



UNIVERSIDADE DE BRASÍLIA (UnB)
FACULDADE DE CIÊNCIA DA INFORMAÇÃO (FCI)
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO (PPGCINF)

THIAGO ALMEIDA RODRIGUES BORGES

**ARQUIVOS E WEB SEMÂNTICA:
ANÁLISE DOS IMPACTOS DO MODELO CONCEITUAL *RECORDS IN CONTEXTS*
(*RIC*) NA DESCRIÇÃO ARQUIVÍSTICA**

BRASÍLIA

2023

THIAGO ALMEIDA RODRIGUES BORGES

**ARQUIVOS E WEB SEMÂNTICA:
ANÁLISE DOS IMPACTOS DO MODELO CONCEITUAL *RECORDS IN CONTEXTS*
(*RIC*) NA DESCRIÇÃO ARQUIVÍSTICA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Ciência da Informação da Universidade de Brasília, como requisito parcial para obtenção do título de Mestre em Ciência da Informação.

Orientadora: Prof.^a Dra. Cynthia Roncaglio.

BRASÍLIA

2023

B732a Borges, Thiago Almeida Rodrigues
 Arquivos e web semântica : análise dos impactos do modelo
 conceitual Records in Contexts (RIC) na descrição
 arquivística / Thiago Almeida Rodrigues Borges; orientador
 Cynthia Roncaglio. -- Brasília, 2023.
 137 p.

 Dissertação (Mestrado em Ciência da Informação) --
 Universidade de Brasília, 2023.

 1. Records in Contexts (RiC). 2. Descrição arquivística.
 3. Web semântica. 4. Linked Open Data (LOD). I. Roncaglio,
 Cynthia, orient. II. Título.

UNIVERSIDADE DE BRASÍLIA

PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Ata Nº: 38

Aos dezenove dias do mês de dezembro do ano de dois mil e vinte e três, instalou-se a banca examinadora de Dissertação de Mestrado do aluno Thiago Almeida Rodrigues Borges, matrícula 210007192. A banca examinadora foi composta pelos professores Dr. Dalton Lopes Martins / membro interno / PPGCINF/UnB, Dr. Vitor Manoel Marques da Fonseca / Membro externo / UFF, Dr. Thiago Henrique Bragato Barros / UFRGS, Suplente e Dra. Cynthia Roncaglio / orientadora/presidente / PPGCINF/UnB. O discente apresentou o trabalho intitulado "Arquivos e Web semântica: análise dos impactos do modelo conceitual *Records in Contexts* (RiC) na descrição arquivística".

Concluída a exposição, procedeu-se a arguição do candidato, e após as considerações dos examinadores o resultado da avaliação do trabalho foi:

() Pela aprovação do trabalho;

(X) Pela aprovação do trabalho, com revisão de forma, indicando o prazo de até 30 dias para apresentação definitiva do trabalho revisado;

() Pela reformulação do trabalho, indicando o prazo de **(Nº DE MESES)** para nova versão;

() Pela reprovação do trabalho, conforme as normas vigentes na Universidade de Brasília.

Conforme os Artigos 34, 39 e 40 da Resolução 0080/2021 - CEPE, o(a) candidato(a) não terá o título se não cumprir as exigências acima.

Dra. Cynthia Roncaglio, PPGCINF/UnB
(Presidente/orientador)

Dr. Dalton Lopes Martins, PPGCINF/UnB
(Membro interno)

Dr. Vitor Manoel Marques da Fonseca, UFF
(Membro externo)

Dr. Thiago Henrique Bragato Barros, UFRGS
(Suplente)

Thiago Almeida Rodrigues Borges
(Mestrando)



Documento assinado eletronicamente por **Cynthia Roncaglio, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 06/02/2024, às 19:59, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Dalton Lopes Martins, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 07/02/2024, às 09:18, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Thiago Almeida Rodrigues Borges, Usuário Externo**, em 07/02/2024, às 09:48, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Clovis Carvalho Britto, Coordenador(a) da Pós-Graduação da Faculdade de Ciência da Informação**, em 07/02/2024, às 10:42, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Vitor Manoel Marques da Fonseca, Usuário Externo**, em 07/02/2024, às 11:53, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



A autenticidade deste documento pode ser conferida no site http://sei.unb.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **10889874** e o código CRC **2D6FACD2**.

AGRADECIMENTOS

À Cynthia Roncaglio, pela parceria e orientação, assim como por me estimular a investigar um tema inovador para a arquivística brasileira.

Ao Vitor Manoel Marques da Fonseca e Dalton Lopes Martins, pelas contribuições enriquecedoras e críticas construtivas feitas ao longo da pesquisa.

À Eveline de Abreu, pela edição e revisão crítica do texto, que tanto aprimoraram a escrita desta dissertação, e pelos momentos de bom-humor que suavizaram os períodos de tensão.

Ao Rafael Barcelos, pelo apoio na formatação desta dissertação.

Aos colegas Ester Eiko Duarte Kimura, Alexandre Olimpio Barbacena e Mariana Sande Marques Moreira, por dividir as suas experiências no mestrado da FCI/UnB.

Ao arquivista e amigo Antoine Meissonnier, pelas valiosas trocas e pela inspiração na escolha do tema deste mestrado.

Às arquivistas e amigas Marta Lira Oliveira Veras, Vania Lucia Alheiro Rosa e Ligia Cristina Pinheiro da Silva, pela convivência e preciosos ensinamentos no Arquivo Histórico da Câmara dos Deputados.

Aos colegas Regina Rocha, Elias Eliger, Bruna dos Santos, Tailane Fonseca, Paloma Mesquita, David Vasconcelos, Elaine de Lima, Marcia Pereira, Renata Fernandes e demais profissionais do Arquivo da Câmara dos Deputados, a minha admiração e reconhecimento pelo compromisso diário com um serviço público de excelência.

Ao Daniel Facci, pela escuta.

À Mari Brandt, agradeço pela generosidade, amizade sincera e pelas conversas que enriqueceram este percurso.

Aos amigos Gui, Luiz, Jana, Jared, Gê, Diego, Anésio, Marcelo e Bárbara, por tornarem cada momento desta jornada acadêmica mais agradável.

Às minhas irmãs Dessa e Pri, pelo amor e pelo significado que têm em minha vida. Ao Cássio, Léo e Julia, por me fazerem amar ainda mais a nossa família.

Aos sobrinhos Vinícius, Laura, Victor e Gustavo, pelos sorrisos e momentos de alegria.

À minha mãe, por ser fonte de incentivo aos estudos e a minha maior referência de vida.

Ao meu pai (*in memoriam*), amor eterno, por participar de cada conquista, ainda que em outro plano.

À minha avó Maria (Bodoquinha), pelo amor mais precioso, com imensa saudade.

RESUMO

Ao considerar os novos pontos de vista sobre a descrição arquivística, confirmada pela adesão do Conselho Internacional de Arquivos (ICA) à Web semântica por meio do desenvolvimento do modelo conceitual *Records in Contexts (RiC)*, esta pesquisa busca preencher, em parte, a falta de sistematização nos estudos sobre os benefícios que tal inovação tecnológica oferece aos arquivos, bem como compreender a introdução de conceitos emergentes, tais como a descrição multidimensional e a visão expandida do princípio da proveniência. O estudo se dedica à análise e à sistematização tanto das renovações em termos de descrição arquivística quanto do ecossistema das tecnologias semânticas incorporadas ao modelo *RiC*, incluindo o modelo Entidade-relacionamento (ER), a própria Web semântica e o *Linked Open Data (LOD)*, junto aos seus respectivos padrões e protocolos. Investiga-se, também, o estado da arte da aplicação do modelo *RiC* em iniciativas internacionais. Por intermédio de uma análise exploratória e qualitativa, utilizando-se de técnicas como a análise de conteúdo e a aplicação de questionários, constata-se que, embora o modelo *RiC* apresente dificuldades de implementação, quando confrontado com as normas de descrição do ICA, e exija ampla qualificação dos arquivistas e profissionais da área, ele proporciona vantagens substanciais para a estruturação de novas entidades e para o mapeamento de seus inter-relacionamentos. No ambiente dos documentos digitais, o *RiC* revela-se como um modelo que intensifica a preservação de contextos inerentes à produção, manutenção e uso de arquivos, reforça o acesso e amplia a utilização e a reutilização dos dados arquivísticos.

Palavras-chave: *Records in Contexts (RiC)*; descrição arquivística; web semântica; *Linked Open Data (LOD)*.

ABSTRACT

Considering new approaches in archival description, highlighted by the International Council on Archives (ICA) adoption of the semantic web through the development of the conceptual data model Records in Contexts (RiC), the research addresses the lack of systematization in studies on the benefits that such technological innovation offers to archives, as well as the introduction of emerging concepts such as multidimensional description and an expanded view of the principle of provenance. The study is dedicated to the analysis and systematization of both the significant updates in archival description and the ecosystem of semantic technologies incorporated into the RiC model, including the Entity-Relationship (ER) model, the semantic web itself, and Linked Open Data (LOD), along with their respective standards and protocols. The state of the art of the RiC model's application in international initiatives is also investigated. By employing an exploratory and qualitative approach, using techniques such as content analysis and the application of questionnaires, the study observes that while the RiC model presents implementation challenges compared to the ICA's traditional description standards and requires new qualifications for archivists and professionals in the field, it offers substantial advantages for the structuring of new entities and mapping their interrelationships. In the realm of digital records, the RiC emerges as a potent technology. It enhances the preservation of contexts inherent to the production, maintenance, and use of archives, thereby strengthening access and expanding the possibilities for the use and reuse of archival data.

Keywords: Records in Contexts (RiC); archival description; semantic web; Linked Open Data (LOD).

LISTA DE FIGURAS

Figura 1: Estimativa de crescimento de dados de 2020 a 2025 em Zettabytes, segundo a IDC	25
Figura 2: Previsão do armazenamento de dados até 2025, segundo o relatório Data Age 2025 da IDP	26
Figura 3: Representação de estratos de localização dos dados.....	27
Figura 4: Mundo digital em 2019, segundo a Global Digital Overview	28
Figura 5: Conjunto de páginas conectadas por links na Web de documentos (1.0).....	34
Figura 6: Conjunto de dados conectados, com semântica, na Web de dados (3.0).....	36
Figura 7: Esquema de publicação de dados abertos de acordo com a escala de cinco estrelas	38
Figura 8: Descrição de um dossiê do fundo da Constituinte de 1823 em formato PDF (uma estrela)	39
Figura 9: Descrição do mesmo dossiê do fundo da Constituinte de 1823, em formato CSV, padrão EAD (três estrelas).....	40
Figura 10: Descrição do mesmo dossiê do fundo da Constituinte de 1823 em formato RDF (quatro estrelas)	42
Figura 11: Exemplo de grafo de dados do Arquivo Nacional da França com fundos, documentos e autoridades relacionadas.....	45
Figura 12: Representação abstrata da descrição em RDF	46
Figura 13: Representação concreta do RDF para o exemplo do requerimento de Maria Joaquina	46
Figura 14: Representação gráfica da descrição de um documento de arquivo, segundo o modelo RiC.....	54
Figura 15: Perfis de expressividade das versões OWL 1 e OWL 2	58
Figura 16: Perda desde a percepção do mundo até a representação do conhecimento	59
Figura 17: Modelo Entidade-relacionamento (ER) aplicado aos arquivos no modelo RiC	62
Figura 18: Pilha de tecnologias da Web semântica.....	64
Figura 19: Representação comparativa de dados em estrutura hierárquica, como XML (à esquerda) e estrutura em triplas, como grafos (à direita)	72
Figura 20: Evolução das normas e modelos conceituais em descrição arquivística	78
Figura 21: Relacionamentos entre entidades do modelo RiC	89

LISTA DE GRÁFICOS

Gráfico 1: Tipos de instituições que aplicam o modelo RiC	108
Gráfico 2: Estágios de implementação dos projetos do modelo RiC	109
Gráfico 3: Duração dos projetos de aplicação do modelo RiC.....	109
Gráfico 4: Visão sobre o princípio da proveniência nas instituições que aplicam o modelo RiC	113
Gráfico 5: Fases da Web nos sistemas de descrição das instituições que aplicam o modelo RiC	115
Gráfico 6: Motivações das instituições para criação dos projetos de aplicação do modelo RiC	117
Gráfico 7: Benefícios obtidos pelos projetos de aplicação do modelo RiC	118
Gráfico 8: Contexto arquivístico no modelo RiC mais representativo que nas normas de descrição do ICA	120
Gráfico 9: Estimativa de uso do modelo RiC pela comunidade arquivística	121
Gráfico 10: Nível de institucionalidade de projetos de implementação do RiC.....	121

LISTA DE QUADROS

Quadro 1: Comparativo entre o modelo atual centrado em aplicações e o modelo futuro centrado em dados	29
Quadro 2: Principais características de cada fase da Web	37
Quadro 3: Classificação dos conjuntos de dados do portal data.gov.uk, segundo o esquema de classificação cinco estrelas	43
Quadro 4: Formatos de serialização do <i>RDF</i> e suas características de uso.....	48
Quadro 5: Uso de ontologias por campos científicos.....	50
Quadro 6: Cronologia do desenvolvimento das normas internacionais de descrição arquivística	66
Quadro 7: Entidades do modelo <i>RiC</i> em quatro níveis de hierarquia	84
Quadro 8: Exemplo de quadro explicativo de entidade no modelo <i>RiC</i>	85
Quadro 9: Exemplo de quadro explicativo do atributo classificação.....	87
Quadro 10: Exemplo de quadro de atributos da entidade documento	88
Quadro 11: Lista geral de fontes de referência do modelo <i>RiC</i> em que foi feita a leitura flutuante, na etapa de seleção de documentos para a análise de conteúdo	92
Quadro 12: Lista final de fontes selecionadas para a análise de conteúdo	96
Quadro 13: Resumo das etapas da seleção de documentos para a análise de conteúdo	97
Quadro 14: Exemplo de trecho de análise e identificação de temas	98
Quadro 15: Sistematização das críticas à descrição e à teoria arquivísticas	100
Quadro 16: Sistematização das críticas às tecnologias em uso nos arquivos	102
Quadro 17: Sistematização das soluções do modelo <i>RiC</i> para a descrição arquivística	103
Quadro 18: <i>Ranking</i> de soluções do <i>RiC</i> bem-sucedidas nos projetos de aplicação do modelo	119

