.” Ao que o autor complementa

Há uma simetria no entendimento entre doutrinadores e juristas que tratam estes novos bens informáticos como uma complementação linear da perspectiva inicial de proteção dos ramos de Direito Autoral e Industrial, determinando seu enquadramento jurídico aos princípios previamente existentes. (WACHOWICZ, 2014, p. 11).

Da Lei de Direitos Autorais brasileira, é importante destacar a natureza dúplice do direito, que protege, por um lado, o autor, por meio dos Direitos Morais e, por outro, a obra, por meio dos Direitos Patrimoniais, que possuem formas e tempo de fruição distintas.

[...] a partir do século XIX, passou a doutrina a analisar o bem intelectual em duas ordens distintas: como direitos patrimoniais passíveis de alienação ligados às características econômicas e pecuniárias, que consistem na faculdade de fruir, de modo exclusivo, todas as vantagens materiais que a reprodução da obra possa oferecer; e como direitos morais do autor, inerentes à sua personalidade, direitos inalienáveis, ligados à paternidade da obra, nomeação ou alteração. (WACHOWICZ, 2014, p. 8)

Assim, no contexto da discussão desta seção, sobretudo no que tange ao uso de licenças nos metadados descritivos, têm lugar os Direitos Patrimoniais e as formas de transferência desses direitos, estabelecidas no Art. 49, considerando as limitações apresentadas nos seus incisos.

Art. 49. Os direitos de autor poderão ser total ou parcialmente transferidos a terceiros, por ele ou por seus sucessores, a título universal ou singular, pessoalmente ou por meio de representantes com poderes especiais, por meio de licenciamento, concessão, cessão ou por outros meios admitidos em Direito [...]” (BRASIL, 1998b).

Nesse sentido, de acordo com a Lei de Direitos Autorais brasileira (BRASIL, 1998b), para que haja a adoção de licenças como uma das formas de transferência dos direitos, faça-se por escrito. Além disso, a referida Lei estabelece, no Art. 29, que a utilização da obra depende da autorização prévia e expressa do autor.

Em paralelo, no início dos anos 2000 tem início o Movimento de Acesso Aberto que atinge sobretudo os resultados de pesquisa materializados em objetos de comunicação científica, seguido do Movimento Transparência e Dados Abertos Governamentais e, mais recentemente, o de Ciência Aberta e, com ele, os Dados Abertos de Pesquisa.

É nesse contexto que as licenças ganham destaque, quer seja a partir da sua compreensão como licenças de Direitos Autorais ou como licenças de uso. Destaca-se que as licenças de direitos autorais dependem de um sistema de proteção autoral, uma vez que apenas as obras protegidas por direitos autorais podem ser licenciadas. Com isso, as licenças possibilitam que o(s) titular(es) dos direitos autorais possam determinar formas de uso autorizadas pelo público da obra, reservando todos os demais usos para si.

No que tange às licenças, no campo científico, as que mais se popularizaram foram as disponibilizadas pela *Creative Commons* (CC), “[...] uma organização mundial sem fins lucrativos que visa promover o compartilhamento do conhecimento e da criatividade como forma de transformação social.” (CREATIVE COMMONS BRASIL, 2022).

A *Creative Commons* disponibiliza um conjunto de seis licenças que podem ser aplicadas a qualquer obra que esteja protegida pelos direitos autorais. Dessa forma, cumpre esclarecer que as referidas licenças não se aplicam a obras que não estejam protegidas por Direitos Autorais, a exemplo de obras que entraram em domínio público, sobre as quais não há mais proteção dos Direitos Patrimoniais, além de patentes, marcas, desenho industrial, *software* ou qualquer outra forma de proteção não abarcada pela Lei de Direitos Autorais.

Destaca-se ainda que, em 2013, a *Creative Commons* disponibilizou a última versão do seu conjunto de licenças, designada como “4.0 Internacional”.

Essas licenças são as mais atualizadas que o CC oferece, sendo recomendadas no lugar de todas as versões anteriores. [...] A 4.0 foi elaborada para ser válida internacionalmente, e traduções oficiais têm sido disponibilizadas após a sua publicação. (CREATIVE COMMONS BRASIL, 2022).

Nesse sentido, é importante esclarecer que, quando do licenciamento, o licenciante, ou seja, o titular dos Direitos Autorais Patrimoniais, que nem sempre é o autor, deve escolher os direitos de uso que pretende estabelecer para a obra e, além disso, a versão na qual deseja licenciar, o que deve ser incluído junto à licença na obra licenciada, visando dar conhecimento ao usuário.

Essa é uma informação importante, visto que as licenças, embora se materializem em uma marca imagética, um breve resumo e um link, são de fato instrumentos jurídicos. Assim, até a versão 3.0, as licenças se aplicavam a jurisdições específicas, uma vez que “[...] são baseadas no conjunto internacional de licenças, mas foram modificadas para refletir nuances locais na expressão de termos e condições jurídicos, redação de protocolos e linguagem.” (CREATIVE COMMONS BRASIL, 2022).

Assim, a adoção da versão 4.0 Internacional é altamente recomendável, sobretudo quando a visibilidade da obra ultrapassa as fronteiras de determinada jurisdição. Nesse sentido, “[...] as licenças internacionais são essencialmente neutras quanto à jurisdição, permanecendo eficazes a nível mundial. A natureza neutra das licenças internacionais agrada a muitas pessoas e organizações, especialmente para o uso em projetos que transcendem fronteiras políticas.” (CREATIVE COMMONS BRASIL, 2022).

A forma de licenciamento pela *Creative Commons* é simples. Basta responder a duas perguntas disponíveis em um seletor de licença²⁹¹, e a licença escolhida será informada, bem como será disponibilizada a marca que deve ser utilizada e, caso necessite, o código legível por máquina, que pode ser inserido, por exemplo, em uma página web.

Contudo, é relevante que o licenciante tenha clareza de quais direitos quer licenciar, para que faça o uso adequado da licença. De igual maneira, é importante

²⁹¹ Disponível em: <https://creativecommons.org/choose/>.

que o licenciante informe, na obra, se há limites para a licença estabelecida. “A CC alerta potenciais licenciantes de que eles precisam ser os titulares de todos os direitos necessários antes de aplicar uma licença CC a uma obra.” (CREATIVE COMMONS BRASIL, 2022).

Destaca-se que a licença adotada deve ser registrada na obra, uma vez que é ela o objeto do licenciamento, e quando inserida em um ambiente informacional que disponibilize metadados descritivos, é imprescindível que a licença seja registrada em um campo de metadado específico, o que possibilita que o conteúdo seja facilmente identificado por máquinas como licenciado e, com isso possa ser recuperado em buscas específicas, a exemplo do filtro de busca por obras licenciadas do Google.

No que tange à aplicação da licença, a *Creative Commons* esclarece:

Para material online: Selecione a licença apropriada para o seu material no seletor de licença CC e, em seguida, siga as instruções para incluir o código HTML. O código automaticamente gerará um botão de licença e uma declaração de que seu material está licenciado sob uma licença CC. Caso você só esteja licenciando parte de uma obra (por exemplo, se você houver criado um vídeo sob uma licença CC, mas estiver usando uma música sob uma licença diferente), não se esqueça de marcar claramente quais as partes estão sob a licença CC e quais não. O código HTML também incluirá metadados que permitem que o material seja descoberto por mecanismos de busca sensíveis o Creative Commons.

Para material off-line: Identifique qual licença deseja aplicar à sua obra e: ou (a) marque sua obra como uma declaração como “Esta obra está licenciada com uma Licença Creative Commons [inserir descrição]”. Para ver uma cópia da licença, visite [inserir url]”; ou (b) insira os botões de licença aplicáveis com a mesma instrução e o link URL.

Para plataformas de terceiros: Muitas plataformas de mídia, como Flickr, YouTube e SoundCloud, possuem ferramentas destinadas a obras em Creative Commons, permitindo que os usuários marquem seu material com uma licença CC por meio de suas configurações de conta. A vantagem de usar esta funcionalidade é que ela permite que outras pessoas encontrem o seu conteúdo ao pesquisar por materiais licenciados sob CC nessas plataformas. Se a plataforma na qual você está publicando o seu conteúdo não suporta licenciamento em CC, você ainda pode identificá-lo como licenciado na descrição em texto dele. (CREATIVE COMMONS BRASIL, 2022, grifo nosso).

No campo científico, destaca-se que ambientes informacionais como bibliotecas digitais, repositórios digitais e periódicos científicos digitais possibilitam a inserção da licença conforme especificada por *Creative Commons* no caso de plataformas de terceiros.

Destaca-se ainda da Lei de Direitos autorais brasileira (BRASIL, 1998b), no Art. 7º, a menção à proteção de bases de dados: “XIII - as coletâneas ou compilações,

antologias, enciclopédias, dicionários, bases de dados e outras obras, que, por sua seleção, organização ou disposição de seu conteúdo, constituam uma criação intelectual.” Destaca-se do *caput* do referido artigo e do inciso citado que as bases de dados podem ser objeto de proteção pelos Direitos Autorais, desde que entendidas como uma criação do intelecto humano, compreensão que encontra lugar no Art. 11 que estabelece “Autor é a pessoa física criadora de obra literária, artística ou científica.”, sendo admitida a coautoria.

Wachowicz (2014, p. 16-17) afirma: “Ressalte-se que a proteção concedida no dispositivo citado não contempla os dados ou materiais em si mesmos e se entende sem prejuízo de quaisquer direitos autorais que subsistam a respeito dos dados ou materiais contidos nas obras.”

Tais exposições justificam o tratamento das bases de dados como direitos *sui generis*, ou seja, considerando certas especificidades. A *Creative Commons Brasil* (2022) esclarece que

Direitos *sui generis* de banco de dados conferem aos fabricantes de banco de dados o direito de proibir a extração e reutilização de uma parte substancial de uma base de dados. Os direitos são concedidos aos fabricantes de banco de dados que fazem um investimento substancial de tempo e recursos para criar o banco de dados.

Ao que complementa: “Direitos *sui generis* de banco de dados impedem a cópia e reutilização de partes substanciais de um banco de dados (incluindo a extração frequente de partes não substanciais).” (CREATIVE COMMONS BRASIL, 2022).

Nesse aspecto, Wachowicz (2014) alerta que é necessário determinar quando as bases de dados se constituem em obras intelectuais passíveis de proteção por Direitos Autorais, e quando incorporam obras protegidas, o que claramente requer tratamentos diferentes considerando a legislação vigente. De forma que Wachowicz (2014), entendendo a necessidade de delimitar com rigor o que é armazenado, apresenta três tipos de bases de dados:

- a) **composta de informações primárias**, ou seja, dados oriundos de tabelas de estoque, de dados históricos, numéricos, dentre outros;
- b) **composta de informações ditas secundárias**, como catálogos, index, lista, rol de informações que surgem originariamente de determinado local, inclusive sinopses de acontecimentos, de jornais, de resumos, dentre outros; e
- c) **composta de textos originais** de petições ou decisões judiciais, manuais de técnicos, apostilas de ensino, atlas, enciclopédias, artigos de jornais, coletâneas musicais, gravações de palestras,

imagens fotográficas ou vídeos, dentre outros. (WACHOWICZ, 2014, p. 15, grifos do autor).

Com isso, esclarece que as bases de dados são, em geral, dependentes de programas de computador e que, a esse, ainda que haja um processo de seleção inovador, não é possível requerer proteção por Direitos Autorais²⁹². Ao passo que “[...] a base de dados desenvolvida pelo programa de computador poderá ambicionar a tutela do direito do autor, com fundamento na criatividade do conjunto, com fulcro no inciso XIII, do artigo 7.º, da Lei n.º 9.610/98 [...]” (WACHOWICZ, 2014, p. 15).

Diante do exposto, caso à base de dados se apliquem os direitos *sui generis*, é possível licenciá-la sob uma das licenças *Creative Commons*: “Na versão 4.0, os direitos *sui generis* sobre bases de dados aplicáveis são licenciados sob as mesmas condições de licença que os direitos autorais.” (CREATIVE COMMONS BRASIL, 2022). Com destaque a uma recomendação da *Creative Commons* (2022): “CC não recomenda o uso de suas licenças NãoComercial (NC) ou SemDerivações (ND) em bancos de dados destinados ao uso acadêmico ou científico.” A recomendação é seguida pela indicação de dedicação ao domínio público:

Além de nossas licenças, o CC0 pode ser usada em bancos de dados para maximizar a reutilização de bases de dados. Quando aplicado, o efeito é o de renunciar a todos os direitos patrimoniais de autor e direitos conexos no banco de dados e seu conteúdo. Em certos domínios, tais como ciência e administração pública, há importantes razões para considerar o uso CC0. Dispensar os direitos de autor e direitos conexos elimina toda a incerteza para os potenciais utilizadores, encorajando a reutilização máxima e o compartilhamento de informações. (CREATIVE COMMONS BRASIL, 2022).

A Dedicção ao Domínio Público é uma forma de licenciamento que pode ser utilizada pelo autor ainda em vida, considerando que o prazo para que uma obra protegida por Direitos Autorais entre em Domínio Público por força da Lei, no Brasil, é de 70 anos a contar de 1 de janeiro do ano seguinte ao falecimento do autor. Na Dedicção ao Domínio Público pela *Creative Commons*, fica estabelecido de forma resumida que “A pessoa que associou um trabalho a este [...] [instrumento] dedicou o trabalho ao domínio público, renunciando a todos os seus direitos sob as leis de direito de autor e/ou de direitos conexos referentes ao trabalho, em todo o mundo, na medida permitida por lei.” (CREATIVE COMMONS, 2022).

²⁹² Destaca-se, aqui, que a proteção de software, no Brasil, é estabelecida por legislação específica (BRASIL, 1996a).

No que tange ao licenciamento de bancos de dados, outra forma de licenciamento possível é apresentada pela *Open Knowledge Foundation*, cuja missão é “[...] criar um mundo mais aberto – um mundo onde todas as informações não pessoais são abertas, gratuitas para todos usarem, desenvolverem e compartilharem; e criadores e inovadores são justamente reconhecidos e recompensados.”²⁹³ (OPEN KNOWLEDGE FOUNDATION, 2022a).

A *Open Knowledge Foundation* disponibiliza as licenças *Open Data Commons*, claramente aplicáveis apenas a bancos de dados, compostas por duas opções de licenças, e a *Open Data Commons Public Domain Dedication and License* (PDDL), uma licença que possibilita a Dedicção ao Domínio Público (OPEN KNOWLEDGE FOUNDATION, 2022b). O licenciamento e as especificações para adoção das licenças são similares aos apresentados ao longo desta seção.

Destaca-se novamente que tanto a proteção aos Direitos Autorais quanto o licenciamento se aplicam a bases de dados e não aos dados propriamente ditos.

Retomando o contexto científico, a obra *Abertura e compartilhamento de dados de pesquisa subjacentes a artigos científicos: questões do direito autoral* (GUANAES, 2020) apresenta discussões relevantes no que concerne aos direitos autorais aplicados aos dados de pesquisa. Na obra, o autor afirma:

[...] pudemos observar a provável emergência de um novo regime de direito autoral, ligado ao processo de publicação, abertura e compartilhamento de dados subjacentes a um artigo, disponibilizados em repositórios de dados. Durante esse processo podem ocorrer reivindicações sobre direito autoral de conjuntos de dados. Isto porque se o termo “dados” ainda não adquiriu uma definição de consenso, é menos trivial ainda estabelecer a propriedade de dados ou a existência de direitos autorais para conjuntos de dados que, para obterem tal proteção, têm que demonstrar um mínimo de originalidade, para não serem definidos como fatos, que não recebem proteção. (GUANAES, 2020, p. 199).

Ao que o autor, ao se referir aos dados de pesquisa, acrescenta “[...] para efetivar todo o potencial de compartilhamento de dados, ainda é preciso superar barreiras legais estruturais.” (GUANAES, 2020, p. 200).

Nesse contexto, tem-se utilizado como prática a disponibilização de dados de pesquisa licenciados pela *Creative Commons* de forma análoga ao que se faz com os demais objetos de comunicação científica, como artigos. Contudo, é imprescindível,

²⁹³ “[...] create a more open world – a world where all non-personal information is open, free for everyone to use, build on and share; and creators and innovators are fairly recognised and rewarded.”

antes do licenciamento, identificar se o que se objetiva licenciar é considerado passível de proteção por direitos autorais, conforme estabelece a lei brasileira.

No contexto da disponibilização de dados, de acordo com FORCE21 (2021), os dados não licenciados, ainda que disponíveis, possuem consumo evitado, sobretudo por grandes empresas, em função dos riscos legais. Nesse sentido, alertam que há situações em que os dados não devem estar em acesso aberto, sobretudo quando há questões de privacidade e propriedade intelectual envolvidas. Ao contrário disso, reforçam que o licenciamento adequado dos dados é uma das questões primordiais para a disponibilização de dados FAIR (FORCE21, 2021).

As licenças de objeto de dados e as condições de uso (acadêmica e/ou privada/comercial) devem ser bem descritas. Essas licenças podem ser referenciadas com identificadores persistentes, bem como parte dos metadados em Data Objects [composto por: elementos de dados + metadados + identificador persistente]. A comunidade FAIRport fornecerá e recomendará cada vez mais licenças padrão para escolher. A comunidade FAIRport recomenda a publicação de dados em Acesso Aberto completo sempre que possível.²⁹⁴ (FORCE21, 2021, tradução nossa).

Para tanto, é necessário que as condições de acesso e licenças aplicadas estejam registradas em metadados legíveis por humanos e máquinas.

Ao que corroboram Lóscio, Burle e Calegari (2017):

Cabe aos publicadores de dados determinar as políticas por meio das quais os dados devem ser compartilhados - e sob quais circunstâncias. Espera-se que as políticas de compartilhamento de dados avaliem o risco de exposição e estabeleçam as medidas de segurança apropriadas para a proteção de dados sensíveis, tais como a autenticação segura e a autorização.²⁹⁵

E alertam ainda que “[...] [os fornecedores de dados] devem ter em mente que a combinação de dados oriundos de múltiplas fontes pode permitir a identificação inadvertida de indivíduos.” (LÓSCIO; BURLE; CALEGARI, 2017), o que deve ser fortemente combatido.

²⁹⁴ “Data Object Licenses and conditions of use (academic and/or private/commercial) should be well described. Such licenses can be referred to with persistent identifiers as well as part of the metadata in Data Objects [composed by: data elements + metadata + persistent identifier]. The FAIRport community will increasingly provide and recommend standard licenses to choose from. The FAIRport community strongly recommends to publish data in complete Open Access wherever possible.”

²⁹⁵ Lóscio, Burle e Calegari (2017) adotam a expressão “publicação de dados” e suas variações para se referir ao efeito de tornar público, para o qual, no contexto da web o termo usual é publicar. Contudo, destaca-se que a publicação de dados requer formalização, conforme discutido em Torino, Roa-Martínez e Vidotti (2020) e, ao processo de “tornar público” entende-se como disponibilização.

No contexto da Arquitetura de Dados, destaca-se a necessidade de se atentar aos aspectos éticos, legais e técnicos aplicados aos dados, todos discutidos nesta tese. No que tange aos Princípios Legais aplicados aos Dados, é importante identificar as legislações vigentes na(s) jurisdição(ões) na(s) qual(is) o projeto de AD é implementado, bem como legislações relacionadas. Essa é, sem dúvida, uma questão central no processo de coleta, tratamento, armazenamento e compartilhamento de dados de qualquer natureza. Dessa forma, o tratamento dos dados atendendo aos Princípios Legais possibilitará que os fluxos de dados se estabeleçam adequadamente.

4.4.1.5.3 Princípios técnicos aplicados aos dados

Na Arquitetura de Dados, como Princípios Técnicos aplicados aos Dados, são considerados aqueles que favorecem a localização, a coleta, o processamento, o uso e o reuso de dados por humanos e máquinas, otimizando atividades que passam a ser realizadas de forma rápida e eficiente, reduzindo as barreiras existentes.

O avanço das tecnologias computacionais e o acesso à web possibilitaram que ambos os atores supramencionados – humanos e máquinas – assumissem o papel de consumidor e fornecedor de dados de forma cíclica e de difícil separação, uma vez que, no consumo de dados, o usuário também fornece dados e o contrário é verdadeiro.

Visando contribuir para este cenário, Lóscio, Burle e Calegari (2017) apresentam, por meio do World Wide Web Consortium (W3C), boas práticas relacionadas à disponibilização e ao uso de dados na web, com vistas a apoiar o ecossistema de dados na web, quer seja para desenvolvedores, fornecedores e/ou consumidores de dados. Para os autores:

Dados devem ser descobertos e compreensíveis por pessoas e máquinas. Em qualquer circunstância em que os dados sejam utilizados de alguma maneira, seja pelo criador original dos dados ou por uma parte externa, tal utilização deve ser igualmente descoberta, assim como devem ser reconhecidos os esforços do publicador dos dados. Ou seja, a adoção destas Boas Práticas facilitará a interação entre publicadores e consumidores. (LÓSCIO; BURLE; CALEGARI, 2017)²⁹⁶.