LISTA DE ABREVIATURAS E SIGLAS

AI/ML	Inteligência Artificial/Aprendizado de Máquina
API	Application for Programming Interface
AR/ML	Realidade Aumentada e Virtual
CC BY	Creative Commons Attribution
CC0	Creative Commons Zero
CERN	Centro Europeu de Pesquisas Nucleares
CGI	Common Gateway Interface
CIDOC	Comitê Internacional para a Documentação do Conselho Internacional de Museus
CNEDA	Comissão de Normas Espanholas de Descrição Arquivística
CRM	Conceptual Reference Model
CSV	Comma-Separated Values
DFKI	Centro Alemão de Pesquisa em Inteligência Artificial
DPLA	Digital Public Library of America
EAD	Encoded Archival Description
EGAD	Grupo de Especialistas em Descrição Arquivística
ER	Entidade-relacionamento
FH;P	Universidade de Ciências Aplicadas em Potsdam
FRBR	Functional Requirements for Bibliographic Records
FRBRoo	Functional Requirements for Bibliographic Records Object Oriented
GFDL	GNU Free Documentation License
GIS	Geographic Information System
GLAM	Galleries, Libraries, Archives and Museums
HHU	Universidade Heinrich Heine de Düsseldorf
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
IA	Inteligência Artificial
ICA	Conselho Internacional de Arquivos
IDC	International Data Corporation
IFLA	Federação Internacional de Associações e Instituições de Bibliotecas
INA	Institut national de l'audiovisuel
Insee	Institut national de la statistique et des études économiques

IoT	Internet das Coisas
IPN	Instituto Politécnico Nacional Santo Tomás
ISAAR(CPF)	Norma Internacional de Registo de Autoridade Arquivística para Entidades Coletivas, Pessoas e Famílias
ISAD(G)	Norma Geral Internacional de Descrição Arquivística
ISDF	Norma Internacional para Descrição de Funções
ISDIAH	Norma Internacional para Descrição de Instituições com Acervos Arquivísticos
ISFD	Norma Internacional para Descrição de Funções (ISDF);
JSON	JavaScript Object Notation
LAM	Libraries, Archives and Museums
LOCAH	Archives Hub Linked Data
LOD	Linked Open Data
MER	Modelo Entidade-relacionamento
NIC.br	Núcleo de Informação e Coordenação do Ponto BR
NOBRADE	Norma Brasileira de Descrição Arquivística
ODC-By	Open Data Commons Attribution License
ODC-ODbL	Open Data Commons Open Database License
OO	Orientação a Objetos
PDDL	Public Domain Dedication and License
PREMIS	Preservation Metadata: Implementation Strategies
RAD	Rules for Archival Description
RC	Representação do Conhecimento
RDF	Resource Description Framework
RiC-CM	Records in Contexts
RiC-FAD	Records in Contexts: Foundations on Archival Description
RiC-O	Records in Contexts Ontology
SGML	Standard Generalized Markup Language
SKOS	Simple Knowledge Organization System
SPARQL	Protocol and RDF Query Language
SQL	Structured Query Language
SSI	Server Side Includes
TI	Tecnologia da Informação
TICs	Tecnologias de Informação e Comunicação
UCLAB	Laboratório de Complexidade Urbana

UDEC	Universidade de Celaya
UFF	Universidade Federal Fluminense
UFMG	Universidade Federal de Minas Gerais
UFRGS	Universidade Federal do Rio Grande do Sul
UFRJ	Universidade Federal do Estado do Rio de Janeiro
UFSB	Universidade Federal do Sul da Bahia
UFV	Universidade Federal de Viçosa
UNIRIO	Universidade Federal do Estado do Rio de Janeiro
UP	Universidade Panamericana
URI	Unified Resource Identifier
USCS	Universidade Municipal de São Caetano do Sul
W3C	World Wide Web Consortium
XML	eXtensible Markup Language

SUMÁRIO

1 INTRODUÇÃO	15
1.1 Problema	16
1.2 Objetivo geral	17
1.3 Objetivos específicos	17
1.4 Justificativa	17
1.5 Procedimentos metodológicos	19
2 REFERENCIAL TEÓRICO E REVISÃO DE LITERATURA	23
2.1 Inovação tecnológica, Web semântica e arquivos	23
2.1.1 O mundo centrado em dados	23
2.1.2 Da Web de documentos à Web de dados	32
2.1.3 <i>Linked Open Data (LOD)</i> ou Dados Abertos Conectados	38
2.1.4 Ontologias.....	48
2.2 Revisão teórica e prática de descrição arquivística	64
2.2.1 Revisão das normas internacionais de descrição arquivística	65
2.2.2 Princípio da proveniência ampliado e princípio do respeito à ordem original.....	72
2.2.3 Modelos conceituais de arquivos, bibliotecas e museus em descrição arquivística.	78
2.2.4 O modelo conceitual <i>Records in Contexts</i>	83
3 ANÁLISE QUALITATIVA	91
3.1 Análise de conteúdo	91
3.2 Análise e sistematização dos resultados	98
4 ANÁLISE EMPÍRICA	106
4.1 Aplicação dos questionários	106
4.1.1 Análise e sistematização dos resultados	108
5 CONSIDERAÇÕES FINAIS	124
REFERÊNCIAS	127
APÊNDICE A: MODELO DE QUESTIONÁRIO	133

1 INTRODUÇÃO

No mundo digital contemporâneo, marcado pela era da datificação da vida, conforme destacado pelo historiador, filósofo e professor israelense Yuval Noah Harari (2016), a *Web* emerge como epicentro de dados, consolidando-se como instrumento primordial para o uso e o compartilhamento de informações. Com mais de três décadas em funcionamento, a *Web* está prevista para acumular centenas de zettabytes de dados e uma vasta quantidade de páginas, uma profusão que afeta a qualidade das buscas e da recuperação de informações. Neste contexto, as instituições arquivísticas, que desde a década de 1990 adotam padrões internacionais de descrição, intensificam a disponibilização de dados sobre os seus acervos em *sites* institucionais. Contudo, existe o risco de que os dados arquivísticos sejam ofuscados no vasto oceano de informações da *Web*. Torna-se imperativo o desenvolvimento de aparatos que destaquem os dados arquivísticos dentre outros conteúdos, por vezes efêmeros, na Internet, e a instalação da automação computacional no processamento de dados e pesquisas, com o suporte de motores de busca capazes de realizar inferências e utilizar recursos avançados, como os proporcionados pela Web semântica.

Junto ao conceito de Dados Abertos Conectados (*Linked Open Data*), a Web semântica fundamenta o modelo conceitual *Records in Contexts (RiC)*, desenvolvido pelo Conselho Internacional de Arquivos (ICA). Inspirando-se nos modelos conceituais de bibliotecas, museus e arquivos da Espanha, Finlândia e Austrália, o *RiC* surge como uma inovação em descrição arquivística. Ele serve à transição da sintaxe para a interoperabilidade semântica nos arquivos – segundo os professores Dunia Llanes-Padrón, da área de Ciência da Informação da Universidade de Havana, Cuba, e Juan-Antonio Pastor-Sánchez, da área de Informação e Documentação da Universidade de Múrcia, Espanha (2017) – não mais pelo compartilhamento de instrumentos de pesquisa digitalizados ou descrições à disposição em páginas da *Web*, mas pelo intercâmbio de dados estruturados e semantizados empregando ontologias, formatos abertos e vocabulários compartilhados entre as comunidades arquivísticas e a Web semântica. Com isso, os dados dos arquivos adquirem maior expressividade na recuperação, por constituir um sistema aberto que oferece amplas possibilidades de uso e reutilização aos pesquisadores.

Esta pesquisa volta-se para a intersecção entre as práticas de descrição em arquivos e as tecnologias emergentes, enfatizando a Web Semântica e suas repercussões nesta função arquivística. O presente estudo explora, principalmente, as críticas do modelo *RiC* aos métodos atuais de descrição, como a descrição hierarquizada e a desarmonia das normas do ICA, para

propor novos padrões, tais como a descrição multidimensional, a expansão do tradicional princípio da proveniência e a integração de dados das galerias de arte, bibliotecas, arquivos e museus (LAM ou GLAM¹), trazendo novas perspectivas para o campo.

A relevância desta pesquisa reside na escassez de estudos que tratem as propostas do *RiC* de maneira abrangente. O objetivo de tal investigação é analisar as inovações na descrição arquivística e no ecossistema das tecnologias semânticas incorporadas ao *RiC*. Ao utilizar uma metodologia qualitativa e exploratória, o estudo emprega técnicas como análise de conteúdo e aplicação de questionários para aprofundar a compreensão dos impactos do *RiC* nas práticas arquivísticas.

Além disso, a pesquisa examina o estado da arte na aplicação do *RiC* em iniciativas internacionais. Profissionais da França, Suíça, Alemanha, Holanda, Liechtenstein, Noruega e Suécia, países pioneiros na adoção do modelo, compartilham as suas experiências e visões sobre as tecnologias da Web semântica em arquivos.

A compreensão dos benefícios tangíveis do modelo *RiC* para as práticas descritivas no âmbito arquivístico constitui a mais importante motivação para a realização desta investigação. O estudo em questão objetiva aliar-se às discussões de âmbito internacional, originando uma análise mais aprofundada e rigorosa acerca do referido modelo e de sua interação com a Web semântica. O que não apenas enriquece o conhecimento existente na comunidade acadêmica e entre profissionais da área, mas também contribui de maneira positiva para o campo da arquivologia.

1.1 Problema

É na avaliação de aspectos teóricos e práticos da descrição arquivística no contexto contemporâneo digital, e com base nas propostas do modelo *RiC*, que reside o problema central desta pesquisa de mestrado. Especificamente, questiona-se qual o impacto do modelo *RiC* nos princípios fundamentais da arquivística, sobretudo no princípio da proveniência e na preservação do contexto arquivístico. Um problema-chave é compreender se o modelo *RiC* consegue sanar lacunas anteriormente identificadas nas práticas de descrição arquivística, particularmente a falta de integração das normas internacionais de descrição, e se ele promove aprimoramentos na descrição em geral. Por fim, soma-se o problema de até que ponto a

¹Libraries, Archives and Museums (LAM) ou Galleries, Libraries, Archives and Museums (GLAM).

integração de dados de acervos de instituições patrimoniais, como bibliotecas e museus, proposta pelo modelo *RiC*, pode beneficiar a descrição arquivística e os usuários dos arquivos.

1.2 Objetivo geral

Analisar e sistematizar os problemas apontados pelo modelo *RiC* na descrição arquivística contemporânea e os benefícios decorrentes das soluções propostas por ele, bem como explorar as perspectivas de instituições que aplicam o modelo.

1.3 Objetivos específicos

- a) Investigar as principais influências teóricas que embasam o modelo *RiC*.
- b) Identificar e sistematizar pontos sensíveis, questionamentos e críticas à descrição arquivística feitas pelo modelo *RiC*.
- c) Identificar e sistematizar os benefícios constantes no modelo *RiC*, considerando seu potencial uso em instituições arquivísticas.
- d) Examinar projetos de aplicação do *RiC* em arquivos e explorar as perspectivas das instituições sobre o modelo.

1.4 Justificativa

A elaboração do modelo conceitual *Records in Contexts (RiC)* pelo Conselho Internacional de Arquivos (ICA) tem repercutido na comunidade arquivística desde 2012. Com a publicação da primeira versão em 2023, depois de mais de uma década de trabalho do ICA, diversas questões permanecem em aberto para arquivistas e instituições arquivísticas. Entre os pontos controversos, destacam-se a revisão e harmonização das normas internacionais de descrição do ICA, novas considerações sobre o princípio da proveniência e a tendência de integração dos arquivos com bibliotecas e museus, além do complexo aparato tecnológico do modelo, que se define como um dispositivo da Web semântica e do *Linked Open Data*.

Jair Martins Miranda, professor do Departamento de Arquivologia da Universidade Federal do Estado do Rio de Janeiro (Unirio), crê num "cenário de incerteza sobre o impacto futuro do novo padrão *RiC* nos princípios e práticas arquivísticas" (Miranda, 2021). Durante o período de elaboração do *RiC*, a comunidade arquivística internacional teve a oportunidade de

enviar comentários sobre o modelo, num total de 165 sugestões e críticas, tanto ao conteúdo quanto à predominância de especialistas europeus e à falta de contribuições da Ásia e África (InterPARES, 2016).

Em 2020, a revista *Archiviste*, da Associação dos Arquivistas Franceses, publicou um dossiê intitulado *Les archives et le Web sémantique*, provocando discussões sobre a Web semântica nos arquivos da França. O dossiê inclui relatos, depoimentos e entrevistas com diferentes perspectivas de profissionais experientes na aplicação de tecnologias da Web semântica em acervos arquivísticos.

Manonmani Restif, do Serviço Interministerial dos Arquivos da França, ressalta a importância da Web semântica nas buscas de documentos e a relação entre arquivistas e usuários. No projeto *FranceArchives*, é proposto como uma interface intuitiva e simples, que permite aos usuários manejarem os dados de modo diverso e independente, ao valorizar as relações entre os dados e a serendipidade², e ao apresentar-se como alternativa aos instrumentos de pesquisa tradicionais, considerados lineares e hierarquizados (Restif, 2020). Florence Clavaud, do Arquivo Nacional da França, compartilha a experiência da instituição com o *RiC-CM* e a ontologia *RiC-O*, destacando as vantagens do modelo em termos de precisão, uso e aprimoramento dos metadados nos sistemas de gestão (Clavaud, 2020).

No entanto, Gaël Chenard, arquivista paleógrafo, doutor em história medieval e diretor do Arquivo Departamental de Vienne, põe em questão a prematuridade de adotar o *RiC* em arquivos que ainda não descreveram seus acervos ou que não possuem metadados nos formatos *Extensible Markup Language (XML)* e *Encoded Archival Description (EAD)*. Ele argumenta que o custo-benefício da mudança não está clara (Chenard, 2020). Jean-Séverin Lair, do Instituto Nacional de Estatística e Estudos Econômicos da França (Insee), pondera que, apesar de vinte anos de existência, a tecnologia da Web semântica não se popularizou, razão pela qual enfrenta várias adversidades como a complexidade na criação de relacionamentos entre dados e o tratamento de grandes volumes de dados (Lair, 2020). Gautier Poupeau, do Instituto Nacional do Audiovisual (INA), analisa as contribuições e limitações da Web semântica, ao destacar a importância do vocabulário e dos relacionamentos entre silos de dados, bem como os problemas relacionados à escalabilidade, robustez e manutenção dessas tecnologias (Poupeau, 2020).

² “A serendipidade se refere a descobertas feitas ao acaso, capacidade que as tecnologias da Web Semântica e principalmente o *Linked Data*” (Santarém Segundo; Coneglian, 2016, p. 218).

As discussões mencionadas refletem o quadro de incertezas e divergências que permeia os arquivos diante do modelo *RiC*. Consta-se a ausência de publicações que sistematizem de maneira abrangente as propostas do modelo. Esta pesquisa justifica-se pela necessidade de um aprofundamento nos conhecimentos acerca dos benefícios do modelo conceitual *RiC* e do seu aparato tecnológico para as práticas de descrição em instituições arquivísticas. Por fim, além de se nivelar às discussões internacionais, considera-se esta pesquisa oportuna para trazer elementos à discussão do *RiC* e promover estudos mais robustos sobre o modelo e sobre a interação entre arquivos e Web semântica, fortalecendo o conhecimento tanto na comunidade acadêmica quanto na comunidade profissional.

1.5 Procedimentos metodológicos

Existem diversas possibilidades de métodos que fundamentam a lógica de qualquer investigação científica. Conforme Antônio Carlos Gil, professor de metodologia de pesquisa na Universidade Municipal de São Caetano do Sul (USCS), a escolha do método pode depender de variáveis como “a natureza do objeto, os recursos materiais disponíveis, o nível de abrangência do estudo e a inspiração do pesquisador” (Gil, 2008, p. 9). Ao considerar a exequibilidade do emprego de métodos combinados, a metodologia adotada nesta pesquisa é categorizada como mista, de natureza exploratória, qualitativa e descritiva. Segundo Gil (2008),

Pesquisas exploratórias são desenvolvidas com o intuito de fornecer uma visão geral, aproximativa, acerca de um determinado fato. Este tipo de pesquisa é particularmente útil quando o tema escolhido é pouco explorado e se mostra desafiador para a formulação de hipóteses precisas e operacionalizáveis (Gil, 2008, p. 27).

A pesquisa exploratória inclui: o levantamento bibliográfico; entrevistas com pessoas que têm experiências práticas relacionadas ao problema de pesquisa; e a análise de exemplos que estimulam a compreensão do tema estudado (Gil, 2008). Em geral, como demonstram os pesquisadores de psicologia John J. Shaughnessy, Eugene B. Zechmeister e Jeanne S. Zechmeister, as pesquisas exploratórias servem para abrir os caminhos e antecedem as pesquisas descritivas, correlacionais ou explicativas. Tanto que uma pesquisa pode ser exploratória no seu início, passar a descritiva e correlacional e terminar como explicativa (Shaughnessy; Zechmeister; Zechmeister, 2012). O professor Amado L. Cervo, da Universidade de Brasília, o filósofo e economista Pedro A. Bervian e o pesquisador de metodologia científica Roberto da Silva afirmam, ainda, que a pesquisa exploratória descreve

precisamente o tema ou a situação e pretende descobrir as relações entre os seus componentes, como no caso desta pesquisa, quando é explorada a relação entre o *RiC* e a Web semântica e entre o modelo e a atual descrição arquivística. Portanto, a pesquisa exploratória tem o objetivo de familiarizar-se com um fenômeno ou dele obter uma nova percepção (Cervo; Bervian; Silva, 2002).

Este tipo de investigação caracteriza-se ainda como empírica, com a finalidade de formular questões ou problemas. Para Eva Maria Lakatos e Marina de Andrade Marconi, a pesquisa exploratória tem o triplo intento de descrever hipóteses; de aumentar a familiaridade do pesquisador com o ambiente, fato ou fenômeno para a execução de uma pesquisa futura mais precisa; e de esclarecer conceitos (Lakatos; Marconi, 2007).

Segundo os pesquisadores mexicanos Roberto Hernandez Sampieri, da Universidade de Celaya (UDECE), Carlos Fernández Collado, do Instituto Politécnico Nacional Santo Tomás (IPN) e María del Pilar Baptista Lucio, da Universidade Panamericana (UP), as pesquisas exploratórias raramente constituem um objetivo final por si só. Geralmente, desempenham um papel alicerçador ao determinar tendências, identificar áreas, ambientes, contextos e situações relevantes para estudo, além de apontar potenciais relações entre variáveis. Também servem para estabelecer as bases de pesquisas subsequentes, mais elaboradas e rigorosas. Distintas na metodologia, as pesquisas exploratórias se caracterizam pela flexibilidade, em contraste com os métodos mais estruturados das pesquisas descritivas, correlacionais ou explicativas. Devido à sua natureza abrangente e diversificada, tais investigações envolvem um maior risco. Exigem, portanto, uma considerável dose de paciência, serenidade e receptividade da parte do pesquisador (Hernández Sampieri; Fernández Collado; Baptista Lucio, 2013).

Aqui, o tema explorado é o modelo *RiC*, recentemente introduzido pelo Conselho Internacional de Arquivos, por intermédio do Grupo de Especialistas em Descrição Arquivística (EGAD). Depois da publicação de versões preliminares em 2016, 2019 e 2021, a primeira com o estatuto de oficial do modelo foi lançada em 2023. O que explica a insipiência das investigações sobre o *RiC*, tanto na comunidade arquivística internacional quanto na brasileira. Esta investigação, de natureza exploratória, não almeja a obtenção de conclusões acerca dos seus objetivos específicos. Ao invés disso, propõe-se a realizar um levantamento abrangente das questões pertinentes à descrição arquivística sob a perspectiva do modelo *RiC*. Devido ao caráter inovador do tema, percebem-se certas limitações e áreas pouco compreendidas, que podem influenciar na generalização dos resultados.

Por ser a fase inicial de uma pesquisa mais ampla, coincidente com o momento em que a comunidade arquivística brasileira se familiariza com o modelo *RiC*, espera-se que as descobertas preliminares desta investigação sejam usadas em futuros estudos explicativos, ao aprofundar as críticas e os questionamentos à descrição arquivística tradicional à luz do *RiC*, além de avaliar a eficácia do modelo nas instituições de arquivo.

A presente investigação científica foi estruturada em diversas etapas, o que possibilita uma visão organizada e detalhada das fontes consultadas, garantindo a integridade analítica do método. O levantamento bibliográfico e documental, na **Seção 2**, consistiu na elaboração do referencial teórico do trabalho. A primeira parte dedicou-se a atingir o objetivo específico de **investigar as principais influências teóricas que embasam o modelo *RiC*** pelo levantamento e análise do referencial teórico sobre dados, Web semântica e ontologias. Na segunda parte, o referencial teórico sobre a descrição arquivística, incluindo a revisão das normas de arquivo, os princípios teóricos, o mapeamento de modelos conceituais e a análise do modelo *RiC*. Nesta fase, está a análise interpretativa e reflexiva das publicações técnicas e científicas, representativa de uma pesquisa qualitativa.

Na **Seção 3**, para alcançar os objetivos específicos de **identificar e sistematizar pontos sensíveis, questionamentos e críticas à descrição arquivística feitas pelo modelo *RiC*** e de **identificar e sistematizar os benefícios constantes no modelo *RiC*, considerando a seu potencial uso em instituições arquivísticas**, utilizou-se a técnica de análise de conteúdo, conforme a metodologia criada por Laurence Bardin, professora de Psicologia da antiga Universidade Paris V. Aplicada às publicações contendo questionamentos à descrição arquivística atual e às tecnologias consideradas obsoletas, a análise enquadra-se igualmente na pesquisa qualitativa, visando à compreensão de significados e padrões nestas fontes. Bardin (2021) define a análise de conteúdo como um conjunto de técnicas para analisar comunicações, pretendendo obter indicadores que permitam a inferência de conhecimentos sobre as condições de produção e recepção das mensagens. A análise de conteúdo desta seção teve, como referência, aquela feita por Mariana Baptista Brandt, bibliotecária e doutora em Ciência da Informação pela Universidade Estadual Paulista (UNESP) nas normas referentes à informação da Câmara dos Deputados, quando de sua pesquisa de doutorado intitulada “Modelagem da informação legislativa: arquitetura da informação para o processo legislativo brasileiro”.

Na proposta de Bardin (2021), a análise se divide em etapas. A fase de pré-análise inclui: a organização dos documentos a serem analisados; a formulação de hipóteses e objetivos; e a elaboração de indicadores para interpretação final. No presente estudo, a seleção de documentos

para análise de conteúdo ocorreu em várias fases, começando com uma análise geral do referencial teórico do modelo *RiC* e seguindo com a redução progressiva das fontes de referência. A metodologia resultou na seleção de um conjunto de textos que tratam da associação entre arquivos e a Web semântica, assegurando um *corpus* representativo para a análise subsequente.

A seleção de documentos incluiu várias etapas, começando com a análise do referencial teórico do *RiC* e prosseguindo com a exclusão de materiais que se concentravam apenas em áreas específicas da arquivologia, biblioteconomia, museologia ou Web semântica. O conjunto final de documentos, limitado aos textos do *RiC* e aos que discutiam a associação entre arquivos e Web semântica, foi estabelecido. A etapa final abrangeu a formulação de hipóteses e objetivos, seguindo as orientações de Bardin (2021), e a definição de critérios para a análise e interpretação dos dados coletados. A meta era identificar questionamentos, críticas e considerações sobre a descrição arquivística tradicional, assim como as inovações e benefícios oferecidos pelo modelo *RiC*.

Na Seção 4, para **examinar projetos de aplicação do *RiC* em arquivos e explorar as perspectivas das instituições sobre o modelo**, como método de coleta de dados aplicaram-se questionários em instituições que testam o *RiC*. Esta fase incorporou elementos quantitativos, principalmente nas questões fechadas, de múltipla escolha e na utilização da escala de Likert, facilitando a coleta de dados mensuráveis e quantificáveis, típicos da pesquisa quantitativa. As questões abertas, por sua vez, foram enquadradas na pesquisa qualitativa. De qualquer forma, por se tratar de pesquisa exploratória, não se pretendeu chegar a dados em que as amostras fossem capazes de gerar inferências, em especial, pelo número reduzido de projetos de aplicações e testes do *RiC* em instituições arquivísticas, bem como pelo número ainda mais abreviado de respondentes ao questionário – de 31 instituições, apenas 9 responderam. Portanto, seria necessário um maior número de respostas ao questionário para que a amostra dos dados alcançasse um padrão mínimo para representar, com precisão, as opiniões das instituições arquivísticas sobre o modelo *RiC*.

Neste estudo, assim sendo, a metodologia é mais exploratória do que descritiva. Exploratória, pois visa a compreender um fenômeno pouco investigado ou emergente – o uso do modelo *RiC* em instituições arquivísticas. Descritiva, ao detalhar os benefícios do modelo *RiC* e a percepção das instituições acerca de sua aplicação nos arquivos. A combinação de ambas as técnicas de pesquisa proporciona uma análise rica e multifacetada do objeto de estudo, levando a uma compreensão inicial do modelo *RiC* e de sua aplicação na descrição arquivística.

2 REFERENCIAL TEÓRICO E REVISÃO DE LITERATURA

Esta seção se dedica à apresentação do referencial teórico e da revisão de literatura que basearam este trabalho.

2.1 Inovação tecnológica, Web semântica e arquivos

Esta subseção trata de aspectos relativos à inovação tecnológica, Web semântica e arquivos.

2.1.1 O mundo centrado em dados

A humanidade tem constantemente buscado adaptar o ambiente para atender aos próprios anseios de bem-estar. Ao longo da história, com o desenvolvimento técnico, alcançamos extraordinários níveis de crescimento e conforto, apesar das críticas acerca da distribuição assimétrica de recursos materiais e serviços. As tecnologias, que também miram o bem-estar, são marcadas pela evolução, infinitude e mutabilidade, têm em vista minorar os esforços exigidos pela existência (Ortega y Gasset, 2015).

A revolução das Tecnologias de Informação e Comunicação (TICs) é dos fenômenos mais notáveis da sociedade contemporânea na busca pelo bem-estar. Segundo Castells (2000), esta revolução tecnológica é a maior transformação desde a Revolução Industrial, moldando o patrimônio material, a economia, a sociedade e a cultura.