²⁹⁶ Nesta tese entende-se que o processo de publicação de dados requer documentação contextual formalizada que auxilie consumidores de dados, por exemplo, na compreensão do contexto, das formas de coleta e tratamento, dos aparatos utilizados, entre outros (TORINO; ROA-MARTÍNEZ; VIDOTTI, 2020). Assim, ao que

Ao que acrescentam:

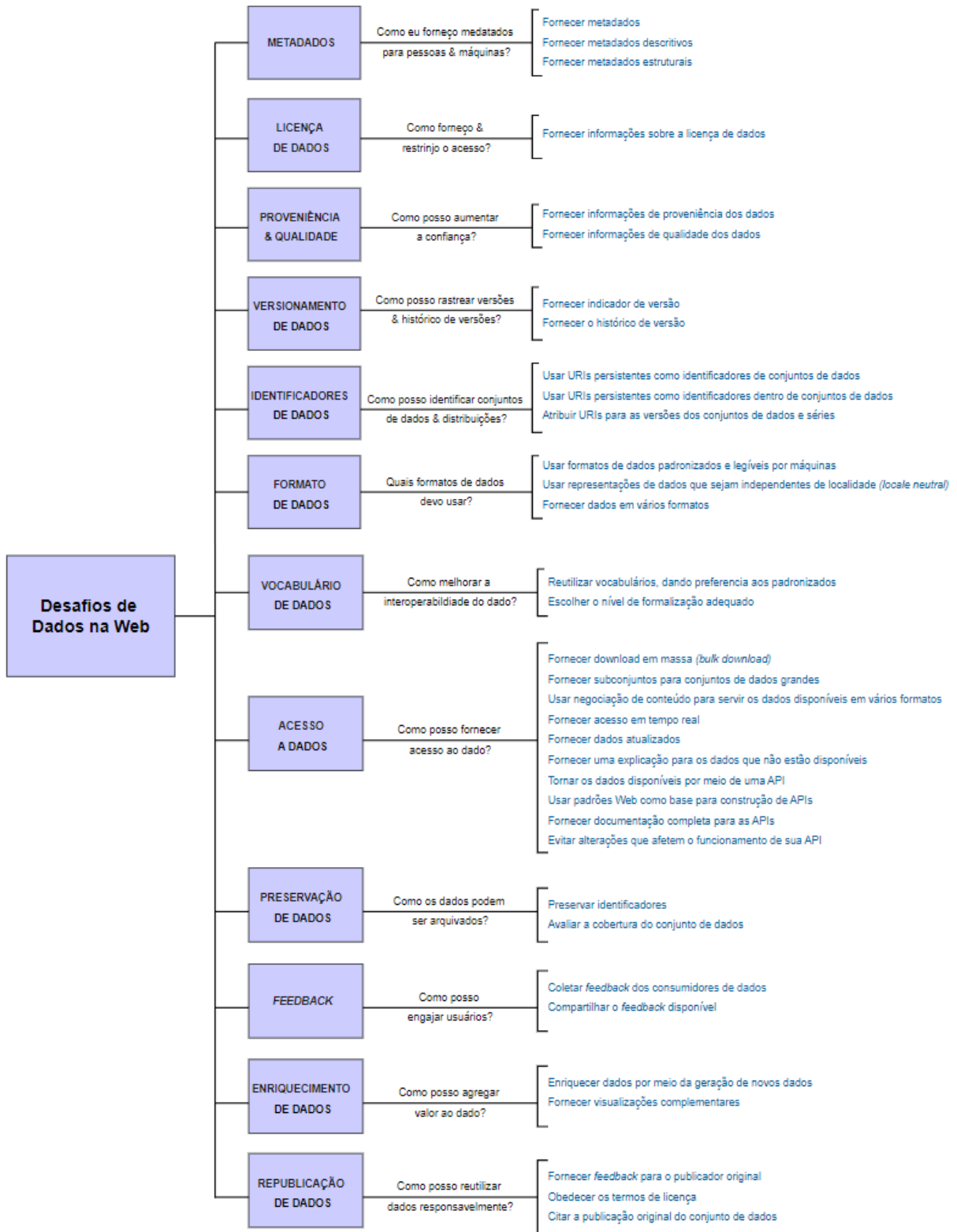
Em termos gerais os [...] [fornecedores] de dados visam compartilhar dados seja de forma aberta ou por meio de acesso controlado. Por sua vez, os consumidores de dados (que também podem eles mesmos ser [...] [fornecedores]) querem ser capazes de encontrar, usar e relacionar os dados - especialmente se estes forem precisos, regularmente atualizados e tiverem disponibilidade garantida a qualquer momento. Isto leva a necessidade de se alcançar um entendimento comum entre os [...] [fornecedores] e os consumidores de dados. Sem este acordo, os esforços dos [...] [fornecedores] de dados podem se revelar incompatíveis com os desejos dos consumidores de dados. (LÓSCIO; BURLE; CALEGARI, 2017).

Tendo em vista a necessidade de assegurar a compreensão dos dados por fornecedores e consumidores, independente do canal de comunicação e da forma de acesso, é que se desenvolvem as boas práticas para dados na web (LÓSCIO; BURLE; CALEGARI, 2017), que são constituídas por quatorze (14) práticas recomendadas para superar trinta e cinco (35) desafios enfrentados no compartilhamento de dados na web (Figura 26).

Os desafios apresentados na Figura 26 podem ser generalizados para o processo de fornecimento e consumo de dados, tendo em vista a necessidade das boas práticas que estão atreladas a metadados, licença de dados, proveniência e qualidade, versionamento, identificadores persistentes, formatos, vocabulários, acesso, preservação, enriquecimento e reuso de dados, além de feedback entre consumidores e fornecedores. Tais práticas evidentemente favorecem o processo de fornecimento e consumo, integração e interoperabilidade dos dados, por humanos e máquinas, para quaisquer finalidades.

Lóscio, Burle e Calegari (2017) chamam “publicação”, compreende-se como “disponibilização”, de igual maneira, os chamados “publicadores” são entendidos como “fornecedores”.

Figura 26 – Desafios enfrentados ao fornecer e consumir dados na web



Fonte: Lóscio, Burle e Calejari (2017).

Tal generalização encontra lugar no escopo das Boas Práticas para Dados na Web (LÓSCIO; BURLE; CALEGARI, 2017), documento que trata de conjuntos de

dados e suas distribuições, que possibilitam o compartilhamento de dados de diferentes volumes em diferentes escalas (Figura 27). Para tanto, o documento considera a base arquitetônica da web, a exemplo do que se faz no uso de *persistent identifier*/identificadores persistentes (PID) no contexto dos Princípios FAIR, no qual há a necessidade de identificação por meio de um identificador utilizando uma base arquitetônica.

Figura 27 – Publicação de dados na Web



Fonte: Lóscio, Burle e Calegari (2017).

Com isso,

Considerando esta heterogeneidade e o fato de que os [...] [fornecedores] e consumidores de dados possam não se conhecer, faz-se necessário o fornecimento de algumas informações sobre os conjuntos de dados e distribuições, de forma a promover a confiabilidade e a reutilização. (LÓSCIO; BURLE; CALEGARI, 2017).

Merece destaque o fato do atendimento às Boas Práticas para Dados na Web, na perspectiva de Lóscio, Burle e Calegari (2017), gerar melhorias na disponibilização de dados para fornecedores e consumidores, conforme apresentado no Quadro 11.

Quadro 11 – Benefícios das Boas Práticas de Dados na Web

Boa Prática	Benefícios
Fornecer metadados	Reuso Compreensão Facilidade de descoberta Facilidade de processamento
Fornecer metadados descritivos	Reuso Compreensão

Boa Prática	Benefícios
	Facilidade de descoberta
Fornecer metadados estruturais	Reuso Compreensão Facilidade de processamento
Fornecer informações sobre a licença de dados	Reuso Confiança
Fornecer informações de procedência dos dados	Reuso Compreensão Confiança
Fornecer informações de qualidade de dados	Reuso Confiança
Fornecer indicador de versão	Reuso Confiança
Fornecer o histórico de versão	Reuso Confiança
Usar URIs persistentes como identificadores de conjuntos de dados	Reuso Capacidade de conexão Interoperabilidade
Usar URIs persistentes como identificadores dentro de conjuntos de dados	Reuso Capacidade de conexão Facilidade de descoberta Interoperabilidade
Atribuir URIs para as versões dos conjuntos de dados e séries	Reuso Facilidade de descoberta Confiança
Usar formatos de dados padronizados legíveis por máquinas	Reuso Facilidade de processamento
Usar representações de dados que sejam independentes de localidade (<i>locale neutral</i>)	Reuso Compreensão
Fornecer dados em formatos múltiplos	Reuso Facilidade de processamento

Boa Prática	Benefícios
Reutilizar vocabulários, dando preferência aos padronizados	Reuso Facilidade de processamento Compreensão Confiança Interoperabilidade
Escolher o nível de formalização adequado	Reuso Compreensão Interoperabilidade
Fornecer <i>download</i> em massa (<i>bulk download</i>)	Reuso Facilidade de acesso
Fornecer subconjuntos para conjuntos de dados extensos	Reuso Capacidade de conexão Facilidade de acesso Facilidade de processamento
Usar negociação de conteúdo para disponibilizar dados em formatos múltiplos	Reuso Facilidade de acesso
Fornecer acesso em tempo real	Reuso Facilidade de acesso
Fornecer dados atualizados	Reuso Facilidade de acesso
Fornecer uma explicação para os dados que não estão disponíveis	Reuso Confiança
Disponibilizar dados por meio de uma API	Reuso Facilidade de processamento Interoperabilidade Facilidade de acesso
Usar padrões Web como base para construção de APIs	Reuso Capacidade de conexão Facilidade de descoberta Facilidade de acesso Facilidade de processamento

Boa Prática	Benefícios
Fornecer documentação completa para as APIs	Reuso Confiança
Evitar alterações que afetem o funcionamento de sua API	Confiança Interoperabilidade
Preservar identificadores	Reuso Confiança
Avaliar a cobertura do conjunto de dados	Reuso Confiança
Coletar feedback de consumidores de dados	Reuso Compreensão Confiança
Compartilhar o feedback disponível	Reuso Confiança
Enriquecer dados por meio da geração de novos dados	Reuso Compreensão Confiança Facilidade de processamento
Fornecer visualizações complementares	Reuso Compreensão Facilidade de acesso Confiança
Fornecer feedback para o publicador original	Reuso Interoperabilidade Confiança
Obedecer os termos de licença	Reuso Confiança
Citar a publicação original do conjunto de dados	Reuso Facilidade de descoberta Confiança

Fonte: Adaptado de Lóscio, Burle e Calegari (2017).

À medida que os envolvidos na gestão de dados compreendem os desafios encontrados ao longo do Ciclo de Vida dos Dados, o atendimento a Princípios Técnicos, embora por vezes se mostre moroso e até mesmo oneroso, tende a ser adotado em razão dos benefícios que serão alcançados, uma vez que, superados os desafios, os dados passam a se comunicar de forma mais fluida e transparente nos diferentes fluxos dos quais fazem parte, com destaque para a constituição de Ecologias Complexas de Dados.

Neste sentido, e no contexto desta tese, tem destaque os FAIR *Principles*/Princípios FAIR, acrônimo em inglês para *Findable*/Localizável, *Accessible*/Acessível, *Interoperable*/Interoperável, *Reusable*/ Reutilizável, publicados por Wilkinson *et al.* (2016) e cujo uso tem se disseminado, sobretudo na Ciência da Informação e nas demais áreas, como decorrentes do movimento de abertura dos dados de pesquisa.

Vale destacar que os Princípios FAIR consistem em facetas que possibilitam orientar o tratamento dos aspectos computacionais relacionados aos dados, metadados e infraestruturas nos quais estão disponíveis, e se aplicam a todos os dados, não apenas aos de pesquisa, visando favorecer a localização e a utilização dos dados disponíveis por humanos e máquinas.

Para Wilkinson *et al.* (2016, tradução nossa) “[...] os Princípios FAIR enfatizam especificamente o aprimoramento da capacidade das máquinas de encontrar e usar os dados automaticamente, além de apoiar sua reutilização por indivíduos.”²⁹⁷.

De acordo com GOFAIR (2021, tradução nossa),

Os princípios enfatizam a capacidade de ação da máquina (ou seja, a capacidade dos sistemas computacionais de encontrar, acessar, interoperar e reutilizar dados com nenhuma ou mínima intervenção humana) porque os humanos dependem cada vez mais do suporte computacional para lidar com dados como resultado do aumento do volume, complexidade e velocidade de criação de dados.²⁹⁸

Cada uma das facetas dos Princípios FAIR pressupõe o tratamento a um conjunto de elementos vinculados aos dados, metadados e infraestruturas utilizadas para o armazenamento e a disponibilização de dados, para que estejam

²⁹⁷ “[...] FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals.”

²⁹⁸ “The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.”

adequadamente preparados e possibilitem que os dados sejam localizáveis, acessíveis, interoperáveis e reutilizáveis (Figura 28).

Figura 28 – Princípios FAIR



Fonte: Australian National Data Service (2020) traduzido por Torino, Coneglian e Vidotti (2020, p. 15).

A Figura 28 elucida que um dos elementos fortemente presentes nos princípios FAIR são os metadados, sobretudo ao considerar que a maioria das facetas depende dos metadados para ser acessível. Corroborar essa afirmação GOFAIR (2021, tradução nossa) “Metadados legíveis por máquina são essenciais para a descoberta automática de conjuntos de dados e serviços, portanto, esse é um componente essencial do processo de FAIRification.”²⁹⁹

De acordo com FORCE21 (2021, tradução nossa):

[...] por meio da definição e suporte generalizado de um conjunto mínimo de princípios e práticas orientadores acordados pela comunidade, provedores de dados e consumidores de dados – tanto máquinas quanto humanos – poderiam descobrir, acessar, interoperar e reutilizar de forma sensata, com citação adequada, as vastas quantidades de informações sendo geradas pela ciência contemporânea com uso intensivo de dados.³⁰⁰

²⁹⁹ “Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.”

³⁰⁰ “[...] through the definition and widespread support of a minimal set of community-agreed guiding principles and practices, data providers and data consumers – both machine and human – could more easily discover, access, interoperate, and sensibly re-use, with proper citation, the vast quantities of information being generated by contemporary data-intensive science.”

Na concepção dos Princípios FAIR, buscou-se ampliar comportamentos integrativos e dar suporte a tecnologias, de forma a fornecer uma camada na qual o fornecimento, o consumo e a visualização de dados na internet fosse favorecida (FORCE21, 2021). Dessa forma, os Princípios FAIR orientam como as implementações devem ser feitas para que os dados e infraestruturas possam se beneficiar com a localização, acesso, interoperabilidade e reuso de dados por máquinas e, conseqüentemente, por humanos.

Vale ressaltar que os Princípios FAIR estão fortemente relacionados à capacidade das máquinas localizarem, acessarem, processarem e utilizarem dados, o que requer a localização e acesso a bases de dados de forma interoperável e facilitada. Com isso, aplicações e agentes computacionais beneficiam os humanos no processo que envolve FAIR e, com isso, otimizam a compreensão e o tratamento dos dados, em qualquer volume.

Tem-se em *Findable*/Localizável que os dados precisam ser localizados (F) para que possam resultar nas demais facetas (AIR). Para tanto, dados e metadados precisam ser localizáveis e identificáveis de forma permanente (GOFAIR, 2021; FORCE21, 2021). *Accessible*/Acessível pressupõe que, uma vez localizados, os dados precisam ser acessados por humanos e máquinas, ainda que necessária a autenticação e o uso de protocolos (GOFAIR, 2021; FORCE21, 2021). Com isso, os dados localizados e acessados podem ser *Interoperable*/Interoperáveis ou integrados a outros dados, por meio de fluxos necessários à análise, processamento e armazenamento, o que requer o uso de metadados (GOFAIR, 2021; FORCE21, 2021). Ao passo que, para que tais dados sejam *Reusable*/Reutilizáveis, precisam atender aos quesitos anteriores, sobretudo utilizando metadados adequados e exaustivos, que possibilitem a compreensão em diferentes aplicações (GOFAIR, 2021; FORCE21, 2021).

Todas as facetas FAIR são imprescindíveis e interdependentes. Contudo, é possível que a implementação dos princípios FAIR seja escalonada. Tais facetas são constituídas por subfacetas que auxiliam na compreensão e na operacionalização das facetas. “Reconhecemos explicitamente que é possível implementar qualquer uma dessas subfacetas sem implementar todas elas.” FORCE21 (2021, tradução nossa).

Visando favorecer a visualização, as facetas e subfacetas dos Princípios FAIR são apresentadas de forma sistematizada no Quadro 12.

Quadro 12 – Facetas e Subfacetas dos Princípios FAIR

Faceta	Subfaceta	Descrição
Localizáveis	F1. (Meta)dados são atribuídos a um identificador persistente (PID) internacionalmente aceito e exclusivo	Adoção de identificadores persistentes internacionalmente aceitos e exclusivos, visando manter a desambiguidade dos dados. No contexto dos dados, cada PID deverá vincular o atributo ao seu conceito.
	F2. Os dados são descritos com metadados ricos (definidos na subfaceta R1)	Utilização de metadados ricos, com ampla granularidade, visando a representação exaustiva das entidades, incluindo dados intrínsecos, de contexto, proveniência, qualidade, condições de acesso, entre outros.
	F3. Os metadados incluem clara e explicitamente o identificador dos dados que descrevem	Deve haver correspondência entre os metadados e os conjuntos de dados representados, cuja associação deve ser formal e pode ser feita por meio de um PID.
	F4. (Meta)dados são registrados ou indexados em um recurso pesquisável	Metadados e conjuntos de dados devem ser indexados em mecanismos de busca e coletados por <i>crawler</i> , o que favorece a localização.
Acessíveis	A1. (Meta)dados são recuperáveis por seu identificador usando um protocolo de comunicação padronizado	Os metadados devem ser, sempre que possível, acessíveis sem barreiras pela internet, utilizando padrões de comunicação bem implementados e documentados. No caso de dados que possuam restrições de acesso, deve-se possibilitar o contato com o responsável.
	A1.1 O protocolo é aberto, gratuito e universalmente implementável	Considerando a subfaceta A1, o protocolo utilizado deve ser aberto e gratuito, possibilitando a implementação e o acesso aos metadados por qualquer consumidor (humano e máquina).
	A1.2 O protocolo permite um procedimento de autenticação e autorização, quando necessário	Considerando a subfaceta A1, quando os dados estiverem em acesso restrito, deve haver uma especificação nos metadados acerca desta condição e, se necessário, da forma de autenticação para acesso.
	A2. Os metadados são acessíveis, mesmo quando os dados não estão mais disponíveis	Se ocorrer a retirada de um conjunto de dados, os metadados e PID devem permanecer disponíveis, válidos e acessíveis. Adicionalmente, é recomendado incluir dados acerca da retirada do conjunto de dados.

Faceta	Subfaceta	Descrição
Interoperáveis	I1. (Meta)dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento	A localização, o acesso e a interoperabilidade necessitam de compreensão acerca dos vocabulários utilizados, que devem ser acionáveis por humanos e máquinas, sem a necessidade de processamento prévio.
	I2. (Meta)dados usam vocabulários que seguem aos princípios FAIR	O vocabulário adotado para descrever os conjuntos de dados deve utilizar PID e estar adequadamente documentado, com fácil acesso aos consumidores de dados.
	I3. (Meta)dados incluem referências qualificadas a outros (meta)dados	Devem ser utilizadas referências cruzadas entre os dados e metadados, visando contribuir para o estabelecimento do contexto dos dados.
Reutilizáveis	R1. (Meta)dados ricos são utilizados para a descrição com uma pluralidade de atributos precisos e relevantes	A descrição dos conjuntos de dados deve ser exaustiva e utilizar metadados ricos que possibilitem aos consumidores, sobretudo as máquinas, a identificarem se os dados correspondem à necessidade. Para tanto, deve incluir dados contextuais e, quando se aplicar, dos aparatos e condições de criação/coleta dos dados.
	R1.1. (Meta)dados são liberados com uma licença de uso de dados clara e acessível	As condições de acesso e uso dos dados e metadados deve ser especificada claramente, para assegurar a “interoperabilidade legal”. A adoção de licenças e seus limites devem ser claramente expressos e compreensíveis por humanos e máquinas.
	R1.2. (Meta)dados estão associados a proveniência detalhada	O reuso dos dados requer compreensão acerca da proveniência dos dados, o que inclui seu histórico ao longo do tempo, desde a criação, os processamentos, as transformações, entre outros. Essas características devem ser descritas de forma legível para a compreensão por humanos e máquinas.
	R1.3. (Meta)dados atendem aos padrões relevantes da comunidade para o domínio	Os padrões de metadados existentes e utilizados no domínio dos dados devem ser utilizados, uma vez que o reuso de dados semelhantes é facilitado.

Fonte: Adaptado de GOFAIR (2021).

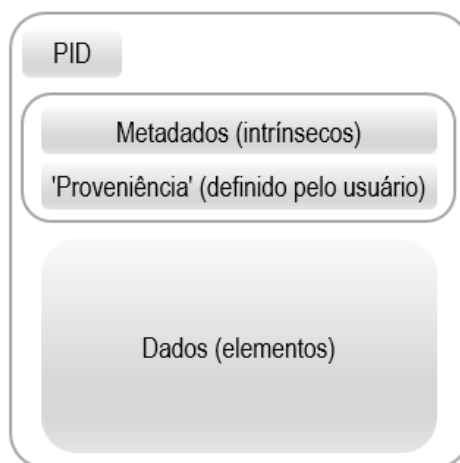
Destaca-se que, embora os Princípios FAIR tenham caráter de orientação, muitas ferramentas tecnológicas têm sido disponibilizadas para auxiliar na FAIRificação dos dados, metadados e infraestruturas, bem como para a verificação do nível de observância aos princípios, o que pode contribuir com consumidores e

fornecedores de dados. Além disso, há iniciativas que buscam capacitar as pessoas interessadas em conhecer e/ou ampliar o conhecimento acerca dos Princípios FAIR (EUROPEAN COMMISSION, 2016; FAIR FUNDERS COLLABORATION, 2019; LIGUE DES BIBLIOTHÈQUES EUROPÉENNES DE RECHERCHE, 2020; FAIRDATABR, 2021).