A dimensão da influência das novas tecnologias – com novos produtos – é um índice incontestável das transformações nas atividades humanas. Aqui, inserem-se as tecnologias de processamento e comunicação de informações que se destacam na revolução contemporânea, incluindo progressos em microeletrônica, computação, telecomunicações/rádiodifusão, optoeletrônica e engenharia genética. A informação, como matéria-prima, a lógica das redes e os sistemas integrados constituem os pilares dessa revolução (Castells, 2000).

O avanço das Tecnologias de Informação e Comunicação tem promovido uma transformação substancial que vai do setor industrial e comercial até o domínio da investigação científica. Atualmente, a sociedade experimenta não apenas a centralidade das tecnologias, mas também a importância preeminente dos dados, produzidos e consumidos em volumes sem precedentes. Estudos recentes indicam que várias indústrias estão investindo na gestão e uso de

dados, atribuindo-lhes uma vital importância para a perpetuação das organizações contemporâneas, período chamado a Era dos Dados (*Data Age*).

O historiador, filósofo e professor Yuval Noah Harari (2016) afirma estarmos na era do dataísmo e investiga como a sociedade contemporânea está adaptando o entendimento de conhecimento e informação pelo uso intensivo de dados e algoritmos. O professor considera o dataísmo uma espécie de estrutura religiosa – a religião dos dados – que confere sentido ao mundo, de maneira análoga ao humanismo e outras ideologias presentes ao longo da história. Além disso, ele sublinha a maior inovação introduzida pelo dataísmo, desde 1789: a liberdade da informação. Para Harari, embora o movimento dataísta compartilhe aspirações humanistas, "no século XXI, o dataísmo pode deslocar o humano, substituindo uma visão antropocêntrica por uma datacêntrica" (Harari, 2016, p. 392).

A atual conjuntura comprova a intensa interconexão entre a sociedade e o volume crescente de dados. No ano de 2020, a corporação estadunidense Splunk Inc., em São Francisco, Califórnia, especializada na criação de *softwares* destinados à pesquisa, monitoramento e análise de dados produzidos por máquinas, conduziu um estudo em colaboração com a True Global Intelligence e divulgou o relatório intitulado “A Era dos dados chegou. Você está preparado?” (Splunk Inc, 2020b). A pesquisa, com a participação de 2.259 gestores de negócios de TI – originários dos Estados Unidos, França, China, Austrália, Reino Unido, Alemanha, Japão e Holanda – evidenciou a importância crítica dos dados. De acordo com o estudo, 67% dos entrevistados estimam um aumento do volume de dados em quase cinco vezes até 2025 e ressaltam que eles “são extremamente ou muito valiosos para a sua empresa, no que tange ao sucesso geral (81%), inovação (75%) e cibersegurança (78%)” (Splunk Inc, 2020b, p. 5). Ademais, a maioria dos inquiridos (57%) manifesta preocupação com a expansão do volume de dados, pois ela tem sido mais veloz do que a capacidade das organizações de geri-la. E 66% relatam que ao menos metade dos dados das próprias empresas permanece subutilizada, desconhecida ou inexplorada (*Dark Data*).

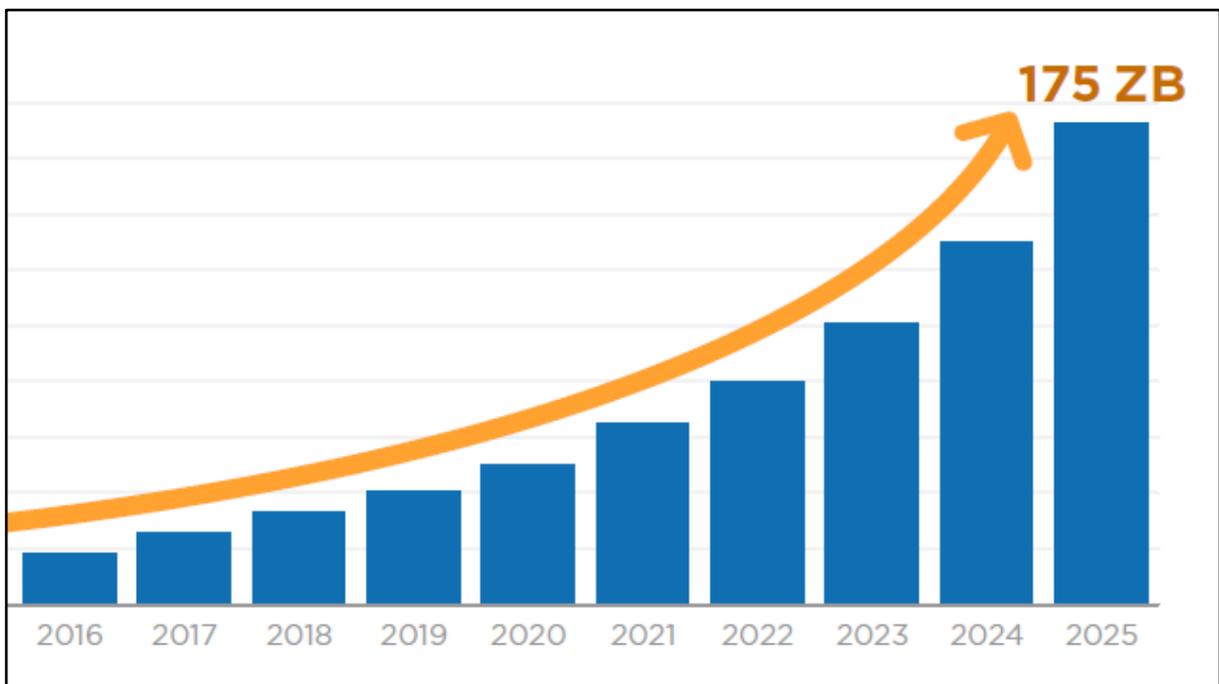
No documento elaborado pela Splunk (2020b), identificam-se as tecnologias de vanguarda que impulsionam o aumento exponencial do volume de dados: *Edge Computing*, 5G, Internet das Coisas (IoT), Inteligência Artificial e Aprendizado de Máquina (AI/ML), Realidade Aumentada e Virtual (AR/VR) e *Blockchain*. Segundo a pesquisa, quase metade dos participantes (49%) espera utilizar tais tecnologias emergentes para amplificar o potencial dos dados. No entanto, apenas 42% afirmam possuir um conhecimento avançado sobre as referidas

tecnologias, sendo que a mais avançada delas – a Internet das Coisas (IoT) – é usada em apenas 28% das organizações. Nas palavras de Doug Merritt, presidente e CEO da Splunk,

A Era dos Dados está aqui. Agora, podemos quantificar como os dados estão no centro do palco das indústrias ao redor do mundo. Como esta nova pesquisa demonstra, as organizações entendem o valor dos dados, mas estão sobrecarregadas com a tarefa de se ajustar às muitas oportunidades e ameaças que a nova realidade apresenta (Splunk Inc., 2020a, tradução nossa).

Conforme o autor e editor Alex Woodie (2022), especialista em Tecnologia da Informação (TI), a International Data Corporation (IDC), em seus relatórios *DataSphere* e *StorageSphere*, calcula que a geração de novos dados aumentará a uma taxa composta anual de 23%, culminando na criação de aproximadamente 175 Zettabytes³ (ZB) até 2025, como demonstra a Figura 1.

Figura 1: Estimativa de crescimento de dados de 2020 a 2025 em *Zettabytes*, segundo a IDC.



Fonte: Woodie (2022).

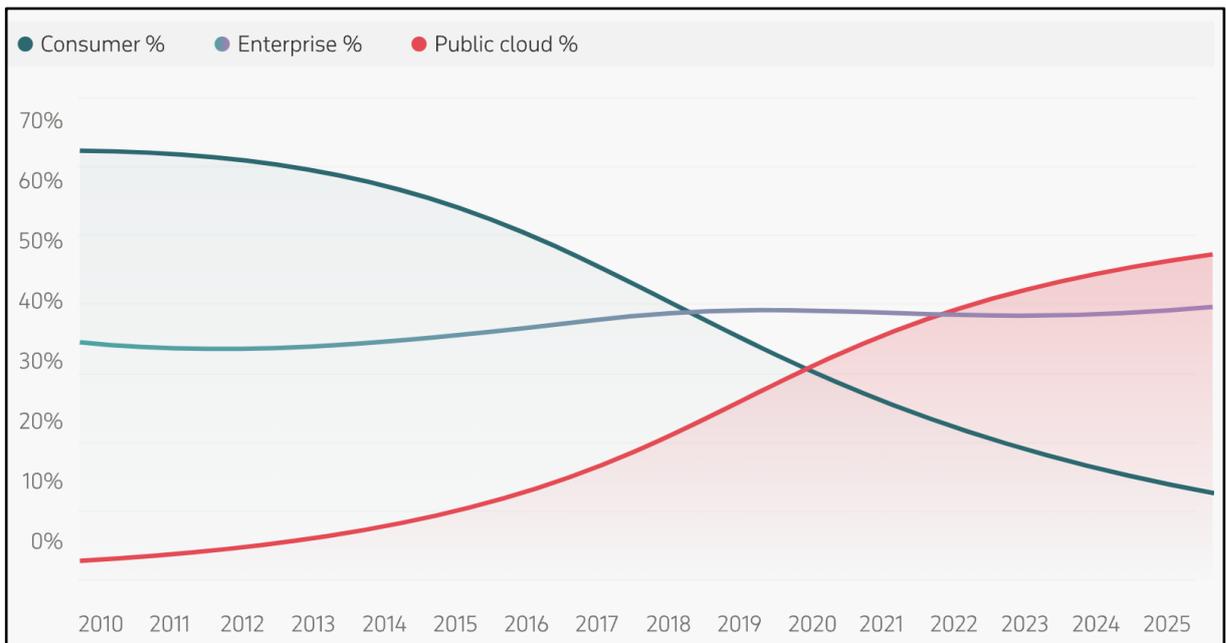
No entanto, destaca-se que, dos 64,1 ZB gerados em 2020, somente 2% foram preservados ou mantidos para 2021. Dave Reinsel, vice-presidente sênior da IDC, afirma que “o volume de dados digitais gerados nos próximos cinco anos superará mais que o dobro da quantidade de dados criados desde o início do armazenamento digital” (Woodie, 2022, tradução

³ Um único *zettabyte* equivale a 1.000.000.000.000.000.000 (1 sextilhão) de *bytes*, que representaria cerca de 250 bilhões de DVDs (Hack, 2021).

nossa). A mesma investigação aponta ainda que o mercado global de *Big Data* e análise de negócios poderá alcançar um valor de 448 bilhões de dólares até 2027, influenciando diretamente a expansão das equipes de análise, engenharia e ciência de dados.

No que diz respeito ao armazenamento, a maior parte dos dados está atualmente armazenada nos aparelhos dos consumidores. Contudo, prevê-se que esta realidade será suplantada pelo armazenamento em nuvem até 2025, conforme consta na Figura 2.

Figura 2: Previsão do armazenamento de dados até 2025, segundo o relatório *Data Age 2025* da IDP.

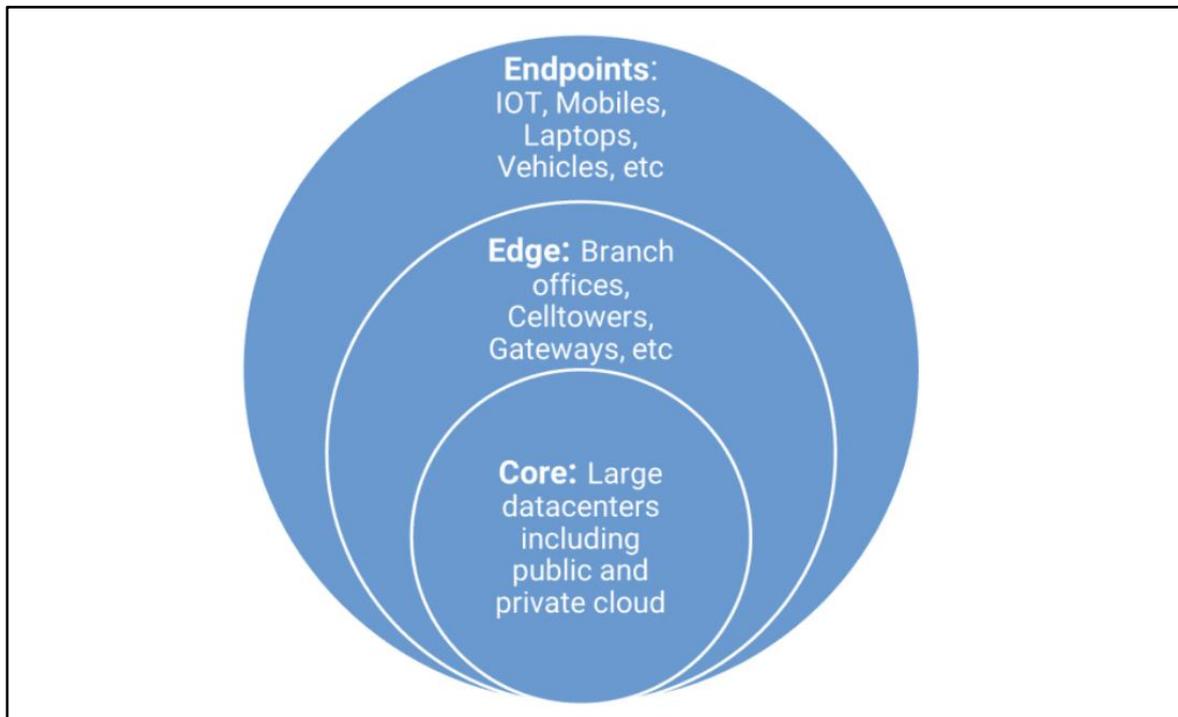


Fonte: Khvoynitskaya (2020).

Ulrike Hack (2021), gerente de marketing de produto do grupo Redgate, observa que, mesmo diante de um crescimento acelerado dos dados, apenas uma fração mínima, de aproximadamente menos de 2%, dos dados gerados, é efetivamente armazenada. Assim, cerca de 98% é destinada ao uso imediato e não para uso futuro, sendo temporariamente preservada “em cache” ou sobreposta por novos dados. Como visto na Figura 3, a maioria dos dados localiza-se fora das redes, em dispositivos como celulares, sensores e aparelhos conectados à IoT – como *smartwatches*, casas inteligentes, carros conectados e *drones*, conhecidos como *Endpoints*. Uma parcela menos expressiva de dados está em servidores das organizações e *datacenters* menores, o chamado *Edge*. Já a menor porção encontra-se em *datacenters* de empresas e provedores de nuvem, onde ocorrem a análise e o processamento. Hack (2021) ainda prognostica que, até 2025, haverá uma centralização dos dados no *Edge* e no *Core*, com as empresas assumindo a responsabilidade por sua geração e armazenamento:

As empresas que utilizam o *data warehouses* para analisar dados estruturados estão migrando para o *data lakes* para analisar dados não estruturados e semiestruturados. Enquanto os bancos de dados relacionais continuam como categoria mais popular, com 72,4% no ranking de setembro de 2021 no site *DB-Engines*, outros estão sendo introduzidos para lidar com o *Big data* (Hack, 2021, tradução nossa).

Figura 3: Representação de estratos de localização dos dados.



Fonte: Hack (2021).

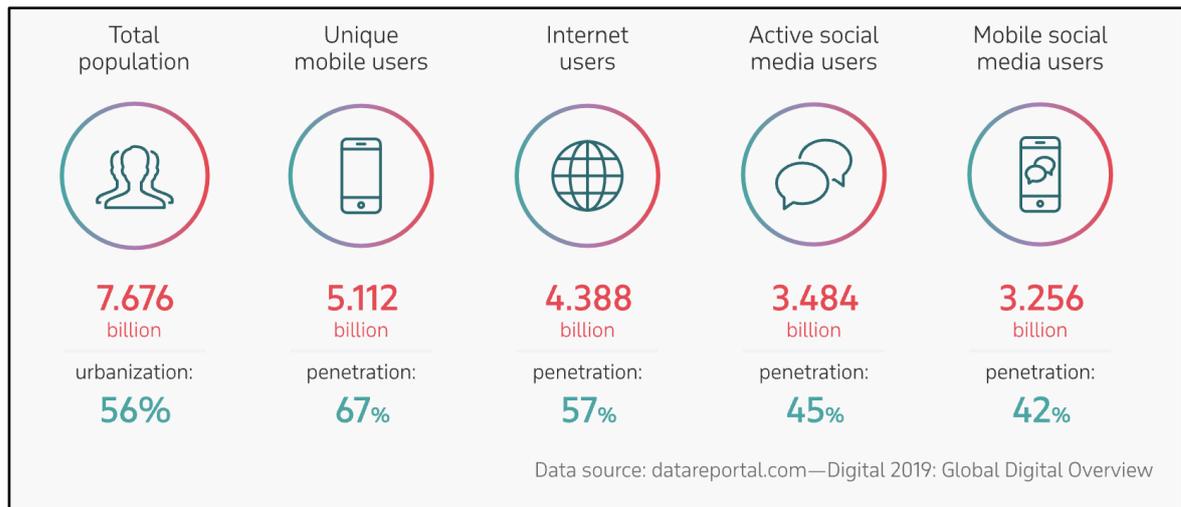
Portanto, independente de os dados serem classificados como *Small Data*⁴, *Wide Data*⁵ ou *Big Data*⁶, é incontestável a sua crescente importância na sociedade contemporânea. Como apontou Khvoynitskaya (2020), a IDC calcula que 75% da população mundial vai interagir com dados on-line até 2025. Na Figura 4, observa-se a profunda penetração do universo digital na sociedade, já no ano de 2019.

⁴ *Small Data* é um procedimento “que requer menos dados, mas ainda oferece insights úteis, incluindo técnicas de análise de séries temporais, aprendizado com poucos exemplos, dados sintéticos ou aprendizado autossupervisionado” (Gartner Inc., 2021, tradução nossa).

⁵ O *Wide Data* “permite a análise e sinergia de uma variedade de fontes de Big Data e Small Data, estruturados e não estruturados” (Gartner Inc., 2021, tradução nossa).

⁶ Para Kampakis (2020 *apud* Freitas, 2020), *Big Data* associa-se a bancos de dados NoSQL e à computação em nuvem, além de considerado termo genérico para conjuntos de dados, que, de tão vastos e complexos, os softwares tradicionais não conseguem processá-los.

Figura 4: Mundo digital em 2019, segundo a Global Digital Overview.



Fonte: Khvoynitskaya (2020).

A multiplicidade de categorias e a ilimitada escalabilidade dos dados conferem a eles o status de objeto de estudo em diversas áreas técnico-científicas, em que se destacam a Ciência de Dados, Computação, Ciência da Informação, Estatística, Economia, Inteligência Artificial, Bioinformática, Ciências Sociais e Administração – que dão suporte à gestão e à aplicação de dados com vistas à sua estruturação, armazenamento, processamento, análise e interpretação. Não de outro modo, surge o *The Data-centric Manifesto*⁷, uma iniciativa que reúne milhares de profissionais dessas áreas e que faz a crítica à cultura institucional tradicionalmente centrada em aplicações e *softwares*, em detrimento de uma cultura orientada para os dados, sugerindo que “[eles] são o centro do universo; as aplicações e *softwares* são efêmeros” (*The Data-Centric Manifesto*, [2023], tradução nossa).

Por trás do Manifesto, os profissionais denunciam que os dados se encontram aprisionados nas organizações, dispersos em silos de dados nos sistemas em geral excessivamente controlados por *softwares* e aplicações (*The Data-Centric Manifesto*, [2023]). Tal observação desafia a validade das boas práticas vigentes no desenvolvimento de sistemas e destaca o contraste com o êxito de empresas inovadoras, como Pinterest, Instagram, Facebook e Google, que operam sem fragmentar as próprias informações e sem investir em projetos dispendiosos de integração de dados.

⁷ Disponível em: <http://datacentricmanifesto.org/>.

O *Data-Centric Manifesto* estabelece sete princípios para uma posição centrada em dados: (1) os dados como recurso essencial de qualquer entidade organizacional; (2) a prevalência do atual padrão tecnológico nas organizações, centrado em aplicações; (3) o equívoco de confinar dados em sistemas proprietários complexos; (4) o custo e a complexidade desmesurados consequentes da relação das aplicações com os dados; (5) o compromisso dos signatários do Manifesto em alterar essa realidade; (6) a existência de lucros no modelo centrado em aplicações; e (7) o potencial de economia de recursos no modelo centrado em dados. Em suma, o Manifesto compara as duas perspectivas e incentiva uma transformação relevante na gestão dos dados organizacionais, como mostra o Quadro 1:

Quadro 1: Comparativo entre o modelo atual centrado em aplicações e o modelo futuro centrado em dados.

	Atualmente: perspectiva centrada em aplicações e softwares	Futuro: perspectiva centrada em dados
Custos Operacionais	Custos proibitivos de mudança.	Custos razoáveis de mudança.
Propriedade dos Dados	Dados vinculados a aplicações específicas.	Dados como recurso aberto e persistente.
Gestão de Projetos	Projetos novos requerem grandes conversões de dados.	Projetos novos acessam dados existentes.
Padronização de Dados	Dados em formatos e estruturas heterogêneas.	Dados integrados com significado comum.
Orçamento de TI	Integração de dados consome grande parte do orçamento.	Integração de dados com custo quase nulo.
Integração de Dados	Integração de dados internos e externos desafiadora.	Integração facilitada entre dados internos e externos.

Fonte: Adaptado de *The Data-Centric Manifesto*, ([2023]).

Apesar das diversas vertentes no campo dos dados, neste estudo, o ponto principal recai sobre uma área específica da Web semântica: o *Linked Open Data (LOD)* – ou seja, Dados

Abertos Conectados, cujo conceito provém do termo inglês *Linked Data*, originalmente proposto pelo cientista da computação Tim Berners-Lee, amplamente difundido pelo *World Wide Web Consortium (W3C)*, organização encarregada da padronização da *Web*.

Caroline Guimarães e Vagner Diniz, respectivamente responsável pelas relações institucionais da W3C Brasil e integrante do Núcleo de Informação e Coordenação do Ponto BR (NIC.br), defendem que os

Dados abertos de alta qualidade são dados publicados e distribuídos na Internet, compartilhados em formato aberto para que possam ser lidos por qualquer pessoa e por máquinas, permitindo o cruzamento com outros dados de diferentes fontes, para serem livremente reutilizados pela sociedade. Dados abertos governamentais são dados produzidos pelos governos, que devem ser colocados à disposição de qualquer cidadão e para qualquer fim (Guimarães; Diniz, 2011 *apud* Isotani; Bittencourt, 2015, p. 13).

Os Dados Abertos Conectados são princípios e práticas ligados à noção teórica mais ampla de Web semântica, que emprega uma variedade de técnicas e instrumentos oriundos de campos como Inteligência Artificial (incluindo agentes inteligentes e representação de conhecimento), Engenharia de *Software* (abrangendo *frameworks* e plataformas) e Computação Distribuída (envolvendo *web services*), sobretudo. O objetivo é executar atividades na *Web* que, anteriormente, apenas seres humanos eram capazes (Isotani; Bittencourt, 2015, p. 27).

Os Dados Abertos Conectados têm conexão com teorias precedentes, como as do sociólogo estadunidense Robert King Merton, acerca das vantagens dos dados científicos abertos, que seguem três preceitos elementares para a cooperação e colaboração institucional, conforme sugestão da *Open Knowledge Foundation* (2010 *apud* Isotani; Bittencourt, 2015): (1) “a disponibilidade e o acesso irrestrito”, a baixo custo, permitindo o *download* e a manipulação conveniente; (2) “a reutilização e a redistribuição”, sob condições que favoreçam tais práticas; e (3) “a participação universal,” assegurando a todos a capacidade de empregar, reempregar e disseminar tais dados.

Isotani e Bittencourt (2015) relatam a existência de diversos movimentos globais voltados para a abertura de dados. Entre eles, sobressai a iniciativa dos Dados Abertos Governamentais, impulsionada pela administração de Barack Obama nos Estados Unidos, e pelo lançamento do Data.gov., que torna acessíveis os dados governamentais para consulta pública. Outro item de referência é a parceria *Open Government Partnership*, de 2011, com a presença de 75 nações, incluindo o Brasil como membro fundador. Tal adesão resultou na criação do portal dados.gov.br, que põe à disposição dos cidadãos informações governamentais, seguindo os princípios dos dados abertos, alinhados à política de acesso à informação do

governo federal brasileiro e à Lei de Acesso à Informação⁸. Existem diversos exemplos práticos da aplicação de Dados Abertos Conectados em setores industriais, conforme ilustrado por Isotani e Bittencourt (2015). Os autores destacam, por exemplo, que a *Best Buy*, no contexto do comércio eletrônico, observou um aumento de 15% a 30% nos cliques originados pelo buscador Google em direção ao seu site, depois de implantar o formato de serialização de Dados Conectados, conhecido como *RDF*. Além disso, a adoção de padrões relacionados à Web semântica e ao *Linked Open Data* foi obra de grandes corporações e entidades, incluindo a Google, a IBM, o Facebook, a BBC e o Governo Britânico.

Ao examinar a conexão e a abertura de dados, destaca-se o seu uso e reuso. No vasto espectro dedicado à análise de dados, as Humanidades Digitais sobressaem-se. Burdick *et al.* (2020) definem as Humanidades Digitais como um conjunto de práticas acadêmicas abrangentes, incluindo ensino, pesquisa e publicação. Este campo é notável por sua natureza colaborativa e transdisciplinar, incorporando tecnologias computacionais.

Burdick e os demais coautores ressaltam habilidades fundamentais nas Humanidades Digitais relacionadas a dados, como a familiaridade com tipos de dados e formatos de arquivos, conhecimento de bases de dados, compreensão de dados estruturados em *eXtensible Markup Language (XML)*, padrões de metadados, plataformas *Geographic Information System (GIS)* e dados espaciais, principalmente. Assim, eles enfatizam que os resultados do aprendizado em Humanidades Digitais devem dar prioridade a habilidades como “a capacidade de compreender, analisar e utilizar dados [e] desenvolver uma avaliação crítica para fontes e dados” (Burdick *et al.*, 2020, p. 95).

Os pesquisadores sustentam que, embora o conteúdo seja o aspecto mais evidente no plano intelectual, nas Humanidades Digitais existem diversas maneiras de apresentar argumentos e experimentos. Neste campo, afluem novos conhecimentos a partir de “complexas interações, visualizações, dados e estrutura de dados, e até mesmo códigos” (Burdick *et al.*, 2020, p. 93). Os autores acrescentam que:

Projetos em Humanidades Digitais não são somente sobre conteúdo (embora isso é [sic] geralmente primordial), mas também sobre o design de múltiplos níveis de conhecimento e argumento que vão das operações na base de dados interna (*back-end*) aos pontos de acesso externo (*front-end*) de uma interface de usuário (Burdick *et al.*, 2020, p. 93).

⁸ Lei nº 12.527, de 18 de novembro de 2011.

No mundo orientado para dados, os arquivos, as bibliotecas e os museus integram-se a tal modelo. Eles compartilham a informação como núcleo central de suas atividades e geram, há séculos, dados sobre os acervos que custodiam. A conjuntura presente, marcada pela transformação digital e pelos avanços tecnológicos, tem impulsionado revisões nas práticas e teorias da gestão do patrimônio documental e cultural da humanidade.