No que tange à legibilidade dos dados por máquina, FORCE21 (2021) afirma que metadados legíveis por máquinas são imprescindíveis para que os dados sejam FAIR. Da mesma forma, se os dados forem FAIR, a interoperabilidade é facilitada. Isso se aplica aos elementos de dados e/ou atributos de metadados, mas também aos valores, para os quais a terminologia utilizada deve ser conhecida por fornecedores e consumidores, por meio de declarações acessíveis.

Isso posto, é possível que dados, metadados e infraestruturas estejam em diferentes níveis de observância aos Princípios FAIR. Assim, considerando que *Data Object* são compostos por elementos de dados, metadados e identificador persistente, e que, no processo de consumo de dados, é imprescindível que a proveniência seja conhecida, a Figura 29, representa os principais elementos de um *Data Object*.

Figura 29 – Elementos de objeto de dados



Fonte: FORCE21 (2021, tradução nossa).

Para considerar o nível de observância aos princípios FAIR, FORCE21 (2021) propõe seis possibilidades:

- a) em desconformidade com os Princípios FAIR: quando nenhuma das facetos está disponível;

- b) localizável e utilizável por humanos: quando apenas os identificadores persistentes (PID) estão disponíveis;
- c) metadados FAIR: quando PID, metadados e proveniência estão disponíveis;
- d) dados FAIR com acesso restrito: quando PID, metadados e proveniência estão disponíveis, e, os dados estão armazenados com restrição de acesso;
- e) dados FAIR com acesso aberto: quando PID, metadados e proveniência estão disponíveis, e, os dados estão armazenados sem restrição de acesso;
- f) dados FAIR com acesso aberto e ligados funcionalmente: quando PID, metadados e proveniência estão disponíveis, e, os dados estão armazenados sem restrição de acesso, com possibilidade de interoperar com outros dados.

Considerando a necessidade de auxiliar no processo de tornar dados, metadados e infraestruturas FAIR, o GOFAIR (2022) elucida que localização e acesso estão fortemente atrelados aos metadados, enquanto a interoperabilidade e a reutilização estão vinculadas aos dados. Destaca-se que os processos de localização e acesso também são dependentes da infraestrutura de armazenamento, que deve adotar padrões, protocolos e demais ferramentas que possibilitem que os conteúdos armazenados sejam FAIR.

No que tange à observância aos Princípios FAIR, também conhecida como *FAIRification*/FAIRificação ou a ação de tornar FAIR, GOFAIR (2022), aponta sete passos:

- a) recuperar dados que não sejam FAIR, ou seja, os dados que estão enquadrados “em desconformidade com os Princípios FAIR” (FORCE21, 2021);
 - b) analisar os dados recuperados: identificar as características dos dados, considerando atributos utilizados interna e externamente, como estão estruturados e a forma de armazenamento;
 - c) definir o modelo semântico para o conjunto de dados: para descrever o significado das entidades, relacionamentos, atributos e valores, sem ambiguidade e processável por máquina;
-

- d) tornar os dados vinculáveis: com base na etapa anterior, possibilitar a interoperabilidade e a reutilização dos dados por diferentes sistemas, considerando as características dos próprios dados e considerando a necessidade de convertê-los para as definições do Ciclo de Vida dos Dados na base de consumo;
- e) atribuir licença: a licença deve figurar nos metadados, para serem facilmente localizadas por máquinas, porém, é imprescindível que as mesmas informações sobre a licença adotada estejam disponíveis junto aos dados, sobretudo nos casos em que seu uso é externo à base de fornecimento dos dados, ou seja, em uma base terceira que, pode não ser a detentora inicial dos dados;
- f) definir metadados para os conjuntos de dados: é um dos pilares fundamentais, uma vez que metadados adequados e enriquecidos são imprescindíveis para que os dados sejam FAIR;
- g) implantar o recurso de dados FAIR: adotar os Princípios FAIR para dados, metadados e infraestrutura visando a recuperação por mecanismos de busca e favorecer os demais Princípios FAIR.

Assim, no contexto da Arquitetura de Dados, é importante que a observância às Boas Práticas para Dados na Web e aos Princípios FAIR seja parte das definições tomadas pelo detentor do Ciclo de Vida dos Dados, visando possibilitar o atendimento, considerando os Princípios Técnicos aplicados aos Dados, em todos os níveis. Destaca-se, contudo, que quando houver qualquer restrição imposta pelos requisitos éticos e/ou legais, medidas devem ser tomadas para assegurar, antes de tudo, que os dados sejam preservados.

Embora a presente tese utilize como bases para os Princípios Técnicos aplicados aos Dados as Boas Práticas para Dados na Web (LÓSCIO; BURLE; CALEGARI, 2017) e os Princípios FAIR (WILKINSON *et al.*, 2016; FORCE21, 2021; GOFAIR, 2022), eles são orientativos, de forma que outras bases podem ser utilizadas no sentido de favorecer os aspectos técnicos que favoreçam a localização, o acesso, o processamento, o uso e o reuso de dados.

Todos os elementos anteriormente descritos auxiliam no planejamento e no projeto da camada de dados de sistemas de informação e/ou ambientes informacionais digitais, beneficiando, sobretudo, os fornecedores e consumidores de dados, na identificação dos Princípios Técnicos necessários ao fornecimento e

consumo de dados, em diferentes Ciclos de Vida dos Dados, considerando o fluxo estabelecido pela Arquitetura de Dados.

Para a Arquitetura de Dados no contexto da Ciência da Informação, os Princípios Éticos, Legais e Técnicos aplicados aos dados governam as formas de coleta, armazenamento e as condições de disponibilização para acesso e reuso, considerando o tratamento ético e legal dos diferentes processos, para que os processos técnicos sejam realizados.

Nesse sentido, há diferentes atores envolvidos no processo de criação, coleta, armazenamento, tratamento, disponibilização, acesso, uso e reuso dos dados, sendo, eles responsáveis pelo tratamento ético, legal e técnico dos dados.

Nesta tese, utilizam-se leis e resoluções brasileiras para discutir os aspectos éticos e legais envolvidos aos dados, quer seja por quem os cria/coleta, trata, armazena, disponibiliza e/ou (re)usa. Contudo, é importante destacar que, muitas vezes, há insciência ou pouca compreensão das pessoas acerca desses processos relacionados aos dados que fornecem, bem como de quando e como esses dados são coletados e que tipo de tratamento se faz sobre eles, o que pode constituir uma situação que merece atenção, sobretudo diante da crescente interação de pessoas com aparatos tecnológicos.

No que tange aos dados pessoais, à LGPD e aos elementos do CVD, é preponderante considerar que, embora a Lei estabeleça que compete ao referenciado a decisão sobre os seus dados, é de fato o detentor do CVD quem possui gestão sobre os dados e, em conjunto com o intermediário, operacionaliza as decisões tomadas. Essa situação é destacada, pois embora na Lei a “autoridade para decidir” sobre os dados seja do indivíduo referenciado nos dados, uma vez coletados, os dados passam a estar sob posse de um terceiro, cujo tratamento pode ter conduta ética e legal ou ser realizado de forma a simular esse tratamento, contudo, mantendo os dados inalterados em algum local do banco de dados, sobre o qual apenas o detentor e o intermediário possuem gestão.

É inegável que algumas alterações e descartes de dados podem acarretar situações-problema. Contudo, tal fato deve ter previsão legal. A esse respeito, é possível ter como exemplo, no Brasil, a alteração de nome pessoal conforme determina a Lei nº 14.382, de 27 de junho de 2022 (BRASIL, 2022) que, uma vez que não estabelece a possibilidade de manutenção do nome anterior, entende-se esse

regido pela LGPD que prevê, no Art. 18, inciso “III - correção de dados incompletos, inexatos ou desatualizados”. Contudo, tal correção pode acarretar em “apagamento” de dados anteriores e que podem, inclusive, ser utilizados de forma comprobatória no espaço temporal anterior à “atualização”, cuja brecha para a manutenção pode estar no Art. 16, inciso “IV - uso exclusivo do controlador, vedado seu acesso por terceiro, e desde que anonimizados os dados.”, o que poderia embasar a manutenção dos dados anteriores para uso do controlador (conforme especificado na referida Lei).

A LGPD busca tratar os dados de forma abrangente, embora haja especificidades que são difíceis de legislar e fiscalizar, ainda que se pretenda. Além disso, há lacunas na Lei, questões que possibilitam diferentes interpretações e, ainda e sobretudo, uma espécie de descuido ou falta de atenção das pessoas quanto ao valor e ao impacto que pode ser causado pelo fornecimento de dados.

Ainda, no que tange aos seres humanos no âmbito da pesquisa, há resoluções brasileiras que buscam contribuir com o tratamento ético entre pesquisadores e sujeitos de pesquisa. Nesse sentido, os aspectos éticos relacionados aos dados – sejam eles pessoais, de pesquisa ou de qualquer outra natureza – requerem compreensão e boa fé dos envolvidos, uma vez que são também de difícil legislação e fiscalização.

Compreende-se, assim, que há forte relacionamento entre os princípios éticos e legais aplicados aos dados e que o seu tratamento depende de legislação, mas também de cultura, para que as condutas adotadas nos processos de criação/coleta, tratamento, armazenamento, disponibilização, acesso, uso e reuso sejam de fato éticas e legais.

No que tange aos princípios técnicos aplicados aos dados, há claro relacionamento com os princípios éticos e legais, uma vez que esses são representados nos metadados acionáveis pelos princípios técnicos. Contudo, há também especificidades.

Assim, é importante destacar que os Princípios Técnicos aplicados aos dados atuam no processamento técnico e tecnológico que possibilita que os dados, uma vez coletados, armazenados e disponibilizados, possam ser facilmente localizáveis, acessíveis, interoperáveis e reutilizáveis, por processamento computacional, em benefício de máquinas e humanos. Para tanto, os aspectos a eles relacionados são fortemente atrelados às atividades de identificação – sobretudo das condições que

são aplicadas aos dados pelos Princípios Éticos e Legais – para a comunicação de dados entre máquinas, o que ocorre pelos Princípios Técnicos aplicados aos dados.

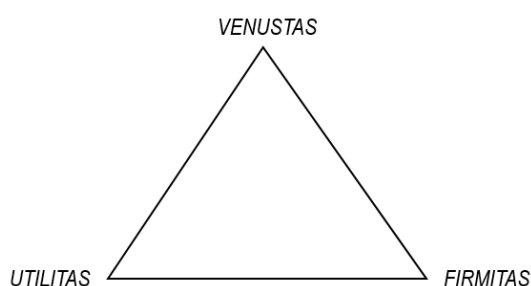
Dessa forma, os Princípios Éticos, Legais e Técnicos aplicados aos Dados, enquanto elementos da Arquitetura de Dados, possibilitam que a comunicação de dados entre máquinas e entre máquinas-humanos ocorra em atendimento aos aspectos éticos, considerando a hegemonia das pessoas, atendendo aos aspectos legais cabíveis vigentes nas jurisdições aplicáveis, para que esses dados estejam “em conformidade” para armazenamento, disponibilização e uso, o que ocorre e se manifesta nos aspectos técnicos utilizados na representação e comunicação entre máquinas.

4.4.2 Arquitetura de Dados no contexto da Ciência da Informação: Tríade e Fluxo

A tríade vitruviana apresenta três princípios básicos por meio dos quais a Arquitetura deve se basear – *venustas*, *utilitas*, *firmitas* – visto que a ausência de um desses elementos descaracteriza a arquitetura (VITRUVIUS POLLIO, 2007).

Nesse sentido, há de se considerar, com base em Vitruvius Pollio (2007) cada um dos elementos que compõem a tríade vitruviana (Figura 30). Inicialmente, é necessário compreender que, de acordo com Vitruvius Pollio (2007, p. 81, grifos do autor), “As partes da própria arquitetura são três: *edificação*, *gnomônica* e *mecânica*.”, e que essas partes devem considerar “[...] os princípios da solidez, da funcionalidade e da beleza”. Na obra, traduzida por Maciel, a forma latina da escrita desses elementos é mencionada em nota de rodapé, na qual constam: “*Firmitas*: solidez, firmeza, consistência, robustez.”, “*Utilitas*: utilidade, uso, funcionalidade, proveito, vantagem.” e “*Venustas*: beleza, elegância e estética” (VITRUVIUS POLLIO, 2007, p. 82, grifos do autor).

Figura 30 – Tríade Vitruviana



Fonte: Autoria própria.

O autor complementa:

O princípio da solidez estará presente quando for feita a escavação dos fundamentos até o chão firme e se escolherem diligentemente e sem avareza as necessárias quantidades de materiais. O da funcionalidade, por sua vez, será conseguido se for bem realizada e sem qualquer impedimento a adequação do uso dos solos, assim como uma repartição apropriada e adaptada ao tipo de exposição solar de cada um dos gêneros. Finalmente, o princípio da beleza será atingido quando o aspecto da obra for agradável e elegante e as medidas das partes corresponderem a uma equilibrada lógica de comensurabilidade. (VITRUVIUS POLLIO, 2007, p. 82).

O Conceito de Arquitetura de Dados no contexto da Ciência da Informação, cunhado nesta tese, busca estrutura na tríade vitruviana. Com isso, faz-se necessário o estabelecimento da tríade, por meio da qual o projeto da infraestrutura deve ser equilibrado, em alusão à Arquitetura e a Vitruvius Pollio.

Assim, a Arquitetura de Dados considera os princípios *Firmitas*, *Utilitas* e *Venustas*, presentes na tríade vitruviana da Arquitetura, para, a partir do Conceito de Arquitetura de Dados cunhado nesta tese, estabelecer a sua tríade.

Dessa forma, apreende-se que a Arquitetura de Dados tem como tríade **Forma**, **Utilidade** e **Robustez**, conforme explicitado na Figura 31.

Figura 31 – Tríade da Arquitetura de Dados



Fonte: Autoria própria.

A tríade da Arquitetura de Dados, composta pelos princípios de Forma, Utilidade e Robustez, considera como **Forma** os aspectos sintáticos e semânticos relacionados aos dados, cuja **Utilidade** se materializa no uso e reuso dos dados em diferentes ambientes informacionais digitais e sistemas de informação, por meio da **Robustez** das tecnologias e técnicas empregadas.

O princípio da **Forma** se relaciona ao contexto no qual a Arquitetura de Dados se insere. Por isso, deve ser considerado no momento da criação, coleta, tratamento e armazenamento dos dados, considerando ainda o armazenamento para favorecer o consumo e fornecimento por diferentes ambientes informacionais digitais e sistemas de informação, possibilitando a equidade entre as partes que compõem o todo.

Destaca-se que, no caso da aplicação da Arquitetura de Dados na estruturação de dados já existentes, recomenda-se o uso das Oito fases do modelo de Ciclo de Vida dos Dados de Yoon, Aiken e Guimaraes (2005), apresentado na Figura 18, para que os dados existentes sejam avaliados, caso necessário, corrigidos e refinados, para que estejam adequados ao projeto de AD e, ainda, como forma de avaliar a necessidade da criação de novos dados.

Na Arquitetura de Dados, o princípio da **Utilidade** considera os requisitos necessários para a comunicação de dados (entrada e saída) entre Ciclos de Vida dos Dados, convertendo-os, quando necessário, para diferentes esquemas e padrões, visando uso e reuso. Por isso, atenta-se aos princípios éticos, legais e técnicos necessários para o consumo, o fornecimento e a comunicação de dados.

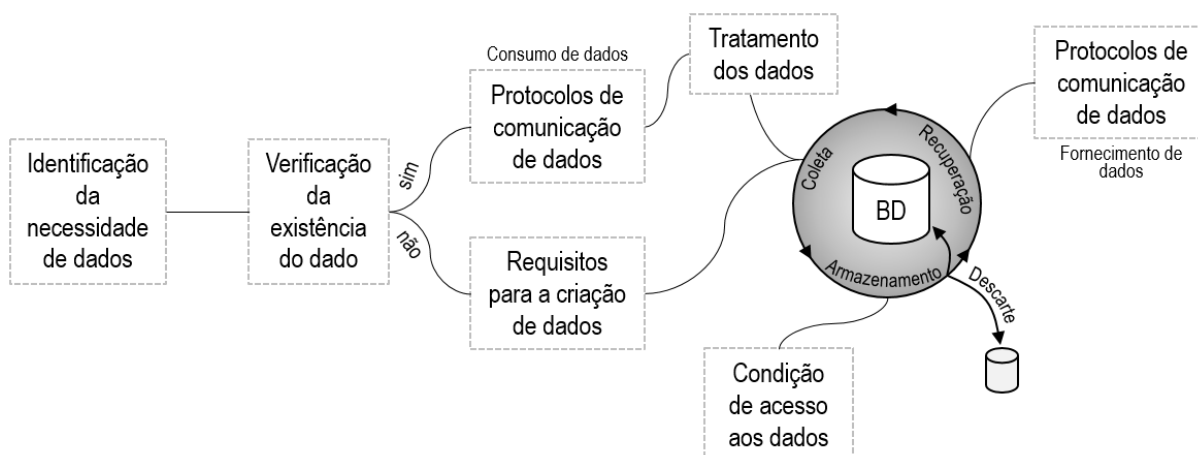
Por fim, o princípio da **Robustez** consiste no uso adequado das técnicas e tecnologias existentes no que tange à infraestrutura dos dados, seus metadados, padrões, esquemas e protocolos, bem como aos aspectos éticos, legais e técnicos aplicados aos dados.

Tendo em vista a tríade – Forma, Utilidade e Robustez –, é importante compreender que os dados são parte de uma estrutura macro, que consiste no contexto em que sua existência é necessária e que pode, por exemplo, consistir em uma empresa, pesquisa e/ou função. Essa estrutura macro pode requerer uma grande quantidade de dados, que estarão disponíveis para finalidades específicas, das quais decorrem o Ciclo de Vida dos Dados.

Nesse sentido, na Arquitetura de Dados há um Fluxo de Dados que orienta a identificação da necessidade de dados, a coleta/criação, o tratamento, o armazenamento, a recuperação/fornecimento e o descarte, quando necessário, a partir do Ciclo de Vida dos Dados (Figura 32).

O fluxo de dados na Arquitetura de Dados no contexto da Ciência da Informação (Figura 32) se inicia na identificação da necessidade de um dado, seguida da verificação de existência, que consiste na busca pelos dados na estrutura do contexto ao qual o CVD se relaciona ou ainda em um CVD externo.

Figura 32 – Fluxo de Dados na Arquitetura de Dados



Fonte: Autoria própria.

Considerando que o dado necessário esteja disponível, ele pode ser consumido para atender à necessidade do CVD, o que requer que ambas as instâncias, a que possui o dado e o fornece, e a que necessita do dado e o consome, disponibilizem protocolos de comunicação de dados, por meio dos quais os dados podem ser coletados. Neste caso, identificado o dado necessário e disponível o protocolo para a comunicação dos dados, eles são consumidos para que possam ser armazenados no CVD que identificou a necessidade e, para tanto, devem passar por uma etapa de tratamento, na qual os dados são estruturados de acordo com os parâmetros estabelecidos pela Arquitetura de Dados e implantados pelo detentor do CVD em questão. Dessa forma, no fluxo, o tratamento dos dados idealmente deve ocorrer na etapa de coleta, para que seja armazenado no CVD de acordo com os padrões por ele estabelecidos.

Vale considerar que há etapas que não estão exibidas no Ciclo de Vida dos Dados e que possuem especificidades. Delas, estão destacadas no Fluxo de Dados na Arquitetura de Dados (Figura 32), o tratamento, relacionado à coleta do CVD, para que os dados coletados atendam aos requisitos do CVD no que tange, por exemplo, aos padrões de metadados, estruturas sintática e semântica dos dados; e a condição de acesso aos dados, que se constitui em uma forma de tratamento no armazenamento e para a disponibilização dos dados na fase de recuperação do CVD e que pode estar relacionada a aspectos éticos, legais e técnicos.

Nesse sentido, Date (2004) alerta que, quando os dados são extraídos de outras fontes, é necessário que a extração esteja acompanhada dos processos de limpeza, transformação e consolidação antes de serem armazenados no CVD que

executou a coleta, e que cada um deles possui especificidades – a esses processos, consideradas pelo autor como preparação de dados, designamos tratamento.

Nesse contexto, a extração é apoiada por programas específicos, ao passo que a limpeza, atrelada à qualidade dos dados na fonte original, consiste, por exemplo, na adequação da representação, no que tange ao preenchimento inadequado de valores, valores omitidos, correção de erros de digitação (DATE, 2004), bem como na padronização dos valores considerando as características sintáticas e semânticas adotadas no CVD de destino. A transformação consiste na adequação dos dados ao modelo definido pela Arquitetura de Dados e utilizado no CVD, etapa na qual podem ser identificadas falhas que não foram corrigidas na etapa de limpeza, e a consolidação é justamente a fusão dos dados (DATE, 2004). Por outro lado, caso o dado não esteja disponível, ele deverá ser criado, considerando os requisitos para a criação de dados. Nesse sentido, o detentor do CVD que necessita do dado deverá proceder às prerrogativas do contexto ao qual está vinculado para que o dado necessário seja criado. Com isso, atende aos requisitos para a criação do dado, que pode acontecer diretamente no CVD, na fase de coleta, ou na estrutura macro, passando a ser identificado como um dado existente e, neste caso, já criado de acordo com os parâmetros especificados, o que torna desnecessário o tratamento dos dados para que possa fazer parte do CVD.