No que concerne aos arquivos, as funções fundamentais, como a descrição dos acervos, adaptam-se ao mundo digital e possibilitam a elaboração de formas inovadoras de acesso aos dados dos arquivos e dos documentos. Tal circunstância requer um reexame dos conceitos e práticas arquivísticas tradicionais. Com a expansão da Internet e da Web semântica, que possibilita a integração de dados e sua interpretação por máquinas, as instituições responsáveis pelo patrimônio documental têm procurado modernizar os seus sistemas e colocar à disposição os acervos on-line. O professor de Gestão e Organização do Conhecimento da Universidade Federal de Minas Gerais (UFMG), Carlos Henrique Marcondes (2016, p. 78) constata a potencialidade das tecnologias semânticas e dados abertos na publicação de acervos na *Web*.

Em escala global, museus e bibliotecas se adaptaram à Web semântica antes dos arquivos, com destaque para os modelos *Conceptual Reference Model (CRM)*, do Comitê Internacional para a Documentação do Conselho Internacional de Museus (CIDOC), e *Functional Requirements for Bibliographic Records (FRBR)*, da Federação Internacional de Associações e Instituições de Bibliotecas (IFLA). Entretanto, os arquivos estão se expandindo em plataformas on-line, melhorando os serviços, especialmente na busca e na recuperação de documentos.

O Conselho Internacional de Arquivos (ICA), por intermédio do Grupo de Especialistas em Descrição Arquivística (EGAD), tem se dedicado à investigação da relação entre os arquivos e a Web semântica, o que veio a culminar no desenvolvimento do modelo conceitual de dados destinado à descrição arquivística *Records in Contexts (RiC-CM)*. O modelo almeja a integração das normas internacionais vigentes, bem como a consideração das práticas vistas em áreas afins do patrimônio, tais como bibliotecas, museus e centros de documentação.

2.1.2 Da Web de documentos à Web de dados

O modelo conceitual para descrição de arquivos *Records in Contexts* apresenta inovações possibilitadas pelas novas tecnologias de informação e comunicação, mais precisamente as tecnologias semânticas. Ao analisar o modelo *RiC* é necessário entender que

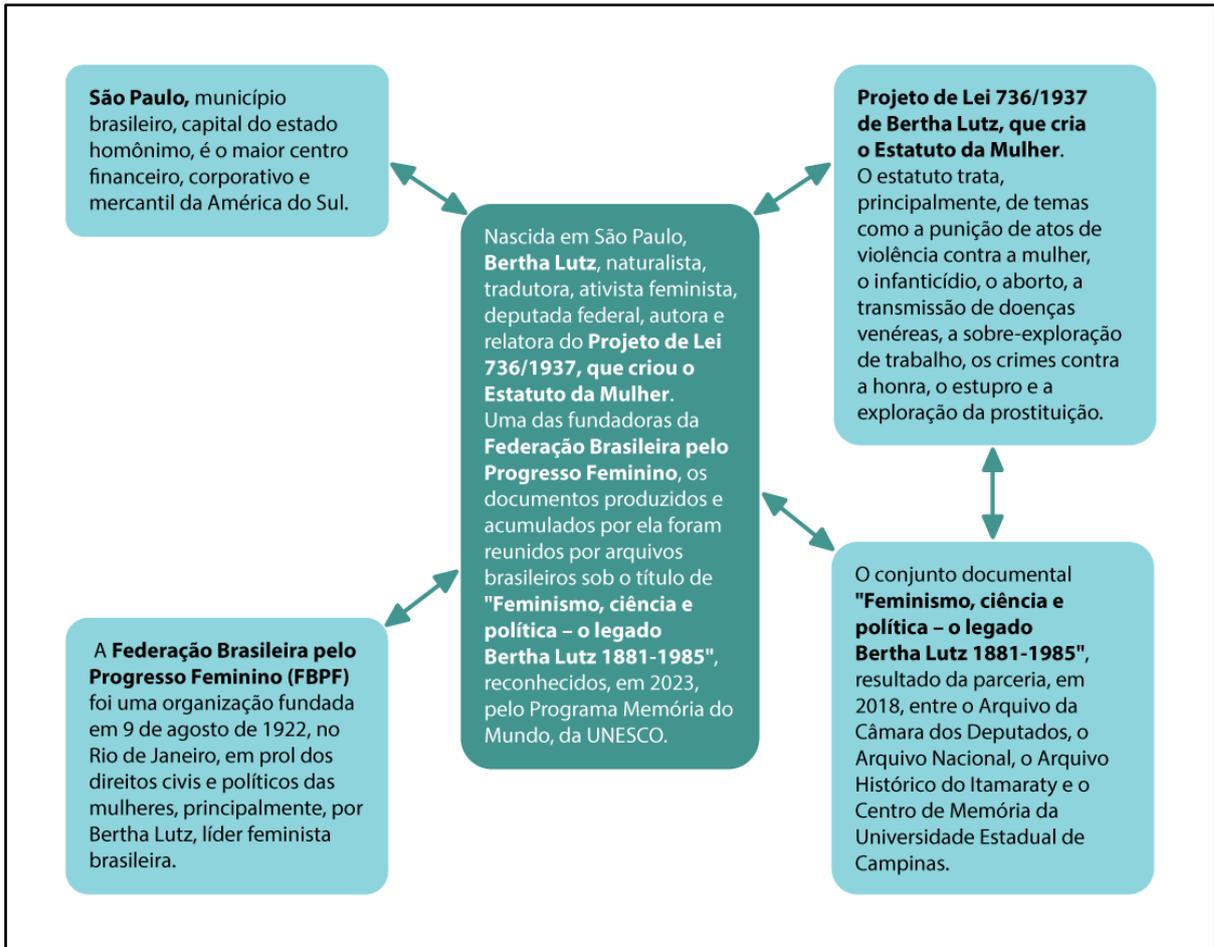
há a adesão à tecnologia específica da Web semântica por parte do Conselho Internacional de Arquivos. O conceito foi introduzido, em 2001, por Tim Berners-Lee, pelo pesquisador em inteligência artificial James Hendler e pelo cientista da computação Ora Lassila: “a Web semântica não é uma *Web* separada, mas uma extensão da atual, na qual a informação possui um significado bem definido, permitindo que computadores e pessoas trabalhem de maneira melhor e em cooperação” (Berners-Lee *et al.*, 2001, p. 29). Assim, a Web semântica visa a tornar, na *Web*, os dados mais compreensíveis e utilizáveis por humanos e máquinas.

Web semântica (Decker *et al.*, 2000 & Berners-Lee *et al.*, 1999) é o nome genérico deste projeto, capitaneado pelo W3C, que pretende embutir inteligência e contexto nos códigos XML utilizados para confecção de páginas Web, de modo a melhorar a forma com que programas podem interagir com estas páginas e também possibilitar um uso mais intuitivo por parte dos usuários (Souza *et al.*, 2004, p. 133).

Para maior abrangência do entendimento da Web semântica é essencial assimilar as transformações da Internet. A *World Wide Web*, ou *Web*, foi fundada entre 1989 e 1991, quando Tim Berners-Lee aprimorou conceitos dos cientistas Bush e Ted Nelson relacionados à ideia de hipertexto (Pickler, 2007).

O web designer MadhurKant Sharma ([2022]) descreve a *Web*, em sua etapa inicial, como *Web 1.0* ou *Web* de documentos, pela natureza estática de seu conteúdo, veiculado em páginas pessoais e atualizadas somente por seus proprietários. A arquitetura desta versão assentava-se em um servidor central de arquivos, com uso de linguagens como *Server Side Includes (SSI)* e *Common Gateway Interface (CGI)*. Segundo os professores Seiji Isotani, da Universidade de São Paulo, e Ig Ibert Bittencourt, da Universidade Federal de Alagoas, as bases da *Web* foram estabelecidas por três tecnologias: *Hypertext Markup Language (HTML)*, o protocolo *Hypertext Transfer Protocol (HTTP)* e o *Unified Resource Identifier (URI)*, para facilitar o intercâmbio documental no Centro Europeu de Pesquisas Nucleares (CERN) (Isotani; Bittencourt, 2015, p. 25).

Figura 5: Conjunto de páginas conectadas por *links* na Web de documentos (1.0).



Fonte: Adaptada de Isotani e Bittencourt (2015).

Em 1994, houve a primeira conferência internacional sobre a *Web*, marcada pelo anúncio do estabelecimento do W3C, entidade encarregada da normatização das tecnologias da *Web*. Entende-se que, na *Web* de documentos, que:

Os recursos como páginas web, figuras e vídeos estão conectados por meio de URIs. Essa conexão é fundamental para fazer associações entre diferentes recursos disponíveis na *Web* e para consumo massivo da informação destes recursos por pessoas [...]. Assim, podemos dizer que na *Web* de Documentos são os recursos, e não as informações contidas nestes recursos, que estão conectados (Isotani; Bittencourt, 2015, p. 43).

A *Web* 2.0, ou *Web* social, simboliza a transformação da Internet em plataforma participativa, destacando-se pela criação de conteúdo pelos usuários e pela ênfase na interação e na colaboração. Na *Web* 2.0, houve aperfeiçoamento na forma de como a *Web* era percebida

e utilizada, com destaque para a personalização avançada pelo *Application for Programming Interface (API)* e a multiplicidade de ferramentas on-line, como *podcasts* e redes sociais.

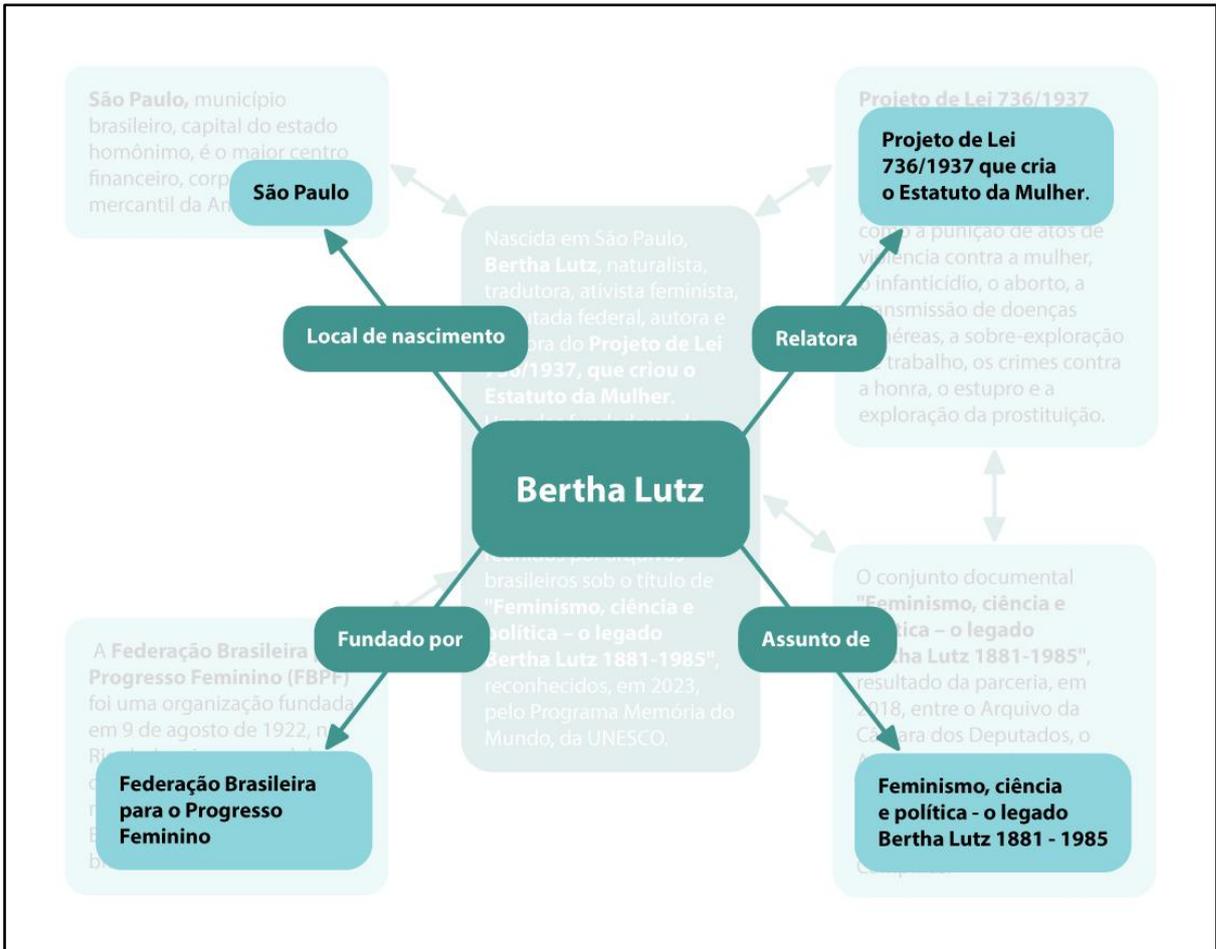
A *Web 3.0*, ou *Web* de dados, é a terceira geração da *Web*. Segundo Sharma ([2022]), ela se distingue pela aplicação da inteligência artificial para organizar e recuperar dados, transcendendo a simples busca por páginas. Abrange a extração de dados da *Web* para formar bancos de dados, favorecendo o avanço das tecnologias de *back-end* (infraestrutura), que, durante anos, deram prioridade a inovações do *front-end* (interface). Uma notável característica da *Web 3.0* é a natureza aberta e compartilhada dos dados, permitindo que um único *site* ou conjunto de dados seja interpretado de diversas maneiras. Os pesquisadores Tom Heath e Christian Bizer (2011) afirmam que a *Web* de Dados se constrói pela publicação dos dados conectados, possibilitando a transformação da *Web* de Documentos para a *Web* de Dados.