A etapa de coleta deve considerar, ainda, a frequência de atualização da coleta dos dados, visando assegurar a atualidade e a qualidade dos dados que compõem o CVD.

No Ciclo de Vida dos Dados, é na fase de coleta que o princípio **forma**, parte da tríade da Arquitetura de Dados, está presente. Assim, a criação ou o tratamento dos dados deve considerar aspectos sintáticos e semânticos que dão a sustentação à forma dos dados no CVD.

Uma vez coletados, os dados são armazenados, de acordo com os parâmetros e requisitos definidos pela Arquitetura de Dados e implantados pelo detentor do CVD, considerando a modelagem de dados, a estrutura de banco de dados e demais aspectos envolvidos no armazenamento. Ainda, relacionado ao armazenamento, serão determinadas as condições de acesso aos dados armazenados no CVD, sendo que aqueles que não possuem restrições poderão ser acessados na fase de recuperação do CVD.

No que tange à fase de recuperação dos dados armazenados no CVD, como mencionado, os dados que não possuem restrição de acesso podem ser disponibilizados, atendendo aos parâmetros definidos pela Arquitetura de Dados e implantados detentor do CVD, dos quais devem fazer parte os princípios éticos, legais e técnicos para a disponibilização, visando maximizar a capacidade de localização, acesso e processamento dos dados.

Pode-se considerar, com isso, que a fase de recuperação consiste na disponibilização. Nesse sentido, cumpre esclarecer que o CVD favorece a localização e a recuperação dos dados, sobretudo por máquinas, não se ocupando das ações dos sujeitos, como proposto na Encontrabilidade da Informação, que não pode ser assegurada pelo CVD.

Ainda na fase de recuperação, devem ser disponibilizados protocolos de comunicação de dados que possibilitem o fornecimento de dados pelo CVD, visando que os dados sejam consumidos, de forma automática ou pré-definida, por outros CVD que os localizem e deles necessitem, gerando, com isso a recuperação dos dados.

Considerando a possibilidade de haver necessidade de exclusão ou desativação dos dados do Ciclo de Vida dos Dados, a fase de descarte pode ser utilizada para que os dados sejam desativados ou definitivamente eliminados. Tupper (2011) alerta que o custo de coleta de um conjunto de dados pode ser superior ao custo de mantê-lo armazenado, ao passo que Sant'Anna (2016) menciona que a redução dos custos de armazenamento tende a diminuir o descarte de dados.

Nesse sentido, e considerando ainda a ampliação das discussões acerca da preservação digital de longo prazo de objetos digitais e dados, parece evidenciar-se que o descarte tende a ser pouco utilizado. Porém, há a necessidade de mantê-lo, para ocorrências em que os dados necessitem ser descartados ou desativados.

Caso haja o descarte de dados, é importante atentar-se à necessidade de manter um registro do processo, de forma legível por humanos e máquinas, para gestão do próprio CVD, para informar usuários interessados que possam buscá-los de forma direta ou indireta, evitando com isso que o ID do dado seja reutilizado, o que poderia acarretar inconsistência nas coletas previamente configuradas.

O princípio **utilidade**, parte da tríade da Arquitetura de Dados, está presente no fluxo de dados na Arquitetura de Dados (Figura 32) desde a etapa de identificação

de necessidade até a fase de recuperação, considerando os processos de uso e reuso de dados para diferentes finalidades.

O fluxo de dados requer, ainda, **robustez**, parte da tríade da Arquitetura de Dados, uma vez que necessita de um conjunto de técnicas e tecnologias para o funcionamento. Este aspecto requer atualização constante para acompanhar os avanços das técnicas e tecnologias existentes, para empregá-las quando necessário.

Um aspecto a ser considerado é que a Arquitetura de Dados estará em permanente evolução, visando acompanhar o desenvolvimento de todas as etapas que compõem seu fluxo, bem como a evolução do contexto ao qual está inserida. Essa possibilidade constante de evolução, considerando inclusive a necessidade de refinamento dos dados, aponta a flexibilidade da Arquitetura de Dados de acordo com os limites estabelecidos no projeto.

Para Tupper (2011, p. 14, tradução nossa), a arquitetura deve acompanhar o desenvolvimento do contexto no qual está inserida

A técnica que cada corporação deve aprender é que suas arquiteturas devem permanecer organismos vivos que evoluem com a necessidade do negócio. Negligenciando isso, a arquitetura ficará velha e inflexível e eventualmente morrerá.³⁰¹

Assim, é evidente que a Arquitetura de Dados necessita acompanhar o desenvolvimento tecnológico. Contudo, seu projeto não deve fixar aparatos, ferramentas, linguagens de programação e outros aspectos que possam reduzi-la ou ampliar a possibilidade de obsolescência do projeto.

Alternativas para a abordagem da arquitetura estão sendo constantemente procuradas devido à evolução em software e hardware que está acelerando com o tempo. [...] A arquitetura fornece diretrizes e frameworks pelos quais essas decisões podem ser tomadas com mais facilidade, uma vez que um “inventário” do que já existe está disponível.³⁰² (TUPPER, 2011, p. 14, tradução nossa).

Para Tupper (2011), a Arquitetura de Dados, no contexto da Arquitetura Empresarial, consiste em um modelo de design estratégico, capaz de auxiliar em diferentes atividades de desenvolvimento. E enfatiza que “O modelo de dados corporativos é de particular importância na compreensão da arquitetura de dados porque é nesse nível que todos os aplicativos, redes, organizações, processos,

³⁰¹ “The technique that each corporation must learn is that its architectures must remain living organisms that evolve with the business need. Neglecting this, the architecture will grow old and inflexible and eventually die.”

³⁰² “Alternatives to the architecture approach are constantly being sought because of the evolution in both software and hardware that is accelerating with time. [...] Architecture provides guidelines and frameworks by which these decisions can be made more easily, since an “inventory” of what already exists is available.”

projetos, sistemas e tecnologias diversificados se reúnem.”³⁰³ (TUPPER, 2011, p. 43, tradução nossa).

Tal afirmação reforça a característica da Arquitetura de Dados enquanto projeto da infraestrutura de dados de forma macro, para que os dados possam ser utilizados para todas as finalidades, por todas as ferramentas disponíveis e de forma atualizada, uma vez que a AD deve ser um processo contínuo, capaz de conduzir o crescimento sistemático, ordenado e sustentável da estrutura de dados. Com isso, a Arquitetura de Dados pode beneficiar a gestão de dados como ativos, menor custo no (re)uso de dados e ampliação das possibilidades de análises e visualização.

4.4.3 Entregáveis da Arquitetura de Dados

Artefatos de Arquitetura de Dados, segundo DAMA-DMBOK (2017, p. 98, tradução nossa) consistem em “[...] especificações usadas para descrever o estado existente, definir requisitos de dados, orientar a integração de dados e controlar ativos de dados conforme apresentado na estratégia de dados.”³⁰⁴

No contexto da Ciência da Informação, apresentam-se como entregáveis da Arquitetura de Dados, de forma análoga ao que se tem na Arquitetura da Informação. O objetivo desses entregáveis é fornecer o projeto estrutural documentado da camada de dados dos ambientes informacionais, visando a adequada gestão dos dados, metadados, padrões e estruturas utilizadas, de forma a otimizar a compreensão por humanos e máquinas. Por isso, possui base sintática e semântica.

Neste sentido, como entregáveis da Arquitetura de Dados, tem-se:

- a) projeto da Arquitetura de Dados;
- b) requisitos para inclusão e exclusão de metadados;
- c) perfil de aplicação de metadados;
- d) características sintáticas e semânticas dos dados e do(s) vocabulário(s) de metadados utilizados;
- e) repositório digital de metadados;
- f) modelo lógico de dados;

³⁰³ “The enterprise data model is of particular importance in understanding the data architecture because it is at this level that all diversified applications, networks, organizations, processes, projects, systems, and technologies come together.”

³⁰⁴ “[...] specifications used to describe existing state, define data requirements, guide data integration, and control data assets as put forth in a data strategy.”

- g) mapas de fluxos de dados;
- h) protocolos de comunicação de dados disponíveis;
- i) documentação dos sistemas de informação com descrição da finalidade de uso, dos dados armazenados, da linguagem utilizada, do Ciclo de Vida dos Dados (coleta, armazenamento, recuperação e descarte) e do atendimento a requisitos éticos e legais.

Os entregáveis da Arquitetura de Dados estão relacionados ao conceito e aos elementos da AD e serão desenvolvidos de acordo com as especificações do projeto de Arquitetura de Dados e do contexto em que ele será desenvolvido.

4.4.4 Arquiteto de Dados

De acordo com Tupper (2011, p. 3, tradução nossa):

A palavra “arquiteto” é de origem grega e tem duas raízes: arch e tect. Arch significa “primário”, “primitivo” ou “vindo antes” e tect significa “fazer” ou “passar a existir”. Assim, quando as duas raízes são combinadas para formar a palavra “architectos”, pode significar “vir antes” + “vir a existir”. Usando esses vários significados, podemos descobrir as definições de arquiteto e arquitetura. Uma boa definição de arquitetura de negócios seria “um plano primitivo” ou “um plano antes da construção.” Esta definição captura sua essência: Arquitetura é um esforço analítico que é criado antes da ocorrência de qualquer atividade de construção real. É a estrutura ou esboço abstrato que fornece diretrizes para a construção do início ao fim.³⁰⁵

Morville e Rosenfeld (2006) destacaram algumas áreas de conhecimento que possuem abordagens vinculadas à informação e são importantes na atuação com Arquitetura da Informação. Dentre elas, destacam Ciência da Informação e Biblioteconomia pela experiência na organização, indexação, acesso e disseminação da informação, bem como no processo de busca e recuperação da informação e no domínio de tecnologias relacionadas, e Ciência da Computação, por suas habilidades na modelagem de conteúdos, metadados, bancos de dados, aplicação de sistemas de informação e tecnologias.

³⁰⁵ “The word “architect” is of Greek origin and is composed of two roots: arch and tect. Arch means “primary,” “primitive,” or “coming before,” and tect means “make” or “come into being.” Thus, when the two roots are combined to form the word “architectos,” it can mean “coming before” + “come into being.” Using these multiple meanings, we can figure out the definitions of architect and architecture. A good working definition of architecture would be “a primitive plan” or “a plan before construction.” This definition captures the essence of it: Architecture is an analytical effort that is created prior to the occurrence of any real construction activity. It is the abstracted framework or outline that provides guidelines for the construction from the beginning to the end.”

No que tange à atuação na Arquitetura de Dados, de acordo com DAMA-DMBOK (2017, p. 99, tradução nossa):

Os arquitetos de dados criam e mantêm o conhecimento organizacional sobre os dados e os sistemas pelos quais eles se movem. Esse conhecimento permite que uma organização gerencie seus dados como um ativo e aumente o valor que obtém de seus dados, identificando oportunidades de uso de dados, redução de custos e mitigação de riscos.³⁰⁶

No contexto desta tese, tendo em vista o percurso teórico realizado e a compreensão da Arquitetura de Dados no contexto da Ciência da Informação, entende-se que o Arquiteto de Dados é o sujeito que atua no projeto de estruturação dos dados e do Ciclo de Vida dos Dados, de forma integrada, para que sejam compreensíveis por máquinas, visando a localização, o acesso, o processamento e o (re)uso de forma automática ou automatizada, e por humanos, cuja compreensão ocorre pela estrutura cognitiva.

O Arquiteto de Dados deve ter em mente que a estruturação dos dados e do Ciclo de Vida dos Dados, por meio da Arquitetura de Dados, visa favorecer a compreensão dos dados, em primeira instância por máquinas, que trará benefícios para o processamento por aplicações e agentes computacionais, e em última instância para humanos, que terão dados estruturados e contextualizados para aproveitamento por processamento cognitivo e/ou computacional, de forma a se constituírem em informação.

Compete ao Arquiteto de Dados gerar convergência entre diferentes padrões, esquemas e estruturas de dados, tornando a compreensão e processamento facilitados, atendendo a requisitos e princípios internacionalmente aceitos e validados, bem como aqueles locais e atrelados ao contexto para o qual a Arquitetura de Dados foi desenvolvida e implementada.

Dessa forma, o Arquiteto de Dados auxilia na determinação de estruturas locais (micro), que podem se comunicar por integração e/ou interoperabilidade com outras estruturas de dados (macro), com barreiras mínimas, como as de conversão dos dados para a estrutura do Ciclo de Vida dos Dados micro.

Cabe ainda ao Arquiteto de Dados a criação e manutenção da documentação referente aos dados, dos requisitos de criação, manutenção, alteração e descarte de

³⁰⁶ *“Data architects create and maintain organizational knowledge about data and the systems through which it moves. This knowledge enables an organization to manage its data as an asset and increase the value it gets from its data by identifying opportunities for data usage, cost reduction, and risk mitigation.”*

dados de forma integrada, das estruturas de comunicação de dados e dos aspectos éticos, legais e técnicos envolvidos, independente da forma de implementação, desde que estejam facilmente acessíveis, visando facilitar a comunicação de dados entre diferentes Ciclos de Vida de Dados, internos e/ou externos. Essa documentação se materializa nos Entregáveis da Arquitetura de Dados, que devem refletir o projeto de Arquitetura de Dados, e estar constantemente atualizados.

Além disso, o Arquiteto de Dados deve se ocupar da manutenção da integridade da Arquitetura de Dados, atentando-se para que os entregáveis do projeto que necessitem de gestão sejam armazenados em um repositório digital, preferencialmente confiável. Esse repositório consistirá em fonte de informação para gestores, Arquitetos de Dados, Arquitetos da Informação e desenvolvedores, de forma a contribuir para que as estratégias definidas sejam refletidas do projeto lógico e físico dos dados, reduzindo as inconsistências e possibilitando o mapeamento e correção, quando necessário.

Destaca-se, nesse contexto, que no âmbito da Arquitetura de Dados não são especificadas áreas do conhecimento às quais devem estar vinculados os Arquitetos de Dados. Assim, a atuação como Arquiteto de Dados não está atrelada a nenhum domínio do conhecimento específico, mas se vale de conhecimentos técnicos e tecnológicos relacionados ao campo da Arquitetura de Dados e suas especificidades.

Nesse contexto, é importante destacar que a área da Ciência da Informação, desde sua gênese, tem se especializado em estruturas e formas de representação, o que implica em metadados e no tratamento de dados, em diferentes contextos, agregando os conhecimentos já aplicados pela área no tratamento da informação. Assim, é possível afirmar que a Ciência da Informação está apta a trabalhar com os processos inerentes à gestão de dados em suas diferentes tipologias, sobretudo na coleta, tratamento, armazenamento, acesso, (re)uso e preservação. Nesse processo, o tratamento inclui, além da representação, as questões legais, éticas e de segurança.

Evidencia-se assim que os profissionais da Ciência da Informação, sobretudo aqueles que se especializam no campo da tecnologia, são profissionais adequados e qualificados para a atuação como Arquiteto de Dados.

4.5 ECOLOGIA COMPLEXA DE DADOS

Uma busca exploratória no contexto da Ciência da Informação (Quadro 1) aponta a inexistência de discussões que envolvam o que se entende nesta tese por Ecologia Complexa de Dados. Com isso, busca-se, a partir do percurso realizado pela pesquisa, das lacunas encontradas na Arquitetura da Informação e do conceito de Arquitetura de Dados, tecer considerações iniciais acerca da Ecologia Complexa de Dados no contexto da Ciência da Informação.

Inicialmente, destaca-se que não há pretensão de discutir o emprego dos termos ecologia ou ecossistema, visto que essa abordagem já foi realizada anteriormente por outros pesquisadores, a exemplo de Oliveira (2014).

Para discutir Ecologia Complexa de Dados, diante da ausência de literatura específica, conforme mencionado anteriormente, utiliza-se da literatura existente na área de Ciência da Informação, acerca de Ecologia da Informação (DAVENPORT, 1998; MORVILLE; ROSENFELD, 2006; ROSENFELD; MORVILLE; ARANGO, 2015) e Ecologia Informacional Complexa (OLIVEIRA, 2014; OLIVEIRA; VIDOTTI, 2016), cuja discussão já está assentada na área.

No contexto da Arquitetura de Dados, embora abordada no domínio da tecnologia da informação, incorporamos à discussão Araújo Machado, Costa e Santos (2022) que afirmam haver um descontentamento oriundo do descompasso entre as necessidades organizacionais e as respostas possibilitadas pela Arquitetura de Dados, que ocasionam sobrecarga às equipes, em razão da crescente necessidade de dados por parte das organizações. Tal fato, segundo os autores, com base em Zhamak Dehghani (2020), gerou uma crise na AD, que culminou no surgimento de um novo paradigma, atrelado à *Data Mesh* (ARAÚJO MACHADO; COSTA; SANTOS, 2022).

A *Data Mesh* indica que haja governança distribuída dos dados, de forma que a gestão ocorra nas diferentes áreas específicas que compõem o negócio. Assim, a gestão, o consumo e o fornecimento de dados são descentralizados e obedecem à área específica do negócio. Nesse contexto, os dados são entendidos como produtos e seus consumidores são vistos como clientes, o que muda a perspectiva da relação fornecimento-consumo. Para tanto, os dados devem cumprir a seis princípios atrelados à qualidade dos dados e eficiência da *Mesh*, de acordo com os DATSIS *Principles*: *discoverable*/detectável, *addressable*/endereçável, *trustworthy*/confiável,

self-describing/autodescritivo, *interoperable*/interoperável e *secure*/seguro (ARAÚJO MACHADO; COSTA; SANTOS, 2022).

Ainda na *Data Mesh*, cada produto é entendido como um nó, que possui como componentes principais o código, os dados e metadados e a infraestrutura. Dessa forma, os nós integram o *Data Mesh's Domain Model* (ARAÚJO MACHADO; COSTA; SANTOS, 2022), do qual fazem parte um catálogo, por meio do qual são relacionados todos nós e suas especificidades, um centralizador de modificações realizadas, e um canal de comunicação eficiente entre as equipes.

Por fim, destaca-se ainda da obra que, nesse cenário de mudança de paradigma, no que tange à Arquitetura de Dados, Araújo Machado, Costa e Santos (2022, tradução nossa) afirmam acreditar que há uma crise científica:

Uma forte crença de que, assim como a engenharia de software, a engenharia de dados está caminhando para o abandono das arquiteturas monolíticas, aprimorando a descentralização e a escalabilidade de seus sistemas.³⁰⁷

Esclarece-se que esta tese não objetiva discutir a temática *Data Mesh*, tampouco as referidas crise e crença. Contudo, tem-se o contato com a obra de Araújo Machado, Costa e Santos (2022) como um achado da investigação realizada. Com isso, trechos do estudo são trazidos para a discussão no contexto da Ecologia Complexa de Dados, pela compreensão do relacionamento existente, sobretudo no que tange à descentralização das fontes de dados, no caso da Ecologia Complexa de Dados.

Inicialmente, apreende-se do estudo de Araújo Machado, Costa e Santos (2022) que o descontentamento das organizações acerca da Arquitetura de Dados pode estar atrelado à abordagem comumente dada à AD no campo de Tecnologia da Informação, na qual a ênfase está nos sistemas de informação. Nesse caso, a governança distribuída entre as diferentes áreas do negócio tende a ocasionar duplicação de esforços, bem como de (infra)estruturas e de (meta)dados. É possível, ainda, vislumbrar que os nós propostos na *Data Mesh* possuem estrutura similar à do Ciclo de Vida dos Dados.

Considerando essa interpretação, a Arquitetura de Dados no contexto da Ciência da Informação, como conceituada nesta tese, pode ser uma solução para o

³⁰⁷ "A strong belief that, like software engineering, data engineering is moving towards the unfollow of monolithic architectures, enhancing the decentralization and scalability of its systems."

descontentamento apresentado, uma vez que o foco está nos dados, seus fluxos e Ciclos de Vida dos Dados, na estruturação da camada de dados, para que se possa ter comunicação facilitada entre os ambientes informacionais e/ou sistemas de informação, com vistas ao processamento computacional em favor do benefício humano.

Nesse sentido, destaca-se que a Ecologia Complexa de Dados não busca discutir a governança dos dados de forma monolítica ou descentralizada no contexto organizacional, mas se vale da discussão apresentada por Araújo Machado, Costa e Santos (2022) para discutir a Ecologia Complexa de Dados a partir da sua constituição descentralizada, que pode ocorrer por meio de Arquiteturas de Dados e Ciclos de Vida dos Dados internos ou externos à organização.

Nesta tese, considerando o contexto organizacional apresentado por Araújo Machado, Costa e Santos (2022) em que a gestão dos dados ocorre, compreende-se que, embora haja crescente demanda interna por dados, o fato de essas estruturas serem responsáveis pelos Ciclos de Vida dos Dados inegavelmente torna a tarefa de gestão, seja ela monolítica ou descentralizada, exequível, uma vez que à organização compete a tomada de decisões acerca dos dados, bem como o papel de detentor do(s) CVD, o que favorece a estruturação e o fluxo dos dados.

Essa discussão ganha complexidade quando a tomada de decisão acerca dos dados é realizada por diferentes instâncias, negócios e organizações, sobre as quais não se tem gestão, o que gera a necessidade de conhecer as estruturas, por meio da análise ou da documentação, para que o fluxo de dados ocorra com o menor ruído e intervenções manuais possíveis.

É nesse contexto, em que os dados estão armazenados em diferentes infraestruturas e Ciclos de Vida dos Dados, nos quais pode haver maior dificuldade na integração dos dados, considerando que, “Os dados são criados em muitos lugares e são movidos entre os locais para uso.”³⁰⁸ e que “[...] em geral a diferença entre aquilo que a aplicação vê e o que está realmente armazenado pode ser considerável.” (DATE, 2004, p. 19), que a Ecologia Complexa de Dados emerge.

Nesse sentido, a Ecologia Complexa de Dados se constitui da ligação de diferentes Arquiteturas de Dados adequadamente implementadas e que, com isso,

³⁰⁸ “Data is created in many places and is moved between places for use.”

possibilitam a ligação dos dados em infraestruturas amplas e enriquecidas a partir da integração de diferentes infraestruturas.

Assim, tem-se nesta tese um conceito operacional de Ecologia Complexa de Dados:

A Ecologia Complexa de Dados é a rede de relacionamentos de dados estabelecida por meio de Arquitetura(s) de Dados e/ou de Ciclo(s) de Vida dos Dados, constituída de forma sistêmica.

A Ecologia Complexa de Dados emerge quando os ambientes informacionais e/ou sistemas de informação possuem dados estruturados que podem ser interligados, intercambiados ou interoperados com outros ambientes informacionais e/ou sistemas de informação de forma transparente, em função da compatibilidade possível em razão dos elementos da Arquitetura de Dados: Ciclo de Vida dos Dados, metadados, estruturas sintática e semântica dos dados, e princípios éticos, legais e técnicos aplicados aos dados.

Com isso, os dados podem ser consumidos e (re)utilizados para diferentes finalidades, dentre as quais se destacam: enriquecimento e ligação de dados, constituição de novos bancos de dados, desenvolvimento de interfaces para humanos e máquinas, criação de visualizações e *dashboards*, análises diversas, além da utilização para o desenvolvimento – a partir do consumo dos dados – de Arquiteturas da Informação e Ecologias Informacionais Complexas.

Parte da Ecologia Complexa de Dados são os metadados e vocabulários padronizados de dados, utilizados para representar os atributos, as ferramentas disponíveis para criá-los e disponibilizá-los, por meio de ontologias, protocolos e serviços de comunicação de dados capazes de suportar, de forma consistente e segura, a oferta e o consumo de dados de múltiplas fontes.

Ainda, no contexto da Ecologia Complexa de Dados é necessário que cada Arquitetura de Dados, quando do fornecimento de dados assegure o atendimento aos Princípios Éticos, Legais e Técnicos aplicados aos Dados, com preponderância para as Boas Práticas para Dados na Web e para os Princípios FAIR para que os atributos e o conjunto de dados (*dataset*), quando houver, possam ser *findable*/localizáveis (vocabulários padronizados de dados, protocolos e serviços de comunicação de dados), *accessible*/acessíveis (vocabulários padronizados de dados, protocolos e

serviços de comunicação de dados), *interoperable/interoperáveis* (vocabulários padronizados de dados, protocolos e serviços de comunicação de dados) e *reusable/reutilizáveis* (metadados, vocabulários padronizados de dados e ontologias).

Nesse sentido, a Ecologia Complexa de Dados é constituída por dados estruturados, que utilizam metadados padronizados e podem estar acompanhados de objetos informacionais e/ou apenas conteúdos.

As Ecologias Complexas de Dados mantêm a proveniência da Arquitetura de Dados e atendem a princípios e práticas que favoreçam a localização e o consumo dos dados automaticamente por máquinas, além daqueles relacionados ao tratamento ético e legal dos dados.

Além disso, a Ecologia Complexa de Dados disponibiliza protocolos e/ou serviços de comunicação de dados para que eles possam ser acionados automaticamente por outras Arquiteturas de Dados ou Ecologias Complexas de Dados, para o consumo e/ou fornecimento de dados. Para tanto, utilizam-se, quando necessário, da conversão de registros para que os dados sejam armazenados de acordo com os padrões estabelecidos na AD que está consumindo os dados.

Sobre as Ecologias Complexas de Dados, é possível realizar visualizações para análises diversas, e a partir delas é possível desenvolver ambientes informacionais digitais e/ou Ecologias Informacionais Complexas.

As Ecologias Complexas de Dados podem realizar a comunicação dos dados com outras estruturas, visando ampliar ou enriquecer os dados. Podem, ainda, estabelecer uma estrutura central, que funcione como um *hub*, para a qual os dados podem ser fornecidos – caso em que se recomenda um processo de deduplicação de registros – e do qual os dados podem ser consumidos, visando igualmente ampliar ou enriquecer os dados, utilizando quando necessário a conversão de registros para o padrão estabelecido pelo consumidor dos dados.

Essa estrutura central, além de agregar dados consumidos das estruturas específicas, pode fornecer dados para a criação de novas estruturas específicas que podem ser Arquiteturas de Dados ou Ecologias Complexas de Dados, bem como para o desenvolvimento de ambientes informacionais digitais e/ou Ecologias Informacionais Complexas.

Destaca-se, assim, que a Arquitetura de Dados (AD) é que propicia a constituição da Ecologia Complexa de Dados (ECD). Por isso, considera-se que o

planejamento e a implementação da ECD são favorecidos pela estruturação proporcionada pela AD, embora não sejam dela dependentes.

Sabe-se que é possível constituir Ecologias Complexas de Dados a partir de ambientes informacionais digitais e/ou sistemas de informação desestruturados, a partir da aplicação de diferentes técnicas e tecnologias. Contudo, com elevado custo – de recursos humanos, de processamento e tecnológico –, sobretudo se consideradas as potenciais falhas no processamento.

Assim, cumpre destacar que é altamente recomendável o desenvolvimento da Arquitetura de Dados diretamente nas instâncias que serão parte da Ecologia Complexa de Dados, visando, sobretudo, manter as estruturas adequadas na fonte primária de armazenamento dos dados.

Dessa forma, os dados, uma vez coletados e armazenados em uma infraestrutura que compõe a Ecologia Complexa de Dados, desde que atendam aos princípios éticos, legais e técnicos, podem ser facilmente comunicados entre máquinas.

4.6 ARQUITETURA DE DADOS: REFLEXÕES

A Arquitetura de Dados parte do pressuposto que a adequada estruturação da camada de dados de sistemas de informação e/ou ambientes informacionais digitais permite que os dados sejam facilmente localizados, acessados, processados, interoperados e (re)utilizados em diferentes contextos. Para tanto, a AD se ocupa de padrões, estruturas e protocolos capazes de permitir que os dados sejam compreensíveis e acionáveis por usuários humanos e máquinas.

Considerando que a comunicação é um dos desafios da Arquitetura de Dados, acredita-se que a compreensão por humanos e máquinas possa se alicerçar na Teoria Matemática da Comunicação, constituída por emissor – mensagem – canal – receptor.

Destaca-se, nesse sentido, que o processo de comunicação entre humanos pode apresentar ruídos, decorrentes, por exemplo, da ausência de conhecimento prévio, da carga emocional ou de outros fatores que são próprios dos contextos humanos.

Para Stonier (1997, p. 189, tradução nossa), “O significado é um estado que define a relação entre duas categorias de informações: informações externas, que são

apresentadas a um sistema receptor, e informações pré-existentes armazenadas pelo sistema receptor, que podem interagir com as informações recebidas.”³⁰⁹

Considerando essa afirmação, o significado depende da relação entre a informação recebida e a pré-existente em um receptor, sendo, conforme o autor, a primeira relacionada à mensagem e a segunda ao contexto.

Na comunicação entre máquinas, é necessário que haja correspondência entre ambas para que o contexto da mensagem seja compreendido e ela tenha significado para o receptor. Com isso,

Quanto maior a interação, mais vínculos são criados entre a nova informação e as diversas estruturas de informação existentes no ambiente interno de informação do receptor. Quanto maior essa interação, mais significativa se torna a informação veiculada pela mensagem.³¹⁰ (STONIER, 1997, p. 189, tradução nossa).

No caso da Teoria Matemática da Comunicação aplicada às máquinas, o emissor/receptor são as máquinas, entendidas como aplicações, agentes computacionais, ambientes informacionais digitais, sistemas de informação; a mensagem são os dados, que se utilizam de estruturas de representação, com o uso de metadados acionáveis, por meio dos quais são transportados, acompanhados de objetos digitais, quando existentes; o canal é entendido como protocolos, portas e webservices; o ruído é reduzido por padrões, vocabulários, normas e por princípios técnicos, que possibilitam criar contexto e compreensão para máquinas.

Ainda, quando da comunicação humano-máquina, são necessários elementos estruturantes da Arquitetura da Informação e da Arquitetura de Dados, sendo a primeira responsável pelo conteúdo, contexto e usuário e, a segunda, pela estrutura dos dados e processos, considerando Forma, Utilidade e Robustez, capazes de fornecê-los adequadamente para consumo, inclusive pela Arquitetura da Informação *bottom-up*.

No cenário de crescente coleta, armazenamento, disponibilização, acesso e uso de dados, é relevante que a Ciência da Informação contribua na estruturação da camada de dados dos sistemas de informação e ambientes informacionais digitais, visando a otimização dos esforços nos processos anteriormente mencionados, bem

³⁰⁹ “Meaning is a state which defines the relationship between two categories of information: external information, which is presented to a recipient system, and pre-existing information stored by the recipient system, which can interact with the incoming information.”

³¹⁰ “The greater the interaction, the more bonds are created between the new information and the various information structures existing within the internal information environment of the recipient. The greater this interaction, the more meaningful becomes the information carried by the message.”

como o favorecimento do acesso e do (re)uso, a partir da comunicação e compreensão efetiva entre humanos, máquinas e humanos e máquinas.

Santos e Vidotti (2009) afirmam que:

Transformar a imensa massa de dados operacionais disponíveis diariamente em informações consistentes que permitam a tomada de decisões e agreguem valor às atividades e aos negócios é um dos desafios da Ciência da Informação. Nesse processo de transformação, tem ocorrido uma rearticulação das relações sociais e de produção em torno das Tecnologias de Informação e Comunicação, gerando um deslocamento das instâncias de mediação política, econômica e social, da dimensão espacial para a temporal, e a construção e instalação do princípio de instantaneidade e de imediatez como base de regulação de nossa experiência significativa.

Ao que corrobora Sant'Ana (2016, p. 119):

A ciência da Informação pode e deve contribuir para que este cenário de acesso e uso intenso de dados se desenvolva da melhor maneira possível, buscando identificar e estudar fatores e características que propiciem ampliação do equilíbrio entre os atores envolvidos no processo e a máxima otimização do uso dos dados.

Nesse sentido, a Arquitetura de Dados no contexto da Ciência da Informação, tem o potencial de contribuir com a estruturação da camada de dados dos sistemas de informação e/ou ambientes informacionais digitais e, por consequência, com a Arquitetura da Informação, com a Recuperação da Informação e com a Encontrabilidade da Informação, com as quais possui relacionamentos direto e indireto.

Além disso, a Arquitetura de Dados deve propiciar que bancos de dados de sistemas de informação e ambientes informacionais digitais estejam adequadamente estruturados e, com isso, possam fornecer e consumir dados de outras estruturas, para que possam ter usos diversos, dentre os quais se destaca a constituição de Ecologias Complexas de Dados.

Destaca-se que, para Tupper (2011), os dados existem independentemente das perspectivas sobre eles. Porém, o autor alerta que o armazenamento de dados, em geral, se preocupa apenas com o atendimento a necessidades atuais, desconsiderando a relevância das necessidades futuras.

Vale lembrar que uma das etapas do Ciclo de Vida dos Dados (SANT'ANNA, 2016) consiste no descarte dos dados, que pode ocorrer com o apagamento efetivo ou com a desativação dos dados. Contudo, o próprio autor destaca que os custos de armazenamento dos dados estão cada vez menores, o que o torna menor do que a

necessidade de nova coleta ou geração, o que parece apontar para a possibilidade de armazenamento definitivo dos dados.

Assim, uma vez que Arquitetura de Dados estabelece o projeto de infraestrutura dos dados, contribui, dentre outros fatores, para as ações necessárias para a preservação digital de longo prazo dos dados, tornando-os passíveis de tratamento sempre que necessário.

Nesse sentido, a preservação digital tem sido amplamente discutida, para os dados registrados em objetos digitais nos mais diversos suportes. Há de se atentar, contudo, para a necessidade de preservação digital dos próprios dados, nos diferentes CVD dos quais são parte, atentando-se às características técnicas necessárias para que sejam passíveis de (re)uso ao longo do tempo e, ainda, para que sejam entendidos como parte da memória da instituição que os gerou ou coletou.

Tupper (2011) aponta a necessidade de manter a integridade da Arquitetura de Dados, para que os projetos estratégicos sejam refletidos no projeto lógico, disponível no repositório, de forma a contribuir para as implementações nos modelos físicos. “Isso garantirá que a análise de negócios corporativa apropriada seja feita antes do projeto de sistemas.”³¹¹ (TUPPER, 2011, p. 199, tradução nossa). Embora o autor atribua essa competência ao gestor do repositório de metadados, nesta tese a entendemos como parte da competência do Arquiteto de Dados.

Para finalizar essa reflexão, esta tese corrobora o entendimento de Tupper (2011), de que é possível desenvolver sistemas de informação sem arquitetura e que esses sistemas se aprimorem, mas que há aumento da eficiência e qualidade, bem como redução de custos com o uso de uma abordagem arquitetônica. Assim, destaca-se que essa afirmação também se aplica aos dados, que podem ser gerenciados sem a Arquitetura de Dados, contudo, os custos de manutenção da estrutura, fornecimento e consumo serão significativamente ampliados, com perda de eficiência; ao passo que a adoção da Arquitetura de Dados consiste em uma atividade que busca coerência e holismo no projeto e desenvolvimento da infraestrutura de dados e de ciclo de vida dos dados, visando favorecer o fornecimento e o consumo de dados.

Dessa forma, pretende-se encorajar gestores, Arquitetos da Informação, Arquitetos de Dados, desenvolvedores, profissionais da informação e demais envolvidos no planejamento da estruturação de dados e Ciclo de Vida dos Dados a

³¹¹ “This will ensure that the appropriate enterprise business analysis is done prior to systems design.”

investir no planejamento ou na adequação da Arquitetura de Dados, considerando a tríade – **forma**, **utilidade** e **robustez** – e os elementos que a compõem, visando assegurar eficiência, eficácia, otimização da comunicação de forma a tornar os dados localizáveis, acessíveis, interoperáveis e reutilizáveis no contexto interno e/ou externo, auxiliar na preservação digital de longo prazo, além de favorecer o processamento dos dados por máquinas em benefício de máquinas e humanos.

5 ARQUITETURA DA INFORMAÇÃO E ARQUITETURA DE DADOS: RELAÇÕES POSSÍVEIS

As discussões realizadas nesta tese permitem afirmar que a Arquitetura da Informação atua principalmente na interface dos ambientes informacionais digitais e sistemas de informação, fato que reforça a relevância da AI *top-down*, inclusive nos estudos desenvolvidos na área de Ciência da Informação, ao passo que a Arquitetura da Informação *bottom-up* é pouco abordada e, com isso, as discussões teóricas e práticas são ainda incipientes.

Na seção 3 desta tese, os precursores da Arquitetura da Informação foram revisitados para iniciar a tessitura necessária à apreensão das lacunas existentes nesse campo. De igual maneira, os antecedentes da Arquitetura de Dados foram discutidos na seção 4.2, para que fosse possível conceituar a Arquitetura de Dados no contexto da Ciência da Informação (seção 4.4). Com isso, foi possível compreender a Arquitetura de Dados e estabelecer seus elementos conceituais essenciais no contexto da Ciência da Informação. Assim, é possível estabelecer as relações entre Arquitetura da Informação e Arquitetura de Dados a partir desta pesquisa.

Tendo em vista a conceituação de Arquitetura de Dados no contexto da Ciência da Informação, realizada na seção 4.4 do presente estudo, essa seção busca discutir as relações entre Arquitetura da Informação e Arquitetura de Dados, que consiste em um dos objetivos específicos (d) e uma das inquietações de pesquisa: como a Arquitetura de Dados se relaciona com a Arquitetura da Informação?

Assim, no intuito de discutir as relações entre a AI e a AD, inicia-se a análise pelo objeto, sendo na Arquitetura da Informação a Informação, e na Arquitetura de Dados, os Dados. Na seção 4.1 desta tese, são apresentadas tessituras teóricas acerca de dados e informação, das quais é possível afirmar que o processamento da informação é realizado apenas por humanos, ao passo que as máquinas processam dados.

No que tange à tríade, na Arquitetura da Informação a tríade é constituída por conteúdo, contexto e usuário, enquanto na Arquitetura de Dados, a tríade consiste em forma, utilidade e robustez.

Nas tríades de ambos os conceitos AI e AD, encontram-se relacionamentos, respectivamente entre conteúdo (AI) e, forma e robustez (AD), sendo os elementos caracterizados de forma mais específica, no caso da AI no que é considerado nesta

tese como Sistema de Representação da Informação, cujo conteúdo, na Arquitetura da Informação *bottom-up* necessita de estrutura, apresentada por Morville e Rosenfeld (2006) e Rosenfeld, Morville e Arango (2015) por meio dos metadados, que estão presentes na Arquitetura de Dados, em “forma”, considerando os elementos “metadados” e “estruturas sintática e semântica dos dados” e, que se utiliza da “robustez” ancorada nos aspectos relacionados às técnicas e tecnologias empregadas para que os dados sejam acionados na camada de dados – estruturada pela Arquitetura de Dados –, e exibidos na interface para humanos, tendo como elo a Arquitetura da Informação *bottom-up*. De igual maneira, os conteúdos que fazem parte de um ambiente informacional digital projetado pela AI podem ser acionáveis por máquinas, desde que haja estrutura e permissão de acesso, a partir da qual a coleta e processamento podem ser facilitados por meio da adoção dos elementos da AD.

Assim, o conteúdo, preponderante para a Arquitetura da Informação encontra forma e robustez na Arquitetura de Dados.

Ainda, no que tange à forma, na Arquitetura de Dados, tem como elementos os metadados e as estruturas sintática e semântica aplicadas aos dados, os quais se relacionam com o Sistema de Representação da Informação da Arquitetura da Informação, constituído por Tesouros, Vocabulários Controlados e Metadados, que consistem em estruturas sintáticas e semânticas de representação da informação, que podem ser utilizadas nos metadados de forma estrutural, para a padronização de atributo e valor, utilizando-se diferentes vocabulários, dados e representação das entidades, a exemplo dos vocabulários controlados e da linguagem natural, que pode se apoiar ainda no Processamento de Linguagem Natural (PLN).

Conforme abordado na seção 3, a Arquitetura da Informação é compreendida de forma *top-down* e *bottom-up*, possui abordagens arquitetural, sistêmica e informacional, e é composta pelos sistemas de organização, rotulagem, navegação, busca e tesouros, vocabulários controlados e metadados.

Por outro lado, a Arquitetura de Dados, conceituada nesta tese, possui elementos: Ciclo de Vida dos Dados, Metadados, Estruturas Sintática e Semântica de Dados e, Princípios Éticos, Legais e Técnicos aplicados aos Dados.

Assim, no que tange à composição, não há correspondência exata entre a anatomia da Arquitetura da Informação e os elementos da Arquitetura de Dados. Porém, é possível identificar relacionamento, uma vez que a AI *bottom-up* aponta a necessidade de ferramentas e soluções tecnológicas para alavancar o conteúdo

(HAGEDORN, 2000). Já a atuação da Arquitetura de Dados na infraestrutura de dados e no Ciclo de Vida dos Dados faz com que a AD possa contribuir com a solução tecnológica necessária para alavancar o conteúdo na Arquitetura da Informação, por meio da *AI bottom-up*.

Nesse sentido, a Arquitetura de Dados tem nos elementos “metadados” e “estruturas sintática e semântica dos dados” a correspondência com “tesauros, vocabulários controlados e metadados” presentes nos sistemas da Arquitetura da Informação. Cumpre destacar que, embora haja correspondência, a discussão acerca dos metadados realizada na AI se relaciona com a representação da informação, sobretudo no campo de valor, ao passo que, na AD, a esse uso é adicionado o elemento atributo, cujas características somadas é que possibilitam alavancar os conteúdos, quer seja pela interface para humanos ou pela comunicação de dados por máquinas.

No que tange aos entregáveis, os desenvolvidos pela Arquitetura da Informação estão atrelados principalmente à interface dos ambientes informacionais digitais e/ou sistemas de informação, ao passo que os desenvolvidos pela Arquitetura de Dados estão relacionados aos dados e aos Ciclos de Vida dos Dados.