Assinalada pelo advento da *Web* semântica, a *Web 3.0* é o campo que almeja estruturar as informações do mundo para serem processadas não apenas por seres humanos, mas também por máquinas. Tim Berners-Lee, conforme, citado por Isotani e Bittencourt (2015), lembra que a *Web* de documentos emprega a linguagem *HTML* estruturada em hipertextos com *links* e nós que interligam documentos (páginas), permitindo a navegação entre eles. Há, porém, uma lacuna: a ausência de distinção de significados entre esses links impede que as máquinas compreendam as nuances de diferentes tipos de relação entre as páginas.

Isotani e Bittencourt (2015) salientam que, diante da massiva produção de dados e do crescimento exponencial de páginas na *Web*, torna-se impraticável aos humanos o acesso a informações de modo eficaz sem a assistência de máquinas ou de entidades de *software*. Neste panorama, a proposta de Berners-Lee é conectar os documentos na *Web* a recursos⁹ com o auxílio de associações semânticas, ou, em outras palavras, “a semântica acessível por máquina é potencializada por meio da especificação de documentos *Web* em uma linguagem que permita que os *links* sejam criados com valor em seu relacionamento” (Isotani; Bittencourt, 2015, p. 26).

⁹ “Qualquer coisa que possa ser identificada por um *URI*, como um vídeo, uma imagem, um serviço, um documento, entre outras coisas” (Isotani; Bittencourt, 2015, p. 57).

Figura 6: Conjunto de dados conectados, com semântica, na Web de dados (3.0).



Fonte: Adaptada de Isotani e Bittencourt (2015).

Na visão de Santarém Segundo e de Coneglian, a Web semântica tem como conceito central “construir uma rede de informações onde os *nós* estejam semanticamente ligados, formando um grande grafo global, com informações advindas de [...] fontes diferentes ao redor do planeta” (Santarém Segundo; Coneglian, 2016, p. 220). Assim, a Web semântica é um conjunto de conceitos, modelos, metodologias e tecnologias em transformação, enriquecido por estudos de caso aplicados ao longo dos anos. Ela é particularizada pela existência do nível semântico, pela automação por máquinas e por mecanismos como serendipidade e motores de inferência¹⁰.

¹⁰ “As inferências, por sua vez, são deduções de conhecimento novo a partir de conhecimento existente” (Noronha, 2021, p. 23).

A Web semântica integra-se totalmente ao contexto da *Web 3.0*. Por tal razão, é imprescindível que arquivistas e profissionais dos arquivos adquiram entendimento essencial das diversas fases da *Web*, elucidadas no Quadro 2:

Quadro 2: Principais características de cada fase da Web.

	Web 1.0	Web 2.0	Web 3.0
Tipo de interação	Somente leitura	Leitura e escrita	Portabilidade e uso pessoal
Alvo principal	Instituições	Comunidades	Indivíduos
Formas de conteúdo	Páginas	Blogs e Wikis	Live streams e Waves tokens
Natureza do conteúdo	Conteúdo proprietário	Conteúdo compartilhado	Conteúdo consolidado
Ferramentas de interface	Web forms	Aplicação Web	Aplicações inteligentes
Métodos de organização	Diretórios	Tagging	Análise de comportamento dos usuários
Métricas de sucesso	Número de acessos	Custo por clique	Engajamento do usuário
Estratégias de publicidade	Publicidade por banner	Publicidade interativa	Publicidade por comportamento
Ícones representativos da fase	Enciclopédia Britannica on-line	Wikipedia	Web semântica
Tecnologias chave de acesso aos dados	HTML / Portais	XML / RSS	RDF / RDFS / OWL

Fonte: Adaptado de Sharma (2022).

Na *Web 3.0* se dá o movimento dos Dados Abertos Conectados, importante componente das propostas da Web semântica para a organização e publicação de dados na rede mundial de computadores.

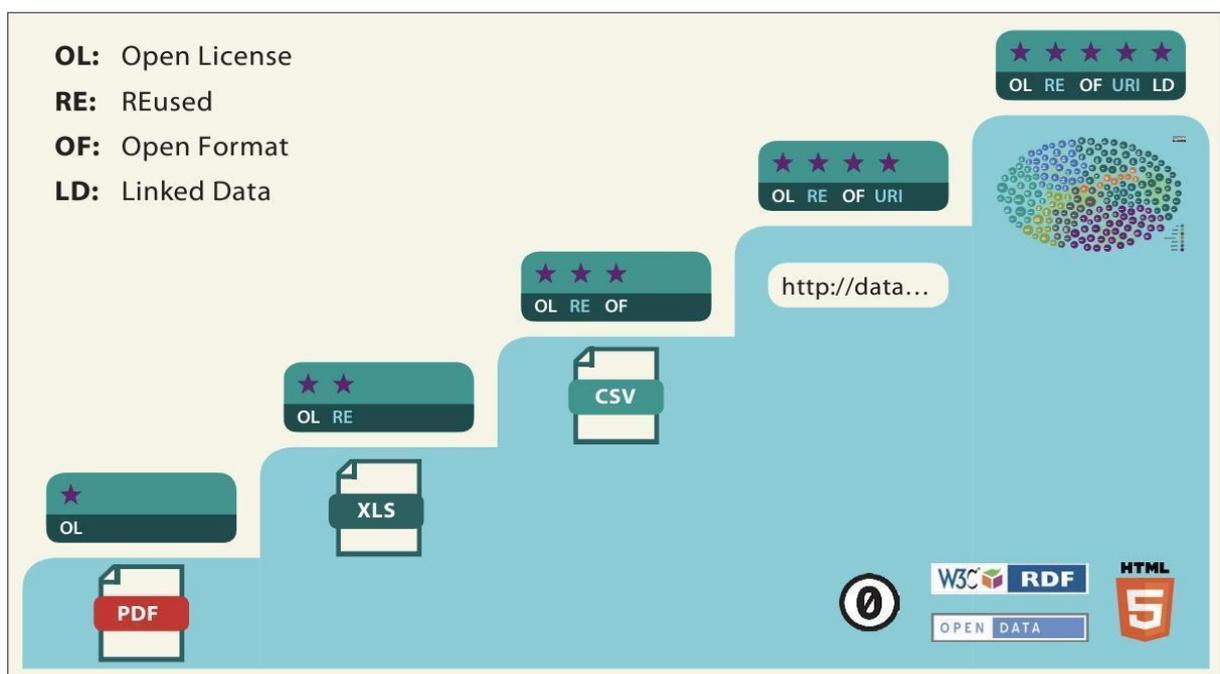
2.1.3 *Linked Open Data (LOD)* ou Dados Abertos Conectados

A conexão e publicação de dados semantizados é proposta do *Linked Open Data*, conceito-chave na visão da Web semântica. Conforme Santarém Segundo e Coneglian (2016),

Essa característica torna os dados não apenas acessíveis e processáveis por máquinas, mas passíveis de processos de organização que podem facilitar a geração de novos dados, apresentação de resultados, relação com outros grupos de dados, aumento do conhecimento para tomadas de decisão, novos modelos de dados gerados a partir do relacionamento e cruzamento de dados de várias esferas governamentais, além da geração de novos modelos mentais de apresentação da informação de forma a facilitar o acesso dos dados pela sociedade civil. (Santarém Segundo; Coneglian, 2016, p. 223).

Como proposto por Tim Berners-Lee, existe uma escala de cinco estrelas a ilustrar o grau de disponibilidade dos dados nos padrões do *LOD*. Esta escala (apresentada na Figura 7) avalia diversos aspectos, incluindo formatos com licenças abertas, o nível de estruturação dos dados, a legibilidade dos dados por máquinas, a publicação de dados na *Web* em formatos não proprietários e a utilização do formato *RDF*. Além disso, verifica-se se os dados em *RDF* contêm *Uniform Resource Identifiers (URIs)*¹¹ e se estão conectados a outras bases de dados.

Figura 7: Esquema de publicação de dados abertos de acordo com a escala de cinco estrelas.



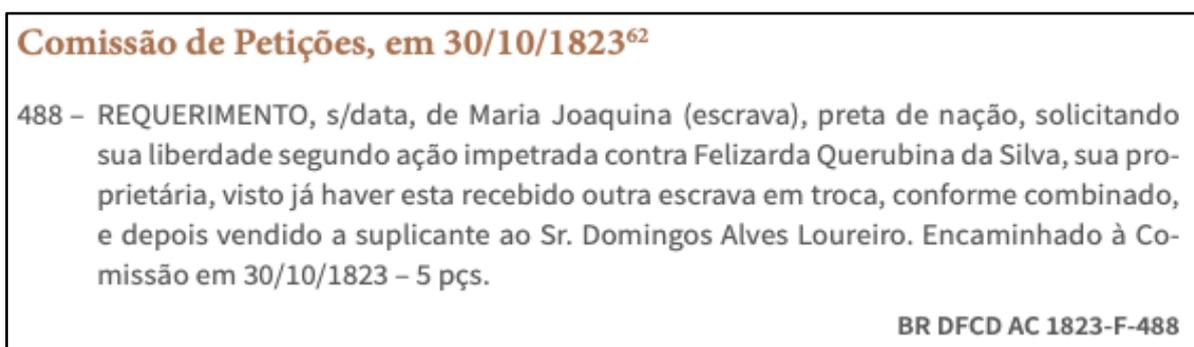
Fonte: Adaptada de Isotani e Bittencourt (2015).

¹¹ “A relação *RDF* com *URI* é um conceito-chave para referenciar e descrever um recurso de forma única e não ambígua” (Isotani; Bittencourt, 2015, p. 59).

No modelo de cinco estrelas, é suficiente os dados estarem disponíveis na *Web* em qualquer formato estruturado para se alcançar a primeira estrela. Contudo, é preciso estarem sob uma licença aberta. Assim, um arquivo divulgado na *Web* nos formatos *PDF*, *PNG* ou *Microsoft Word* pode ser classificado com uma estrela, desde que acompanhado de algum tipo de licença aberta (Paula; Deus; Barbosa, 2020). Como demonstram Isotani e Bittencourt (2015), há inúmeras licenças abertas, como a *Public Domain Dedication and License (PDDL)*, *Attribution License (ODC-By)*, *Open Database License (ODC-ODbL)*, além das recentemente adicionadas *Creative Commons Zero (CC0)*, *Creative Commons Attribution (CC BY)* e *GNU Free Documentation License (GFDL)*.

Quanto às instituições de arquivos, os instrumentos de pesquisa da documentação em formato *PDF* na Internet – inventários, catálogos, listagens e fichas – são classificados com uma estrela pela escala dos dados abertos. É o caso do Inventário Analítico do Arquivo da Assembleia Geral Constituinte e Legislativa do Império do Brasil de 1823¹², fundo documental custodiado pelo Arquivo Histórico da Câmara dos Deputados em Brasília. Como se vê na Figura 8, o trecho extraído do inventário, com a descrição do dossiê referente à petição de Maria Joaquina, escravizada que lutava por sua liberdade:

Figura 8: Descrição de um dossiê do fundo da Constituinte de 1823 em formato *PDF* (uma estrela).



Fonte: Brasil. Congresso Nacional. Câmara dos Deputados. Coordenação de Arquivo (2015).

Apesar da publicação do inventário em *PDF*, sob licença aberta, os dados contidos no instrumento de pesquisa encontram-se aprisionados, com limitada capacidade de manejo e de aproveitamento por usuários em diferentes contextos de uso. Tal situação compromete a acessibilidade dos dados tanto pelas máquinas quanto para indivíduos, consideradas ainda características de apresentação e possibilidade de grande volume. Para alcançar a segunda

¹² Disponível em: <https://arquivohistorico.camara.leg.br/index.php/assembleia-geral-consituente-e-legislativa-do-imperio-do-brasil-1823>.

estrela do modelo de Berners-Lee, além da adoção de licenças abertas, é preciso que os dados estejam publicados de maneira estruturada e em formatos legíveis por máquinas, por exemplo, no formato *Excel*, propriedade da *Microsoft* (Isotani; Bittencourt, 2015; Paula; Deus; Barbosa, 2020).

Caso o Inventário Analítico da Assembleia Constituinte de 1823 tivesse em formato *XLS (Excel)* ou *Google Sheets*, ele poderia elevar o Arquivo da Câmara à condição de possuidor da segunda estrela de dados abertos. Contudo, persistem dois problemas: os dados ainda confinados no documento; e a manipulação dos dados em documentos *XLS* requer o uso de *softwares* proprietários (Isotani; Bittencourt, 2015).

A terceira estrela é alcançada quando os dados estão disponíveis de maneira estruturada na *Web*, em formato não proprietário. Os formatos *XML*, *Comma-Separated Values (CSV)* e *JavaScript Object Notation (JSON)* são exemplos desta categoria, permitindo o uso mais eficaz dos dados, sem a dependência de *softwares* específicos (Isotani; Bittencourt, 2015; Paula *et al.*, 2020).

No exemplo anterior, referente ao dossiê de Maria Joaquina constante no Inventário Analítico da Assembleia Constituinte de 1823, o site do Arquivo da Câmara dos Deputados¹³ oferece a possibilidade de exportar as descrições no formato *CSV*, em padrão *Encoded Archival Description (EAD)*, alcançando, portanto, o nível de três estrelas em dados abertos, como se vê na Figura 9 a seguir.