Nesse sentido, especificamente, o repositório de metadados, apresentado por Brandt (2020) como um entregável da AI, se assemelha ao repositório presente na AD, de forma que pode ser ampliado para um repositório genérico, que possa armazenar dados e informações relacionados, respectivamente, à Arquitetura da Informação e à Arquitetura de Dados. Esse ambiente pode compreender diferentes ferramentas, armazenar entregáveis de ambos os projetos, de forma versionada, que possibilite ainda gerenciar o desenvolvimento da Arquitetura de Dados e da Arquitetura da Informação do(s) sistema(s) de informação(ões) e/ou ambiente(s) informacional(is) digital(is) disponíveis, incluindo políticas envolvidas com o(s) Ciclo(s) de Vida dos Dados. Esse repositório pode se constituir em um repositório digital confiável, que propicie a preservação digital de longo prazo.

A presente tese apresenta como achado da pesquisa o reposicionamento da abordagem Pervasiva da Arquitetura da Informação para a abordagem Informacional. Neste contexto, está presente a Ecologia Informacional Complexa, que consiste no conjunto de ambientes, canais, mídias, tecnologias, sujeitos e comportamentos, interligados pela informação (OLIVEIRA, 2014). Nesse sentido, tendo a informação a

característica pervasiva, o projeto de AI é ampliado para múltiplos ambientes, mídias, canais e tecnologias, por meio da Ecologia Informacional Complexa.

Desse modo, a Arquitetura da Informação se ocupa do planejamento de um ambiente informacional, ao passo que a Ecologia Informacional Complexa, quando da informação pervasiva, atua no planejamento de ambientes, canais, mídias, tecnologias para assegurar que tenham características que os identifiquem como um único projeto de Arquitetura da Informação.

Por outro lado, a Ecologia Complexa de Dados, presente na Arquitetura de Dados, se constitui a partir da interligação de Arquiteturas de Dados ou de infraestruturas de dados que não estejam estruturadas pela AD. Dessa forma, a Ecologia Complexa de Dados possibilita o relacionamento entre diferentes fontes de armazenamento de dados, que podem se constituir em outras Arquiteturas de Dados, Ecologias Complexas de Dados, ambientes informacionais digitais, Ecologias Informacionais Complexas, além de permitir que sejam geradas interfaces de consumo para visualização.

Diante do exposto, e considerando as discussões realizadas nas seções 3 e 4 desta tese, as relações entre a Arquitetura da Informação e a Arquitetura de Dados encontradas nesta tese e que respondem ao objetivo específico (d) são sistematizadas no o Quadro 13.

Quadro 13 – Relações entre Arquitetura da Informação e Arquitetura de Dados

Característica	Arquitetura da Informação (AI)	Arquitetura de Dados (AD)	Relação
Objeto	Informação	Dados	-
Tríade	Conteúdo Contexto Usuário	Forma Utilidade Robustez	Conteúdo (AI) e, Forma e Robustez (AD)
Constituição	Anatomia <i>Top-down</i> e <i>Bottom-up</i> Sistemas de Organização, Rotulagem, Navegação, Busca e, Tesouros, Vocabulários Controlados e Metadados	Metadados Estruturas Sintática e Semântica Princípios Éticos, Legais e Técnicos Ciclo de Vida dos Dados	AD pode contribuir com solução tecnológica para alavancar o conteúdo na AI Metadados como Sistema na AI e Elemento na AD
Abordagens	Arquitetural Sistêmica Informacional	-	-
Entregáveis	Relacionados à interface	Relacionados aos dados e ao Ciclo de Vida dos Dados	Repositório Digital de Metadados
Favorece a constituição de	Ecologia Informacional Complexa	Ecologia Complexa de Dados	A AI se utiliza da estruturação da camada de dados, realizada pela AD, para a constituição de ambiente informacional digital e/ou Ecologia Informacional Complexa A AD se utiliza da AI para possibilitar que os dados disponíveis na camada de dados de ambientes informacionais digitais e/ou sistemas de informação possam ser acessados por humanos em uma interface amigável

Fonte: Autoria própria.

A análise possibilita visualizar que a Arquitetura da Informação e a Arquitetura de Dados possuem características próprias e que, embora algumas delas estejam presentes em ambas, são diferentes em suas essências, o que possibilita entendê-las de forma coexistente e com alguns relacionamentos.

Com isso, e considerando que a Arquitetura da Informação e a Arquitetura de Dados possuem objetos distintos, é possível afirmar que a AD não se caracteriza como uma nova abordagem da AI.

Quanto aos demais relacionamentos já explicitados, eles convergem no sentido de se estabelecerem no conteúdo, na estruturação dos dados e na solução tecnológica, considerando sua atuação. Dessa forma, Arquitetura de Dados pode prover estruturação, dados e soluções tecnológicas para alavancar os conteúdos da AI que, por sua vez pode oferecer projeto e interface gráfica para a exibição dos dados em ambientes informacionais digitais e Ecologias Informacionais Complexas, cujo fornecimento e consumo podem ser automatizados utilizando tecnologias e técnicas emergentes. E a conexão se faz entre a AI *bottom-up* e a infraestrutura de dados na AD.

6 CONCLUSÕES

Nesta tese, o que se buscou foi aproximar a expertise e os aportes teóricos da Ciência da Informação ao tratamento da camada de dados dos ambientes informacionais digitais e/ou sistemas de informação e das estruturas de representação da informação, constituída por dados, metadados, bancos de dados e suas formas de coleta/consumo, tratamento, armazenamento, recuperação/fornecimento, disseminação, acesso, uso, reuso e preservação. Assim, aspectos já trabalhados na área podem ser utilizados de forma a expandir o escopo de atuação.

Essa discussão foi realizada no Brasil por Santos e Vidotti (2009), cujas tendências apresentadas foram de fato incorporadas à área e outras ainda estão em implantação, como todo processo de consolidação.

É por meio da infraestrutura de dados que a Arquitetura de Dados se comunica com a Arquitetura da Informação *bottom-up* e *top-down*, à medida que possibilita a estruturação dos dados para consumo via interface gráfica (consumidores humanos) e via interface de saída de dados (consumidores máquinas).

Nesta tese, as discussões realizadas na seção 3 evidenciam que a Arquitetura da Informação é insuficiente para estruturar a camada de dados dos ambientes informacionais digitais, sobretudo por não ser parte do seu escopo o tratamento dos dados, embora discuta superficialmente tesouros, vocabulários controlados e metadados. O que responde à primeira inquietação desta pesquisa.

Destaca-se do conceito de Arquitetura de Dados no contexto da Ciência da Informação, cunhado nesta tese, que a AD atua no projeto de desenvolvimento da infraestrutura de dados e do Ciclo de Vida dos Dados, de forma holística e coesa. Ainda do conceito, os dados são objeto da Arquitetura de Dados que possui como elementos: Ciclo de Vida dos Dados, Metadados, Estruturas Sintática e Semântica dos Dados e Padrões Éticos, Legais e Técnicos aplicados aos Dados, e que possuem elementos relacionados, a exemplo dos processos de identificação, coleta, tratamento, armazenamento, acesso, recuperação e descarte, além de protocolos de comunicação.

Muitos desses elementos são parte do escopo da Ciência da Informação. Deles, destacam-se, inicialmente – visando o relacionamento entre a Arquitetura da Informação e a Arquitetura de Dados –, os tesouros, vocabulários controlados e metadados que são amplamente utilizados na CI, sobretudo no processo de

representação da informação, cuja relevância é indiscutível quando do emprego no campo de valor dos metadados.

Contudo, a discussão apresentada nesta tese possibilita compreender outras formas de tratamento dos metadados pela Ciência da Informação, por exemplo, entendendo e tratando-os também sob o prisma da modelagem de dados, das características do atributo, que, de igual maneira requer olhar criterioso e manipulação de instrumentos como de vocabulários controlados de dados, preferencialmente a partir de padrões de metadados do domínio dos dados.

Neste sentido, o reconhecimento de padrões adequados de metadados para domínios específicos de dados pode e deve ser campo de atuação da Ciência da Informação, área que, desde a sua gênese atua em processos de identificação, coleta, tratamento, armazenamento, recuperação, disseminação e preservação, o que se fazia anteriormente com aparatos analógicos e na atualidade com aparatos digitais.

Ainda no contexto dos metadados, outras literaturas menos triviais nas discussões realizadas na área são incorporadas à tese, a exemplo da ISO/IEC 11179, que se constitui como um importante instrumento para a Arquitetura de Dados e para a Ciência da Informação, à medida que normatiza todas as etapas concernentes ao registro de metadados.

Os metadados, entendidos de forma similar à de um dos sistemas da Arquitetura da Informação, notadamente na AI *bottom-up*, e presentes na Arquitetura de Dados como um dos elementos estruturantes do conceito, possibilitam estabelecer relacionamentos entre a AI e a AD. Neste sentido, a Arquitetura de Dados, conforme conceituada nesta tese, atua no projeto da infraestrutura de dados e do Ciclo de Vida dos Dados, buscando estruturá-los para favorecer a compreensão dos dados por máquinas.

Destaca-se que, no contexto da Arquitetura de Dados, o que se busca é que os dados sejam compreendidos e não apenas acionados, uma vez que dados acionáveis podem ser coletados, mas apenas dados compreensíveis podem ser processados, sobretudo por máquinas.

A compreensão, cuja discussão nesta tese tem início em Wurman (1996, 2000), que já se preocupava em dar sentido a informação para que fosse compreendida nos diferentes ambientes projetados pela Arquitetura da Informação, tem lugar de destaque na Arquitetura de Dados, uma vez que armazenar dados em

um ambiente informacional e/ou sistema de informação é apenas uma parte do processo, que é antecedida e seguida por outras etapas.

Assim, a compreensão evocada na Arquitetura da Informação requer que uma informação faça sentido para seu usuário, claramente entendido como um indivíduo. Isso requer que haja processamento cognitivo para que aquilo que é por ele recebido faça conexões com sua estrutura cognitiva em busca de sentido. Lembrando que o sentido é parte da estrutura semântica que, segundo Dienstbach *et al.* (2021), se relaciona com a referencialidade entre o termo empregado e o objeto, considerando a língua utilizada, e a composicionalidade que se refere à estrutura gramatical da língua na sentença utilizada.

A compreensão entre humanos, embora muitas vezes pareça simples e intuitiva, possui grandes complexidades. Para que se efetive, é preciso que haja, ao menos, compreensão do idioma, da estrutura da sentença no idioma, do termo utilizado como representação do objeto e do próprio objeto para que o termo faça sentido.

De igual maneira, no contexto das máquinas, a compreensão requer que haja, minimamente, protocolo de comunicação de dados, padrão de dados, estrutura sintática e semântica de dados. Ao que, em razão de imposições legais, requer tratamento, considerando princípios éticos e legais para que sejam comunicados, por meio dos princípios técnicos, apenas os dados permitidos (condição de acesso aos dados).

É essa compreensão entre máquinas que requer o tratamento dos dados e dos Ciclos de Vida dos Dados, que pode na Arquitetura de Dados no contexto da Ciência da Informação – por meio do elemento Estruturas Sintática e Semântica dos Dados –, auxiliar na estruturação da camada de dados e reduzir os ruídos de comunicação que possam existir.

Tal fato contribui para otimizar o fluxo de dados entre diferentes Ciclos de Vida dos Dados e favorecer a comunicação dos dados e sua utilização para diferentes finalidades, dentre as quais se destacam: o consumo e o fornecimento de dados entre diferentes CVD, a criação de novos CVD exclusivamente por meio da coleta de dados de múltiplas fontes, a constituição de Ecologias Complexas de Dados, e o consumo por ambientes informacionais digitais e Ecologias Informacionais Complexas.

Tais afirmações possibilitam compreender que o projeto da infraestrutura de dados por meio da Arquitetura de Dados pode contribuir com a Arquitetura da

Informação *bottom-up* e, com isso, alavancar o conteúdo, o que responde à segunda pergunta desta pesquisa.

Ainda no que tange ao relacionamento entre a Arquitetura da Informação e a Arquitetura de Dados, o referencial teórico da pesquisa apresentado nas seções 3 e 4 tornou possível materializá-lo por meio do Quadro 13, que reafirma que ambas se complementam, coexistem e têm como ponto de relacionamento a AI *bottom-up*, na qual estão contidos os elementos que compõem o Sistema de Representação, o que responde à terceira pergunta desta pesquisa.

Quanto aos elementos da Arquitetura de Dados preponderantes para o projeto da infraestrutura de dados, merecem destaque, ainda, o Ciclo de Vida dos Dados, cuja relação com a Ciência da Informação foi realizada por Sant’Ana (2013, 2016) e que reforça a presença da área no fluxo de comunicação de dados; e os Princípios Éticos, Legais e Técnicos, imprescindíveis para o tratamento adequado dos dados.

Tendo em vista o delineamento metodológico da presente tese, a seção 4.3 conduz à compreensão da Teoria do Conceito (DAHLBERG, 1978a, 1978b) – como aparato teórico e técnico –, e possibilita que os resultados, que compõem o polo morfológico, sejam apresentados:

- a) o Conceito de Arquitetura de Dados no contexto da Ciência da Informação (seção 4.4), e os enunciados da AD;
- b) os enunciados presentes na seção 4.4 se desdobram nos Elementos da Arquitetura de Dados no contexto da Ciência da Informação (seção 4.4.1);
- c) a Tríade da Arquitetura de Dados e o Fluxo de Dados na Arquitetura de Dados (seção 4.4.2);
- d) os Entregáveis da Arquitetura de Dados (seção 4.4.3);
- e) o Arquiteto de Dados (seção 4.4.4);
- f) e, a Ecologia Complexa de Dados (seção 4.5).

No percurso da pesquisa conduzida pelo delineamento quadripolar, tem-se, ainda, como achados:

- a) o Reposicionamento da abordagem Pervasiva da Arquitetura da Informação para a abordagem Informacional (seção 3.4.1);
- b) o Reposicionamento da Ecologia Informacional Complexa para a Arquitetura da Informação, decorrente do reposicionamento da abordagem Pervasiva da Arquitetura da Informação para a abordagem Informacional, (seção 3.5);

- c) a compreensão de que, no projeto de Arquitetura da Informação, tendo a informação a característica pervasiva, é necessário um projeto de Arquitetura da Informação macro, aplicável aos ambientes, canais, mídias e tecnologias que compõem a Ecologia Informacional Complexa (seção 3.4.1 e 3.5);
- d) o conceito operacional de Dados (seção 4.1); e,
- e) o conceito operacional de Ecologia Complexa de Dados (seção 4.5).

O percurso da pesquisa possibilitou ainda confirmar a hipótese apresentada: a Arquitetura da Informação é insuficiente para a adequada estruturação da camada dos dados dos ambientes informacionais digitais, o que requer a aproximação com a Arquitetura de Dados.

Na obra de Capurro (2003), a Ciência da Informação, enquadrada em um paradigma físico, vinculado à Recuperação da Informação, pressupõe que um objeto é transmitido por um emissor a um receptor. Oliveira (2014) destaca a ausência do usuário nesse processo. Contudo, não é possível invalidar a afirmação, sobretudo, no que tange ao fluxo necessário à comunicação, relevante até o momento presente.

Por outro lado, o advento das tecnologias e da própria Ciência da Informação permitem observar que seu objeto, tido apenas como informação, está claramente imbuído de dados, quer sejam considerados como a granularidade da informação, ou ainda pelos dados representacionais que exprimem seu contexto. Ainda, tendo em vista o fluxo, o acesso, o uso, o armazenamento, o processamento, a interoperabilidade, a integração, a representação, a recuperação, realizados por meio de dados, utilizando metadados, é possível reforçar os dados como objeto da Ciência da Informação.

Vale destacar que os dados são objeto de estudo do campo de Sistemas de Informação, de forma que a Arquitetura de Dados pode se constituir como elemento de interdisciplinaridade entre a Ciência da Informação e Sistemas de Informação.

Nesse sentido, para Santos e Vidotti (2009):

O uso intenso das Tecnologias de Informação e Comunicação (TICs) exige um direcionamento de inteligência no interior de sua organização. As estruturas de acesso automático, de armazenamento e de preservação da informação para sua localização e uso, requisitam uma fundamentação garantida pela racionalização, pela sistematização e pela compreensão: das estruturas de sinais e de símbolos contidos nos diferentes tipos de dados; da operacionalização no processo de comunicação e de transferência de informações; de linguagens natural e artificial; e de análises semântica e semiótica.

Esse direcionamento requer que as camadas de dados dos ambientes informacionais e sistemas de informação sejam adequadamente estruturadas, o que possibilita ampliar a capacidade de gerar valor por meio dos dados. Dessa forma, o conceito de Arquitetura de Dados no contexto da Ciência da Informação, pode refletir em múltiplos ganhos: à primeira são apresentados elementos de organização, representação, acesso e uso, visando usuários humanos e máquinas; à segunda, são incorporados conceitos, técnicas e tecnologias emergentes.

O mesmo ocorre com a aproximação entre a Arquitetura da Informação e a Arquitetura de Dados, que possibilita o consumo dos dados por meio da AI *bottom-up* cuja exposição ocorre na AI *top-down*. Além disso, a AI e a AD juntas possibilitam o desenvolvimento de interfaces para consumo e fornecimento automático de informações a partir do consumo de dados de uma ou de múltiplas fontes de dados.

De forma mais ampla, as Ecologias Complexas de Dados podem alicerçar processos de comunicação de dados entre múltiplos bancos de dados, favorecendo o enriquecimento, a disponibilização de dados estruturados em uma infraestrutura central, bem como a criação de bancos de dados a partir da coleta e conversão de registros provenientes de múltiplas fontes de dados, a visualização e análise de dados para diferentes finalidades.

Considera-se, assim, que a conceituação da Arquitetura de Dados no contexto da Ciência da Informação é relevante e necessária para a área de Ciência da Informação e beneficia ainda áreas fronteiriças, uma vez que a coleta/criação, o tratamento, o armazenamento, o acesso e o uso de dados estão presentes de forma massiva e fundamental na sociedade contemporânea. Isso posto, a Ciência da Informação, por sua característica social e aplicada, deve atuar ativamente na infraestrutura dos dados para beneficiar usuários – humanos e máquinas.

A presente tese, desenvolvida no âmbito da Informação e Tecnologia, possui potencial de contribuição e impacto social, educacional, tecnológico e econômico, dada a criação de um conceito no contexto da Ciência da Informação. Tal fato reafirma o impacto educacional, ao apresentar a Arquitetura de Dados como um novo campo de pesquisa para a área. No que tange ao impacto tecnológico, há aportes para a estruturação da camada de dados de sistemas de informação e ambientes informacionais digitais – considerando aspectos éticos, legais e técnicos – para que possam ser facilmente localizados, acessados, processados e (re)utilizados em

diferentes contextos, incluindo a constituição de infraestruturas de dados acionáveis e que possam ser úteis para processos diários e para a tomada de decisão. Essa estruturação da camada de dados e a possibilidade de acionamento e processamento automáticos podem trazer impactos econômicos, oriundos, por exemplo, do reuso de recursos existentes e da tomada de decisão quanto a investimentos, políticas públicas, indicadores e ações que contribuam com o desenvolvimento socioeconômico. Neste contexto, os Objetivos do Desenvolvimento Sustentável (ODS) podem ser apoiados com o uso de dados de múltiplas fontes, por meio dos quais desafios globais podem ser mapeados, monitorados e solucionados.

A aproximação realizada por esta tese entre a Arquitetura da Informação, a Arquitetura de Dados e a Ciência da Informação, suscita possibilidades de pesquisa futuras. Sugere-se:

- a) aprofundar estudos relacionados aos Princípios Éticos, Legais e Técnicos aplicados aos Dados;
- b) estudar o relacionamento entre a Arquitetura de Dados e a Arquitetura de Sistemas de Informação;
- c) investigar se e como a Arquitetura de Dados e a Ecologia Complexa de Dados se relacionam à *Data Mesh*;
- d) averiguar a viabilidade da aplicação de técnicas como Inteligência Artificial a sistemas de informação e/ou ambientes informacionais digitais projetados pela Arquitetura de Dados;
- e) projetar a Arquitetura da Informação de um ambiente informacional digital e/ou Ecologia Informacional Complexa que utilize a Arquitetura de Dados na camada de dados;
- f) projetar a Arquitetura da Informação de um ambiente informacional digital e/ou Ecologia Informacional Complexa utilizando Ecologia Complexa de Dados.

Finalmente, reapresenta-se o sonho de Wurman (2000, p. 286, tradução nossa): “Estamos à beira do casamento da tecnologia da informação e arquitetura de informação. Nossa extraordinária capacidade de armazenar e transmitir dados fará desse sonho um sonho acordado.”

Esse sonho consiste da união entre a Tecnologia da Informação e a Arquitetura da Informação no contexto dos dados que, considerando o percurso desta pesquisa, ainda não se efetivou. Destaca-se que, no que concerne ao projeto de

ambientes informacionais digitais, profissionais de TI, AI e CI devem trabalhar juntos. Infere-se, assim, que o problema dessa união, no contexto dados, possa ser decorrente de dois aspectos: por um lado, a TI tem como objeto os dados, sobre os quais exerce formas de tratamento próprias; por outro, a AI tem como objeto a informação e, de igual maneira, exerce sobre ela formas de tratamento próprias.

O desenvolvimento da AI preferencialmente ocorre em equipes multidisciplinares, das quais fazem parte profissionais de TI, AI e CI, e nelas há comunicação, não apenas entre indivíduos, mas envolvendo seus domínios. Assim, pode ser que algo no processo comunicacional esteja falhando, sobretudo se houver dificuldade na compreensão dos elementos mínimos necessários para que seja possível alavancar as discussões acerca da camada de dados dos ambientes informacionais digitais e sua relação com a Arquitetura da Informação.

Essa afirmação pretende mostrar apenas que, assim como no contexto humano, no contexto das máquinas, a comunicação entre domínios também requer compreensão dos elementos necessários para que se constitua.

Com isso, tem-se como pressuposto que o sonho de Wurman possa se materializar por meio do que se conceitua nesta tese como Arquitetura de Dados no contexto da Ciência da Informação. Espera-se que assim ocorra.

REFERÊNCIAS

- ACKOFF, Russell Lincoln. From data to wisdom. **Journal of applied systems analysis**, v. 16, n. 1, p. 3-9, 1989. Disponível em: <http://www-public.imtbs-tsp.eu/~gibson/Teaching/Teaching-ReadingMaterial/Ackoff89.pdf>. Acesso em: 27 jul. 2022.
- ALVES, Rachel Cristina Vesu. **Metadados como elementos do processo de catalogação**. 2010. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Marília, 2010. Disponível em: https://repositorio.unesp.br/bitstream/handle/11449/103361/alves_rcv_dr_mar.pdf?sequence=1&isAllowed=y. Acesso em: 15 jul. 2018.
- AMARAL, Fernando. **Introdução à ciência de dados**: mineração de dados e big data. Rio de Janeiro: Alta Books, 2016.
- AMDAHL, Gene Myron; BLAAUW, Gerrit Anne; BROOKS, F. P. Architecture of the IBM System/360. **IBM Journal of Research and Development**, v. 8, n. 2, p. 87-101, 1964. DOI: <https://doi.org/10.1147/rd.82.0087>. Disponível em: <https://ieeexplore.ieee.org/document/5392210?arnumber=5392210>. Acesso em: 01 dez. 2021.
- ARANGO, Jorge. **Rediscovering Information Architecture**. 2019. Disponível em: <https://jarango.com/2019/12/03/rediscovering-information-architecture/>. Acesso em: 10 nov. 2021.
- ARAÚJO MACHADO, Inês; COSTA, Carlos; SANTOS, Maribel Yasmina. Advancing Data Architectures with data mesh implementations. **Lecture Notes in Business Information Processing**, n. 452, 2022. DOI: https://doi.org/10.1007/978-3-031-07481-3_2. Disponível em: <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1373&context=isd2014>. Acesso em: 08 out. 2022.
- ARQUIVO NACIONAL (BRASIL). **Dicionário brasileiro de terminologia arquivística**. Rio de Janeiro: Arquivo Nacional, 2005. Disponível em: http://www.arquivonacional.gov.br/images/pdf/Dicion_Term_Arquiv.pdf. Acesso em: 23 jun. 2022.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR ISO 9241-210**: ergonomia da interação humano-sistema: parte 210: projeto centrado no ser humano para sistemas interativos. Rio de Janeiro, 2011.
- AUSTRALIAN NATIONAL DATA SERVICE. **FAIR data training**. Disponível em: <https://www.ands.org.au/working-with-data/fairdata/training>. Acesso em: 09 jul. 2020.
- BARBOSA, Cláudia Soares. **Sintaxe do Português**. São Paulo: SAGAH, 2016.
- BERNERS-LEE, Tim. **Axioms of web architecture**: metadata. 2009. Disponível em: <https://www.w3.org/DesignIssues/Metadata>. Acesso em: 4 jun. 2018.

BERNERS-LEE, Tim. **Linked data**. 2006. Disponível em:

<http://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 4 jun. 2018.

BORKO, Harold. Information Science: what is it? **American Documentation**, v. 19, n. 1, p. 3-5, jan. 1968.

BRANDT, Mariana Baptista. **Modelagem da informação legislativa**: Arquitetura da Informação para o processo legislativo brasileiro. 2020. Tese. (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Marília, 2020. Disponível em:

https://repositorio.unesp.br/bitstream/handle/11449/191740/brandt_mb_dr_mar.pdf?sequence=3&isAllowed=y. Acesso em: 15 out. 2021.

BRASIL. Decreto nº 8.777, de 11 de maio de 2016. Institui a Política de Dados Abertos do Poder Executivo federal. **Diário Oficial da União**, Brasília, DF, 12 maio 2016a. Disponível em: http://www.planalto.gov.br/ccivil_03/ato2015-2018/2016/decreto/D8777.htm. Acesso em: 23 jan. 2020.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. **Diário Oficial da União**, Brasília, DF, 18 nov. 2011. Disponível em: http://www.planalto.gov.br/ccivil_03/ato2011-2014/2011/lei/l12527.htm. Acesso em: 23 jan. 2020.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União**: seção 1, Brasília, DF, ano 157, n. 1, p. 59, 15 ago. 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/ato2015-2018/2018/lei/l13709.htm. Acesso em: 29 ago. 2022.

BRASIL. Lei nº 13.853, de 8 de julho de 2019. Altera a Lei nº 13.709, de 14 de agosto de 2018, para dispor sobre a proteção de dados pessoais e para criar a Autoridade Nacional de Proteção de Dados; e dá outras providências. **Diário Oficial da União**, Brasília, DF, 20 dez. 2019. Disponível em: http://www.planalto.gov.br/ccivil_03/ato2019-2022/2019/Lei/L13853.htm#art1. Acesso em: 29 ago. 2022.

BRASIL. Lei nº 14.382, de 27 de junho de 2022. Dispõe sobre o Sistema Eletrônico dos Registros Públicos (Serp); altera as Leis nºs 4.591, de 16 de dezembro de 1964, 6.015, de 31 de dezembro de 1973 (Lei de Registros Públicos), 6.766, de 19 de dezembro de 1979, 8.935, de 18 de novembro de 1994, 10.406, de 10 de janeiro de 2002 (Código Civil), 11.977, de 7 de julho de 2009, 13.097, de 19 de janeiro de 2015, e 13.465, de 11 de julho de 2017; e revoga a Lei nº 9.042, de 9 de maio de 1995, e dispositivos das Leis nºs 4.864, de 29 de novembro de 1965, 8.212, de 24 de julho de 1991, 12.441, de 11 de julho de 2011, 12.810, de 15 de maio de 2013, e 14.195, de 26 de agosto de 2021. **Diário Oficial da União**, Brasília, DF, 28 ago. 2022. Disponível em: http://www.planalto.gov.br/ccivil_03/ato2019-2022/2022/lei/L14382.htm. Acesso em: 03 nov. 2022.

BRASIL. Lei nº 9.279, de 14 de maio de 1996. Regula direitos e obrigações relativos à propriedade industrial. **Diário Oficial da União**, Brasília, DF, 15 maio 1996. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/19279.htm. Acesso em: 03 nov. 2021.

BRASIL. Lei nº 9.609, de 19 de fevereiro de 1998. Dispõe sobre a proteção da propriedade intelectual de programa de computador, sua comercialização no país, e dá outras providências. **Diário Oficial da União**, Brasília, DF, 20 fev. 1998a. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/L9609.htm. Acesso em: 03 nov. 2021.

BRASIL. Lei nº 9.610, de 19 de fevereiro de 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. **Diário Oficial da União**, Brasília, DF, 20 fev. 1998b. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/19610.htm. Acesso em: 03 nov. 2021.

BRASIL. Ministério da Saúde. Conselho Nacional de Saúde. **Resolução MS/CNS nº 466, de 12 de dezembro de 2012**. Aprova as diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. 2012. Disponível em: <https://conselho.saude.gov.br/resolucoes/2012/Reso466.pdf>. Acesso em: 16 ago. 2022.

BRASIL. Ministério da Saúde. Conselho Nacional de Saúde. **Resolução nº 510, de 07 de abril de 2016**. Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais cujos procedimentos metodológicos envolvam a utilização de dados diretamente obtidos com os participantes ou de informações identificáveis ou que possam acarretar riscos maiores do que os existentes na vida cotidiana, na forma definida nesta Resolução. 2016b. Disponível em: <https://conselho.saude.gov.br/resolucoes/2016/Reso510.pdf>. Acesso em: 16 ago. 2022.

BUCKLAND, Michael Keeble. Information as thing. **Journal of the American Society for Information Science**, v. 42, n. 5, p. 351-360, 1991. DOI: [https://doi.org.ez48.periodicos.capes.gov.br/10.1002/\(SICI\)1097-4571\(199106\)42:5<351::AID-ASI5>3.0.CO;2-3](https://doi.org.ez48.periodicos.capes.gov.br/10.1002/(SICI)1097-4571(199106)42:5<351::AID-ASI5>3.0.CO;2-3). Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/%28SICI%291097-4571%28199106%2942%3A5%3C351%3A%3AAID-ASI5%3E3.0.CO%3B2-3>. Acesso em: 27 ago. 2021.

CAMARGO, Liriane Soares de Araújo de; VIDOTTI, Silvana Aparecida Borsetti Gregorio. **Arquitetura da Informação**: uma abordagem prática para o tratamento de conteúdo e interface em ambientes informacionais digitais. Rio de Janeiro: LTC, 2011.

CAPURRO, Rafael. Epistemologia y Ciencia de la Informacion. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 5., 2003, Belo Horizonte. **Anais** [...]. Belo Horizonte: [s. n.], 2003. Disponível em: <http://www.capurro.de/enancib.htm>. Acesso em: 20 jul. 2020.

CARROLL, Stephanie Russo; GARBA, Ibrahim; FIGUEROA-RODRÍGUEZ, Oscar L.; HOLBROOK, Jarita; LOVETT, Raymond; MATERECHERA, Simeon; PARSONS,

Mark; RASEROKA, Kay; RODRIGUEZ-LONEBEAR, Desi; ROWE, Robyn; SARA, Rodrigo; WALKER, Jennifer D.; ANDERSON, Jane; HUDSON, Maui. The CARE principles for indigenous data governance. **Data Science Journal**, v. 19, n. 1, p. 43, nov. 2020a. DOI: <http://doi.org/10.5334/dsj-2020-043>. Disponível em: <https://datascience.codata.org/articles/10.5334/dsj-2020-043/#:~:text=The%20CARE%20Principles%20are%20a,value%20of%20data%20for%20reuse>. Acesso em: 05 fev. 2021.

CARROLL, Stephanie Russo; HUDSON, Maui; HOLBROOK, Jarita; MATERECHERA, Simeon; ANDERSON, Jane. **Working with the CARE principles: operationalizing indigenous data governance**. 2020b. Disponível em: <https://www.adalovelaceinstitute.org/blog/care-principles-operationalising-indigenousdata-governance/>. Acesso em: 06 fev. 2021.

CARVALHO, Carlos Ponce de Leon Ferreira de; LORENA, Ana Carolina. **Introdução à computação: hardware, software e dados**. Rio de Janeiro, 2017.

CONEGLIAN, Caio Saraiva. **Recuperação da informação com abordagem semântica utilizando linguagem natural: a inteligência artificial na Ciência da Informação**. 2020. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Marília, 2020. Disponível em: <http://hdl.handle.net/11449/193051>. Acesso em: 12 maio 2022.

CONEGLIAN, Caio Saraiva; DIEGER, Rodrigo; SANTAREM SEGUNDO, José Eduardo; CAPRETZ, Miriam Akemi Manabe. O papel da web semântica nos processos do big data. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 23, n. 53, p. 137-146, set./dez. 2018. DOI: [10.5007/1518-2924.2018v23n53p137](https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p137). Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p137>. Acesso em: 14 nov. 2021.

CONEGLIAN, Caio Saraiva; ROA-MARTINEZ, Sandra Milena; FERREIRA, Ana Maria Jensen Ferreira da Costa, VIDOTTI, Silvana Aparecida Borsetti Gregório; SANTAREM SEGUNDO, José Eduardo. Tecnologías de la web semántica en Arquitectura de la Información. **Revista Interamericana de Bibliotecología**, [s. l.], v. 42, n. 1, p. 23-35, 2019. DOI: <https://doi.org/10.17533/udea.rib.v42n1a03>. Disponível em: <https://revistas.udea.edu.co/index.php/RIB/article/view/328023>. Acesso em: 14 nov. 2021.

CREATIVE COMMONS BRASIL. **Perguntas frequentes (FAQ)**. Disponível em: <https://br.creativecommons.net/faq/>. Acesso em: 16 set. 2022.

CREATIVE COMMONS. **CC0 1.0 Universal (CC0 1.0) dedicação ao domínio público**. Disponível em: https://creativecommons.org/publicdomain/zero/1.0/deed.pt_BR. Acesso em: 16 set. 2022.

CUNHA, Murilo Bastos da; CAVALCANTI, Cordélia Robalinho de Oliveira. **Dicionário de Biblioteconomia e Arquivologia**. Brasília: Briquet de Lemos, 2008. Disponível em: <https://repositorio.unb.br/handle/10482/34113>. Acesso em: 23 jul. 2022.

DAHLBERG, Ingetraut. A referent-oriented, analytical concept theory for INTERCONCEPT. **International Classification**, Miinchen, v. 5, n. 3, p. 142-151, nov. 1978a. DOI: <https://doi.org/10.5771/0943-7444-1978-3-142>. Disponível em: <https://www.nomos-elibrary.de/10.5771/0943-7444-1978-3-142.pdf>. Acesso em: 15 ago. 2022.

DAHLBERG, Ingetraut. Teoria do conceito. **Ciência da Informação**, Brasília, v. 7, n. 2, p. 101-107, dez. 1978b. DOI: <https://doi.org/10.18225/ci.inf.v7i2.115>. Disponível em: <https://revista.ibict.br/ciinf/article/view/115>. Acesso em: 15 abr. 2022.

DAMA International. **About**. Disponível em: <https://www.dama.org/cpages/mission-vision-purpose-and-goals>. Acesso em: 15 jan. 2022.

DAMA-DMBOK. **Data management body of knowledge**. New Jersey: Technics Publications, 2017.

DAMIANI, A.; MASCIOCCHI, C.; LENKOWICZ, J.; CAPOCCHIANO, N. D.; BOLDRINI, L.; TAGLIAFERRI, L.; CESARIO, A.; SERGI, P.; MARCHETTI, A.; LURASCHI, A.; PATARNELLO, S.; VALENTINI, V. Building an Artificial Intelligence Laboratory based on real world data: the experience of Gemelli Generator. **Frontiers in Computer Science**, v. 3, p. 1-20, dec. 2021. DOI: <https://doi.org/10.3389/fcomp.2021.768266>. Disponível em: <https://www.frontiersin.org/articles/10.3389/fcomp.2021.768266/full>. Acesso em: 01 out. 2022.

DATE, C. J. **Introdução a sistemas de bancos de dados**. São Paulo: Grupo GEN, 2004.

DAVENPORT, Thomas H. **Ecologia da informação**: por que só a tecnologia não basta para o sucesso na era da informação. São Paulo: Futura, 1998.

DE BRUYNE, Paul de; HERMAN, Jacques; DE SCHOUTHEETE, Marc. **Dinâmica da pesquisa em Ciências Sociais**. Rio de Janeiro: Francisco Alves Editora, 1977.

DIENSTBACH, Dalby; VILHAGRA, Leonardo Teixeira de Freitas Ribeiro; KUHLMANN, Mariana Corallo Mello de Azevedo; GUIMARÃES, Sílvia Adélia; BIZELLO, Henrique Aline Azeredo. **Semântica e Pragmática**. Porto Alegre: SAGAH, 2021.

DOW, Kevin E.; HACKBARTH, Gary; WONG, Jeffrey. Data Architectures for an organizational memory information system. **Journal of the American Society for Information Science and Technology**, v. 64, n. 7, p. 1345-1356, jul. 2013. DOI: <https://doi.org/10.1002/asi.22848>. Disponível em: <https://onlinelibrary-wiley.ez87.periodicos.capes.gov.br/doi/10.1002/asi.22848>. Acesso em: 24 jan. 2022.

DOWNING, Douglas A.; COVINGTON, Michael Aaron; COVINGTON, Melody Mauldin. **Dictionary of computer and internet terms**. New York: Barron's Educational Series, 2009.

EUROPEAN COMMISSION. **H2020 Programme**: Guidelines on FAIR Data management in Horizon 2020. 2016. Disponível em:

https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. Acesso em: 12 jul. 2020.

FAIR FUNDERS COLLABORATION. **A pilot programme to make it easy for funders to require and for grantees to produce FAIR data**. 2019. Disponível em: [https://academia.edu/79672187/The FAIR Funder pilot programme to make it easy for funders to require and for grantees to produce FAIR Data](https://academia.edu/79672187/The_FAIR_Funder_pilot_programme_to_make_it_easy_for_funders_to_require_and_for_grantees_to_produce_FAIR_Data). Acesso em: 12 nov. 2020.

FAIRDATABR: uma ferramenta para a avaliação de conjuntos de dados. Disponível em: <https://wrco.ufpb.br/fair/fair+.html#:~:text=Caracteriza%2Dse%20por%20ser%20de,e%20de%20seu%20posterior%20re%C3%BAso>. Acesso em: 12 fev. 2021.

FARRELL, Brenda; BENGTON, Jason. Scientist and data architect collaborate to curate and archive an inner ear electrophysiology data collection. **PLOS ONE**, v. 14, n. 10, p. 1-26, oct. 2019. DOI: <https://doi.org/10.1371/journal.pone.0223984>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0223984>. Acesso em: 02 ago. 2020.

FEDERAL ENTERPRISE ARCHITECTURE PROGRAM. **The data reference model: version 2.0**. 2005. Disponível em: <http://xml.coverpages.org/FEA-DRMv20Final-2005.pdf>. Acesso em: 28 out. 2022.

FLORIDI, Luciano. **Information: a very short introduction**. New York: Oxford University Press, 2010.

FORCE21. **Guiding principles for findable, accessible, interoperable and reusable data publishing version b1.0**. 2021. Disponível em: <https://www.force11.org/fairprinciples#Annex1-1>. Acesso em: 07 fev. 2021.

GILL, Tony. Metadata and the web. *In*: BACA, Murtha (ed.). **Introduction to metadata**. 3. ed. Los Angeles: Getty Publications, 2016. Disponível em: <https://www.getty.edu/publications/intrometadata/setting-the-stage/>. Acesso em: 02 ago. 2020.

GILLILAND, Anne J. Setting the stage. *In*: BACA, Murtha (ed.). **Introduction to metadata**. 3. ed. Los Angeles: Getty Publications, 2016. Disponível em: <https://www.getty.edu/publications/intrometadata/setting-the-stage/>. Acesso em: 02 ago. 2020.

GLOBAL INDIGENOUS DATA ALLIANCE. **CARE principles for indigenous data governance**. 2019. Disponível em: https://static1.squarespace.com/static/5d3799de845604000199cd24/t/5da9f4479ecab221ce848fb2/1571419335217/CARE+Principles_One+Paggers+FINAL_Oct_17_2019.pdf. Acesso em: 05 fev. 2021.

GOFAIR. **FAIR principles**. Disponível em: <https://www.go-fair.org/fair-principles/>. Acesso em: 07 fev. 2021.

GOFAIR. **FAIRification process**. Disponível em: <https://www.go-fair.org/fair-principles/fairification-process/>. Acesso em: 08 ago. 2022.

GUANAES, Paulo Cezar Vieira. **Abertura e compartilhamento de dados de pesquisa subjacentes a artigos científicos**: questões do direito autoral. 2020. Tese. (Doutorado em Ciência da Informação) – Instituto Brasileiro de Informação em Ciência e Tecnologia; Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2020. Disponível em: <https://www.arca.fiocruz.br/handle/icict/46188>. Acesso em: 10 jul. 2021.

HAGEDORN, K. **The Information Architecture glossary**. USA, 2000. Disponível em: https://argus-acia.com/white_papers/ia_glossary.pdf. Acesso em: 02 ago. 2020.

HARFOUCHE, Antoine L.; NAKHLE, Farid; HARFOUCHE, Antoine H.; SARDELLA, Orlando G.; DART, Eli, JACOBSON, Daniel. A primer on artificial intelligence in plant digital phenomics: embarking on the data to insights journey. **Trends in Plant Science**, p. 1-31, 2022. DOI: <https://doi.org/10.1016/j.tplants.2022.08.021>. No prelo. Disponível em: <https://www-sciencedirect.ez87.periodicos.capes.gov.br/science/article/pii/S1360138522002278?via%3Dihub>. Acesso em: 18 out. 2022.

HUSSERL, Edmund. **A ideia da fenomenologia**. Lisboa: Edições 70, 2008.

INGWERSEN, Wesley W.; HAWKINS, Troy R.; TRANSUE, Thomas R.; MEYER, David E.; MOORE, Gary; KAHN, Ezra; ARBUCKLE, Peter; PAULSEN, Heidi; NORRIS, Gregory A. A new data architecture for advancing life cycle assessment. **International Journal of Life Cycle Assessment**, v. 20, n. 4, p. 520-526, abr. 2015. DOI: <https://doi.org/10.1007/s11367-015-0850-6>. Disponível em: <https://link.springer.com/article/10.1007/s11367-015-0850-6>. Acesso em: 18 jan. 2022.

INMON, William H.; LINSTEDT, Daniel. **Data Architecture**: a primer for the data scientist big data, data warehouse and data vault. Massachusetts: Elsevier, 2015.

INMON, William H.; LINSTEDT, Daniel; LEVINS, Mary. **Data Architecture**: a primer for the data scientist. Massachusetts: Elsevier, 2019.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC 11179-1**: information technology: metadata registries (MDR): part 1: framework. London: BSI Standards Limited, 2015a.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC 11179-3 +A1**: information technology: metadata registries (MDR): part 3: registry metamodel and basic attributes. London: BSI Standards Limited, 2020.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC 11179-4**: information technology: metadata registries (MDR): part 4: formulation of data definitions. Switzerland: ISO copyright office, 2004.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC 11179-6**: information technology: metadata registries (MDR): part 5: naming principles. London: BSI Standards Limited, 2015b.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC 2382**: information technology: vocabulary. 2015c. Disponível em: <https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:ed-1:v2:en>. Acesso em: 12 abr. 2022.

IYAMU, Tiko. The architectures of data and information: their confounded confusion. **South African Journal of Information Management**, v. 21, n. 1, p. 1-9, 2019. DOI: <https://doi.org/10.4102/sajim.v21i1.1060>. Disponível em: <https://sajim.co.za/index.php/sajim/article/view/1060>. Acesso em: 06 jun. 2021.

KHOSROW-POUR, Mehdi. **Dictionary of Information Science and Technology**. Pensilvânia: Idea Group Reference, 2007.

LACERDA, Flávia. **Arquitetura da Informação Pervasiva**: projetos de ecossistemas de informação na internet das coisas. 2015. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Brasília, 2015. Disponível em: <https://repositorio.unb.br/handle/10482/19646>. Acesso em: 01 set. 2017.

LAUN, Alexander; MAZZUCHI, Thomas A.; SARKANI, Shahram. Conceptual data model for system resilience characterization. **Systems Engineering**, v. 25, n. 2, p. 115-132, 2022. DOI: <https://doi.org/10.1002/sys.21605>. Disponível em: <https://incose.onlinelibrary.wiley.com/doi/10.1002/sys.21605>. Acesso em: 18 out. 2022.

LEDFOURD, J. L. **Search engine optimization bible**. Indianapolis: Wiley, 2009.

LEVITIN, Anany V.; REDMAN, Thomas C. A model of the data (life) cycles with application to quality. **Information and Software Technology**, v. 35, n. 4, p. 217-223, abr. 1993. DOI: [https://doi.org/10.1016/0950-5849\(93\)90069-F](https://doi.org/10.1016/0950-5849(93)90069-F). Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S095058499390069F>. Acesso em: 14 ago. 2022.

LIGUE DES BIBLIOTHÈQUES EUROPÉENNES DE RECHERCHE. **Implementing FAIR Data Principles**: the role of libraries. Disponível em: <https://libereurope.eu/article/implementing-fair-data-principles-role-libraries/>. Acesso em: 12 nov. 2020

LÓSCIO, Bernadette Farias; BURLE, Caroline; CALEGARI, Newton (ed.). **Boas práticas para dados na web**. 2017. Disponível em: <https://w3c.br/traducoes/DWBP-pt-br/>. Acesso em: 26 abr. 2018.

MAGLIO, Paul P.; SRINIVASAN, Savitha; KREULEN, Jeffrey T.; SPOHRER, Jim. Service systems, service scientists, SSME, and innovation. **Communications of the ACM**, v. 49, n. 7, p. 81-85, jul. 2006. Disponível em: <http://delivery.acm.org/10.1145/1140000/1139955/p81-maglio.pdf?ip=200.134.25.13&id=1139955&acc=ACTIVE%20SERVICE&key=344E943C9DC262BB%2E2EB0389E020F18FC%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=734797945&CFTOKEN=48750344&acm=1488580801965766f122477b0ffd209a62ad3758ab>. Acesso em: 02 fev. 2017.

MAGLIO, Paul P.; VARGO, Stephen L.; CASWELL, Nathan; SPOHRER, Jim. The service system in the basic abstraction of service science. **Inf Syst E-Bus Manage**,

- n. 7, p. 395-406, 2009. Disponível em: <https://link.springer.com/article/10.1007/s10257-008-0105-1>. Acesso em: 02 fev. 2017.
- MALHEIRO, Armando; RIBEIRO, Fernanda. **Paradigmas, serviços e mediações em ciência da informação**. Recife: Nectar, 2011.
- MARTINS, Gilberto de Andrade; THEÓPHILO, Carlos Renato. **Metodologia da Investigação Científica para Ciências Sociais Aplicadas**. São Paulo: Atlas, 2016.
- MÉNDEZ RODRÍGUEZ, Eva Maria. **Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales**. Gijón, Asturias: Ediciones Trea, 2002.
- MOREIRA, Daniel Augusto. **O método fenomenológico na pesquisa**. São Paulo: Thomson, 2002.
- MORVILLE, Peter. **In defense of search**. 2001. Disponível em: https://semanticstudios.com/in_defense_of_search/. Acesso em: 09 jan. 2022.
- MORVILLE, Peter; ROSENFELD, Louis. **Information architecture for the world wide web**. Sebastopol, CA: O'Reilly, 2006.
- NORTE, Mariângela Braga. **Glossário de termos técnicos em Ciência da Informação: inglês/português**. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2010. Disponível em: <https://www.marilia.unesp.br/Home/Publicacoes/glossario.pdf>. Acesso em: 24 jun. 2022.
- OLIVEIRA, Henry Poncio Cruz de. **Arquitetura da Informação Pervasiva: contribuições conceituais**. 2014. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Marília, 2014. Disponível em: <http://hdl.handle.net/11449/110387>. Acesso em: 01 set. 2017.
- OLIVEIRA, Henry Poncio Cruz de; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Dos ambientes informacionais às ecologias informacionais complexas. **Informação & Sociedade: Estudos**, v. 26, n. 1, p. 91-101, jan./abr. 2016. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/29438>. Acesso em: 27 nov. 2019.
- OLIVEIRA, Henry Poncio Cruz de; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Arquitetura da Informação digital: conexões interdisciplinares dentro da abordagem sistêmica. In: CAVALCANTE, Lidia Eugenia; PINTO, Virgínia Bentes; VIDOTTI, Silvana Aparecida Borsetti Gregorio. **Ciência da Informação e contemporaneidade: tessituras e olhares**. Fortaleza: Edições UFC, 2012. p. 184-202.
- OPEN KNOWLEDGE FOUNDATION. **A fair, free and open future**. Disponível em: <https://okfn.org/>. Acesso em: 20 set. 2022a.
- OPEN KNOWLEDGE FOUNDATION. **Open Data Commons: legal tools for open data**. Disponível em: <https://opendatacommons.org/>. Acesso em: 20 set. 2022b.

OTTO, Boris. How to design the master Data Architecture: findings from a case study at Bosch. **International Journal of Information Management**, v. 32, n. 4, p. 337-346, ago. 2012. DOI: <https://doi.org/10.1016/j.ijinfomgt.2011.11.018>. Disponível em: <https://www-sciencedirect.ez87.periodicos.capes.gov.br/science/article/pii/S0268401211001435?via%3Dihub>. Acesso em: 21 jan. 2022.

PAKE, George E. Research at Xerox PARK: a founder's assessment. **IEEE Spectrum**, v. 22, n. 10, p. 54-61, oct. 1985. DOI: <https://doi.org/10.1109/MSPEC.1985.6370843>. Disponível em: <https://ieeexplore.ieee.org/document/6370843>. Acesso em: 01 dez. 2021.

PINHEIRO, Lena Vania. Processo evolutivo e tendências contemporâneas da Ciência da Informação. **Informação & Sociedade: Estudos**, João Pessoa, v. 15, n. 1, p. 13-48, jan./jun. 2005. Disponível em: <https://ridi.ibict.br/bitstream/123456789/23/1/1%26SPinheiro2005.PDF>. Acesso em: 06 jul. 2022.

REITZ, Joan M. **ODLIS**: online dictionary for Library and Information Science. 2013. Disponível em: https://products.abc-clio.com/ODLIS/odlis_about. Acesso em: 23 jun. 2022.

RESMINI, Andrea; ROSATI, Luca. **Pervasive Information Architecture**: designing cross-channel user experiences. Burlington: Elsevier, 2011.

RIBEIRO, Anna Carolina Mendonça Lemos; SANTOS, Carlos Denner dos. Isso não é uma pirâmide: revisando o modelo clássico de dado, informação, conhecimento e sabedoria. **Ciência da Informação**, Brasília, v. 49, n. 2, 2020. Disponível em: <https://revista.ibict.br/ciinf/article/view/5066>. Acesso em: 27 jun. 2022.

RILEY, Jenn. **Understanding metadata**: what is metadata, and what is it for? Baltimore: National Information Standards Organization, 2017.

RONDA LEÓN, Rodrigo. Arquitectura de Información: análisis histórico-conceptual. **No Solo Usabilidad**, n. 7, 2008. Disponível em: http://www.nosolousabilidad.com/articulos/historia_arquitectura_informacion.htm. Acesso em: 14 dez. 2021.

ROSENFELD, Louis. Bandleaders in the idea business: in conversation with Lou Rosenfeld. [Entrevista cedida a] Andrea Resmini. In: RESMINI, Andrea; RICE, Sarah A.; IRIZARRY, Bernadette (ed.). **Advances in Information Architecture**: the academics/practitioners roundtable 2014–2019. Suíça: Springer, 2021. p. 75-87.

ROSENFELD, Louis; MORVILLE, Peter. **Information Architecture for the world wide web**. Sebastopol, CA: O'Reilly Media Inc., 1998.

ROSENFELD, Louis; MORVILLE, Peter; ARANGO, Jorge. **Information Architecture**: for the web and beyond. Canada: O'Reilly, 2015.

ROWLEY, Jennifer. The wisdom hierarchy: representations of the DIKW hierarchy. **Journal of Information Science**, v. 33, n. 2, p. 163-180 2007. DOI: <https://doi.org/10.1177/0165551506070706>. Disponível em:

<https://journals.sagepub.com/doi/10.1177/0165551506070706>. Acesso em: 28 jun. 2022.

ROWLEY, Jennifer. What is information? **Information Services & Use**, v. 18, n. 4, p. 243-254, 1998. DOI: <https://doi.org/10.3233/ISU-1998-18402>. Disponível em: <https://search-ebSCOhost-com.ez87.periodicos.capes.gov.br/login.aspx?direct=true&db=aph&AN=1935949&lang=pt-br&site=ehost-live>. Acesso em: 28 jun. 2022.

SANT'ANA, Ricardo César Gonçalves. Ciclo de vida dos dados e o papel da ciência da informação. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais [...]**. Florianópolis, 2013. Disponível em: <http://www.brapci.inf.br/index.php/article/download/40488>. Acesso em: 14 set. 2018.

SANT'ANA, Ricardo César Gonçalves. Ciclo de vida dos dados: uma perspectiva a partir da Ciência da Informação. **Informação & Informação**, Londrina, v. 21, n. 2, p. 116-142, maio/ago. 2016. DOI: <https://doi.org/10.5433/1981-8920.2016v21n2p116>. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940>. Acesso em: 14 set. 2018.

SANT'ANA, Ricardo César Gonçalves. Transdução informacional: impactos do controle sobre os dados. *In*: MARTÍNEZ-ÁVILA, Daniel; SOUZA, Edna Alves de; GONZALEZ, Maria Eunice Quilici (ed.) **Informação, conhecimento, ação autônoma e big data: continuidade ou revolução?** Marília: Oficina Universitária; São Paulo: Cultura Acadêmica; FiloCzar, 2019. p. 117-128. DOI: <https://doi.org/10.36311/2019.978-85-7249-055-9.p117-128>. Disponível em: <http://books.scielo.org/id/gfrbh/pdf/martinez-9788572490559-09.pdf>. Acesso em: 20 ago. 2022.

SANTOS, Plácida Leopoldina Ventura Amorim da Costa; SANTANA, Ricardo César Gonçalves. Dado e granularidade na perspectiva da informação e tecnologia: uma interpretação pela Ciência da Informação. **Ciência da Informação**, Brasília, v. 42, n. 2, 2013. DOI: <https://doi.org/10.18225/ci.inf.v42i2.1382>. Disponível em: <https://revista.ibict.br/ciinf/article/view/1382>. Acesso em: 5 jun. 2022.

SANTOS, Plácida Leopoldina Ventura Amorim da Costa; SANTANA, Ricardo César Gonçalves. Transferência da informação: análise para valoração de unidades de conhecimento. **DataGramZero**, v. 3, n. 2, abr. 2002. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/5349>. Acesso em: 5 jun. 2022.

SANTOS, Plácida Leopoldina Ventura Amorim da Costa; VIDOTTI, Silvana Aparecida Borsetti Gregorio Vidotti. Perspectivismo e tecnologias de informação e comunicação: acréscimos à Ciência da Informação? **DataGramZero**, v. 10, n. 3, 2009. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/6513>. Acesso em: 16 jul. 2020.

SAYÃO, Luis Fernando; SALES, Luana. Afinal, o que é dado de pesquisa? **Biblos: Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 34, n. 2, p. 32-51, jul./dez. 2020. DOI: <https://doi.org/10.14295/biblos.v34i2.11875>. Disponível

em: <https://periodicos.furg.br/biblos/article/view/11875/8426>. Acesso em: 16 jul. 2021.

SETZER, Valdemar Waingort. Dado, informação, conhecimento e competência. **DataGramaZero**, v. 0, n. 0, 1999. Disponível em: <https://www.ime.usp.br/~vwsetzer/datagrama.html>. Acesso em: 03 jul. 2022.

SHANNON, Claude Elwood. A mathematical theory of communication. **The Bell System Technical Journal**, v. 27, jul./oct. 1948.

SILVA, Armando Malheiro da. **A informação**: da compreensão do fenômeno e construção do objeto científico. Porto: Ed. Afrontamento, 2006.

SILVA, Armando Malheiro da. Entrevista: Armando Malheiro da Silva. [Entrevista cedida a] Moisés Rockembach. **Em Questão**, Porto Alegre, v. 23, n. 2, p. 11-33, maio/ago. 2017. DOI <http://dx.doi.org/10.19132/1808-5245232.11-33>. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/70025>. Acesso em: 03 jul. 2022.

SILVA, Armando Malheiro da; RIBEIRO, Fernanda. **Das <<ciências>> documentais à Ciência da Informação**: ensaio epistemológico para um novo modelo curricular. Porto: Ed. Afrontamento, 2002.

SIQUEIRA, André Henrique de. **Arquitetura da Informação**: uma proposta para fundamentação e caracterização da disciplina científica. 2012. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Brasília, 2012. Disponível em: <https://repositorio.unb.br/handle/10482/12157>. Acesso em: 01 ago. 2022.

SOARES, Filipi Miranda; MACULAN, Benildes Coura Moreira dos Santos; DRUCKER, Debora Pignatari; SARAIVA, Antonio Mauro. Methodological principles to create a metadata extension to the Darwin Core standard for agrobiodiversity data. **Brazilian Journal of Information Science**, v. 14, n. 4, 2020. DOI: <https://doi.org/10.36311/1940-1640.2020.v14n4.10865>. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/10865>. Acesso em: 07 fev. 2022.

STONE, Paul; CALDERON, Ania. **[Spotlight] CARE Principles**: unpacking indigenous data governance. 2019. Disponível em: <https://opendatacharter.medium.com/spotlight-care-principles-f475ec2bf6ec>. Acesso em: 06 fev. 2021.

STONIER, Tom. **Information and meaning**: an evolutionary perspective. New York: Springer, 1997.

TORINO, Emanuelle; CONEGLIAN, Caio Saraiva; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Estruturas de representação para reuso de dados no contexto da ecologia de pesquisa: CRIS Institucional. **Informação & Informação**, Londrina, v. 25, n. 3, p. 1-27, jul./set. 2020. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/41946>. Acesso em: 03 nov. 2020.

TORINO, Emanuelle; ROA-MARTÍNEZ, Sandra Milena; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Dados de pesquisa: disponibilização ou publicação?. *In*: SHINTAKU, Milton; SALES, Luana Farias; COSTA, Michelli (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 183-201. DOI: <https://doi.org/10.21452/978-85-93910-04-3.cap15>. Disponível em: <https://repositorio.utfpr.edu.br/jspui/bitstream/1/4725/1/dadospesquisapublicacao.pdf>. Acesso em: 03 nov. 2020.

TORINO, Emanuelle; VIDOTTI, Silvana Aparecida Borsetti Gregorio; SANT'ANA, Ricardo César Gonçalves. Ciclo de vida de dados no processo de publicação e acesso à produção científica. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 20. 2019, Florianópolis. **Anais** [...]. Florianópolis: Universidade Federal de Santa Catarina, 2019. Disponível em: <https://conferencias.ufsc.br/index.php/enancib/2019/paper/viewFile/576/612>. Acesso em: 15 out. 2019.

TUPPER, Charles D. **Data architecture: from zen to reality**. Amsterdam: Morgan Kaufmann, 2011.

UNITED STATES. EXECUTIVE OFFICE OF THE PRESIDENT. **Federal enterprise architecture framework: version 2**. 2013. Disponível em: https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/egov_docs/fea_v2.pdf. Acesso em: 04 mar. 2022.

UNITED STATES. EXECUTIVE OFFICE OF THE PRESIDENT. **The common approach to federal enterprise architecture**. 2012. Disponível em: https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/egov_docs/common_approach_to_federal_ea.pdf. Acesso em: 04 mar. 2022.

VECHIATO, Fernando Luiz. **Encontrabilidade da Informação: contributo para uma conceituação no campo da Ciência da Informação**. 2013. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Marília, 2013. Disponível em: <http://hdl.handle.net/11449/103365>. Acesso em: 01 ago. 2017.

VECHIATO, Fernando Luiz; FARIAS, Gabriela Belmont de. Serendipidade no contexto da Ciência da Informação: perspectivas para os estudos com sujeitos informacionais. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 25, p. 01-23, 2020. DOI: <https://doi.org/10.5007/1518-2924.2020.e72056>. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2020.e72056>. Acesso em: 21 mar. 2022.

VECHIATO, Fernando Luiz; VIDOTTI, Silvana Aparecida Borsetti Gregório. **Encontrabilidade da Informação**. São Paulo: Cultura Acadêmica, 2014. Disponível em: <http://hdl.handle.net/11449/126218>. Acesso em: 01 ago. 2017.

VIDOTTI, Silvana Aparecida Borsetti Gregório; CUSIN, César Augusto; CORRADI, Jiliane Adne Mesa. Acessibilidade digital sob o prisma da Arquitetura da Informação. *In*: GUIMARÃES, José Augusto Chaves; FUJITA, Mariângela Spotti Lopes (org.). **Ensino e pesquisa em Biblioteconomia no Brasil: a emergência de um novo olhar**. São Paulo: Cultura Acadêmica, 2008.

VIDOTTI, Silvana Aparecida Borsetti Gregorio; TORINO, Emanuelle; CONEGLIAN, Caio Saraiva. #SejaJUSTOeCUIDADOSO: princípios FAIR e CARE na gestão de dados de pesquisa. *In*: SALES, Luana Farias; VEIGA, Viviane dos Santos; HENNING, Patrícia; SAYÃO, Luís Fernando (org.). **Princípios FAIR aplicados à gestão de dados de pesquisa**. Rio de Janeiro: Ibict, 2021. p. 201-214. DOI: <https://doi.org/10.22477/9786589167242.cap15>. Disponível em: <https://ridi.ibict.br/handle/123456789/1182>. Acesso em: 14 fev. 2022.

VITRUVIUS POLLIO. **Tratado de Arquitetura**. São Paulo: Martins Fontes, 2007.

WACHOWICZ, Marcos. **A proteção jurídica das bases de dados em face da revolução da tecnologia da informação**. 2014. Disponível em: <https://www.gedai.com.br/wp-content/uploads/2014/07/artigo-base-dados-marcos-wachowicz-1.pdf>. Acesso em: 16 ago. 2022.

WILKINSON, Mark D.; DUMONTIER, Michel; AALBERSBERG, IJsbrand Jan *et al.* The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, v. 3, 160018, mar. 2016. DOI: <https://doi.org/10.1038/sdata.2016.18>. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 06 fev. 2021.

WOODLEY, Mary S. Metadata matters: connecting people and information. *In*: BACA, Murtha (ed.). **Introduction to metadata**. 3. ed. Los Angeles: Getty Publications, 2016. Disponível em: <https://www.getty.edu/publications/intrometadata/metadata-matters/#crosswalk-example>. Acesso em: 02 ago. 2020.

WURMAN, Richard Saul. **Ansiedade da informação**. São Paulo: Cultura Editores Associados, 1991.

WURMAN, Richard Saul. IA Summit 10: Richard Saul Wurman keynote. **Boxes and Arrows**, abr. 2010. Disponível em: <https://boxesandarrows.com/ia-summit-10-richard-saul-wurman-keynote/>. Acesso em: 10 mar. 2022.

WURMAN, Richard Saul. **Information anxiety 2**. Indianapolis: Que, 2000.

WURMAN, Richard Saul. **Information architects**. Zurich: Graphis Press Corp, 1996.

WURMAN, Richard Saul; KATZ, Joel. Beyond graphics: the Architecture of Information. **AIA Journal**, p. 40;60, oct. 1975. Disponível em: <https://usmodernist.org/AJ/AJ-1975-10.pdf>. Acesso em: 06 mar. 2022.

YOON, Victoria Youngohc; AIKEN, Peter; GUIMARAES Tor. Applying a metadata framework to improve data quality. *In*: KHOSROW-POUR, Mehdi (ed.). **Encyclopedia of Information Science and Technology**. Hershey, PA: Idea Group Reference, 2005. v. 1, p. 146-151.