Figura 9: Descrição do mesmo dossiê do fundo da Constituinte de 1823, em formato *CSV*, padrão *EAD* (três estrelas).

identifier	title	levelOfDescription	extentAndMedium	eventDates	repository	scopeAndContent
488	Requerimento de Maria Joaquina	Dossiê	5 peças Manuscrito em papel	22-09-1823 a 30-10-1823	Câmara dos Deputados	Requerimento, s/data, de Maria Joaquina (pessoa escravizada), preta de nação, solicitando sua liberdade segundo ação impetrada contra Felizarda Querubina da Silva, sua proprietária, visto já haver esta recebido outra escrava em troca, conforme combinado, e depois vendido a suplicante ao Sr. Domingos Alves Loureiro. Encaminhado à Comissão de Petições em 30.10.1823.

Fonte: Adaptada do site do Arquivo Histórico da Câmara dos Deputados. (Brasil. Congresso Nacional. Câmara dos Deputados. Coordenação de Arquivo, 2023).

Embora o formato *CSV* apresente benefícios, como a ampla utilização em bancos de dados públicos e o investimento do *W3C* no desenvolvimento de vocabulários e mapeamento

¹³ Nota-se que metadados podem ser complementados no momento da conversão dos dados do inventário impresso para o site do Arquivo, como ocorreu com as datas no exemplo.

para os formatos *RDF* e *JSON*, ele não oferece a representação detalhada de dados, nem favorece o desenvolvimento de programas computacionais avançados capazes de “extrair múltiplos catálogos de dados, manipulá-los, visualizá-los e combiná-los de maneira flexível” (Isotani; Bittencourt, 2015, p. 49).

Na quarta estrela, os critérios e os padrões técnicos voltados para a semantização e integração de dados tornam-se centrais, especialmente pelo emprego de *URIs* – identificadores únicos e globais de recursos na *Web* – e apresentação no formato *RDF*. Ao ultrapassar a compreensão básica dos formatos *PDF*, *Excel* e *CSV*, característicos das estrelas anteriores, atingem-se, a partir da quarta estrela, elementos fundamentais à *Web* semântica e ao modelo conceitual *RiC*.

Inicialmente, os *URIs* são essenciais porque “facilitam o compartilhamento de cada dado que publicamos [e] são acessíveis e utilizáveis de maneira mais simplificada por pessoas e máquinas” (Isotani; Bittencourt, 2015, p. 50). Para alcançar a quarta estrela, é imprescindível que os dados possuam *URIs*, que os habilitam a:

- a) conectar e combinar estes dados com outros dados;
- b) reusar estes dados em outros contextos;
- c) melhorar a busca e a compreensão dos dados apresentados;
- d) possibilitar inferência a partir de dados parciais;
- e) permitir navegação entre documentos (Isotani; Bittencourt, 2015, p. 53).

Os dados publicados com *URIs* serão representados pelo *RDF*, formato que incorpora o conceito de serialização e pode estar disponível de diversas maneiras, a exemplo do *RDFa*, que integra atributos às *tags* do conhecido *HTML*. O formato *RDF* confere múltiplas informações suplementares, que permitem às máquinas identificarem e manipular recursos na *Web*.

No exemplo do dossiê referente à petição de Maria Joaquina, a obtenção da quarta estrela pelo Arquivo da Câmara depende da atribuição de *URIs* aos dados disponíveis no formato *RDF*. A implantação do modelo conceitual *Records in Contexts* e da sua ontologia associada constitui uma estratégia apropriada para alcançar o nível de quatro estrelas de dados abertos.

A Figura 10 ilustra o dossiê 488 em formato *RDF*, evidenciando a criação do *URI* específico – <http://example.org/resource/record488>. Como o arquivo em formato *CSV* foi convertido para *RDF*, com base no *RiC-Ontology*, para cada linha foi criado um *URI* (no caso, havia apenas o dossiê 488), e cada coluna foi convertida em uma propriedade *RDF*¹⁴.

¹⁴ Foi usado o aplicativo *Open Refine* para conversão, disponível em: <https://openrefine.org/>.

Figura 10: Descrição do mesmo dossiê do fundo da Constituinte de 1823 em formato *RDF* (quatro estrelas).

```
@prefix ns1: <http://www.ica.org/standards/RiC/ontology#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<http://example.org/resource/record488> a ns1:Record ;
  ns1:hasDate "22-09-1823 a 30-10-1823"^^xsd:string ;
  ns1:hasExtentAndMedium ""5 peças
  Manuscrito em papel""^^xsd:string ;
  ns1:hasScopeAndContent "Requerimento, s/data, de Maria Joaquina (pessoa escravizada), preta
  de nação, solicitando sua liberdade segundo ação impetrada contra Felizarda Querubina da Silva,
  sua proprietária, visto já haver esta recebido outra escrava em troca, conforme combinado, e depois
  vendido a suplicante ao Sr. Domingos Alves Loureiro. Encaminhado à Comissão de Petições em
  30.10.1823."^^xsd:string ;
  ns1:hasTitle "Requerimento de Maria Joaquina"^^xsd:string ;
  ns1:isStoredIn "Câmara dos Deputados"^^xsd:string .
```

Fonte: Elaborada pelo autor (2023).

A quinta e última estrela segue todos os critérios para a quarta estrela, como o formato *RDF* e o estabelecimento de *URIs*, além da conexão com dados de outras bases. Um exemplo da quinta estrela seria vincular os dados do dossiê da petição, contendo o nome da interessada, Maria Joaquina, com a data de 22 de setembro de 1823 e o nome da entidade custodiadora, Câmara dos Deputados, aos dados da DBpedia¹⁵. A DBpedia caracteriza-se pela inovação e por extrair informações estruturadas da Wikipédia para publicá-las em padrões de dados cinco estrelas na *Web*. Pelo uso de tecnologias ligadas à Web semântica, a DBpedia transforma o conteúdo textual da Wikipédia em banco de dados semântico, facilitando o acesso e o manejo de informações em formatos mais estruturados e interconectados.

O portal de dados data.gov.uk, iniciativa do governo britânico, se sobressai como exemplo bem-sucedido nos Dados Abertos Conectados, atualmente contando com a publicação de 47 mil conjuntos de dados. Isotani e Bittencourt (2015) apontam que, em um período anterior, quando o portal gerenciava aproximadamente 19 mil catálogos de dados, perto de 14 mil foram classificados com menos de uma estrela da escala, enquanto apenas 168 atingiram a abertura e a conectividade da classificação de cinco estrelas – ver Quadro 3.

¹⁵ O DBpedia contém 850 milhões de triplas em *RDF* e 55 mil propriedades cadastradas. Ainda existem 1.792.308 pessoas, 748.372 localidades, 610.589 obras e 345.523 organizações, 1.933.436 espécies, 7.718 plantas e 10.591 doenças (Holze, 2023).

Quadro 3: Classificação dos conjuntos de dados do portal data.gov.uk, segundo o esquema de classificação cinco estrelas.

Número de estrelas	Quantidade de catálogos
-	13.965
★	281
★ ★	1.145
★ ★ ★	2.345
★ ★ ★ ★	-
★ ★ ★ ★ ★	168

Fonte: Isotani e Bittencourt (2015).

A classificação menor que uma estrela, ou em “zero” estrela, pode ser hipoteticamente aplicada à maior parte dos arquivos brasileiros, pois eles enfrentam dificuldades fundamentais relacionadas ao tratamento e à preservação de documentos. A descrição documental, com baixa elaboração e publicação de instrumentos de pesquisa na *Web*, é ainda uma realidade em poucas instituições arquivísticas no Brasil. Anna Carla Almeida Mariz, professora de arquivologia na Universidade Federal do Estado do Rio de Janeiro (UFRJ), ao analisar os sites de instituições arquivísticas brasileiras, evidencia a carência de recursos humanos, financeiros, materiais e tecnológicos dos arquivos. A falta de investimento se reflete na escassa presença de sites institucionais para a divulgação de acervos. Mariz (2012) afirma que “se os arquivos não estão organizados, não podem ser disponibilizados na sala de consulta nem na internet”, ressaltando a importância da organização e da descrição dos arquivos para a sua utilização (Mariz, 2012, p. 146).

Na proposta de estruturar dados abertos conectados pelo esquema de cinco estrelas, de modo a facilitar a compreensão e o processamento por computadores, o formato *RDF* é essencial por organizar e etiquetar informações, esclarecendo tanto o seu conteúdo como as relações entre elas.

Para iniciar o entendimento sobre o formato *RDF*, pode-se usar o dossiê relativo à petição de Maria Joaquina. O documento possui metadados descritivos, como o número identificador do dossiê, título, dimensão, suporte, datas e pessoas interessadas. Tradicionalmente, apesar de tais informações estarem registradas em instrumentos de pesquisa

impressos – inventários ou fichas descritivas – ou mesmo em bases de dados e sites, elas, em geral, não estão em formatos que permitam aos computadores compreenderem as conexões desses dados com o dossiê em questão.

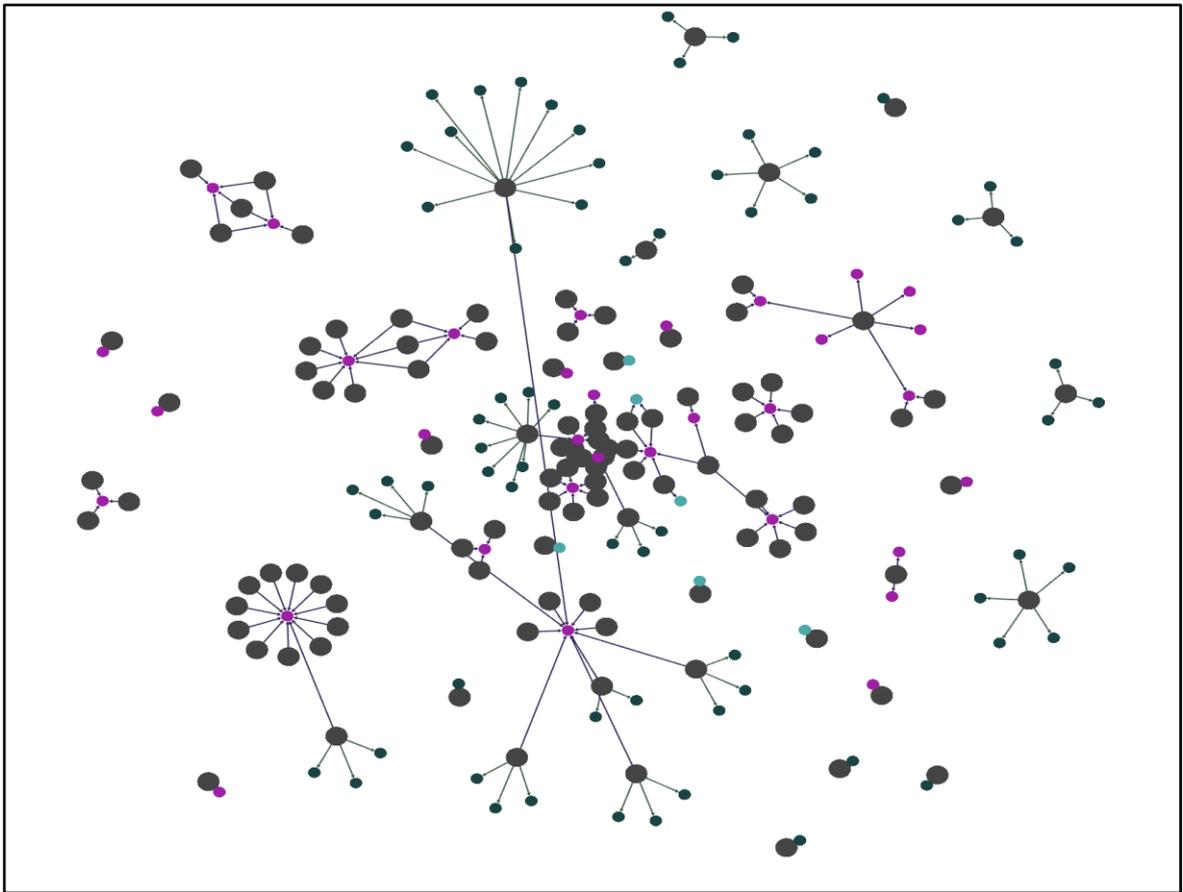
O *RDF* transforma os dados em declarações simples, semelhantes a frases. Cada declaração em *RDF* é composta por três elementos: um sujeito, um predicado e um objeto. Exemplo: na petição de Maria Joaquina, uma declaração *RDF* poderia ser “Esta petição (sujeito) solicita (predicado) justiça e liberdade (objeto)” ou “Esta petição (sujeito) foi solicitada (predicado) no Rio de Janeiro (objeto)”. Tais declarações são chamadas de “triplas” e permitem aos computadores não apenas entender o documento em si como, também, as relações entre os documentos e seus dados.

Então, se houver outros documentos no Arquivo da Câmara dos Deputados relacionados a pedidos de liberdade e justiça ou relativos ao Rio de Janeiro, o computador poderá compreender as conexões, desde que estejam declaradas em formato *RDF*. Isotani e Bittencourt (2015) explicam que à medida que mais vínculos descritivos entre os recursos da *Web* são estabelecidos, originam-se conjuntos de triplas que, por sua vez, compõem grafos de dados. Ademais, é possível a integração de múltiplos grafos, reforçando a interconectividade e a complexidade da informação na *Web* semântica.

Iniciativas têm sido tomadas para a criação de grafos de dados em projetos de *Web* semântica, nos variados segmentos, como bibliotecas, arquivos e museus. Uma experiência notável é a do Ministério da Cultura da França, descrita por Marie-Véronique Leroi, do Departamento de Inovação Tecnológica. Leroi (2021) relata que, em 2014, a instituição lançou a publicação do relatório *Dados culturais e a transição para Web 3.0* [tradução nossa], visando à promoção, entre agentes culturais franceses, da pertinência de aplicar a *Web* semântica aos dados institucionais. A meta principal é aumentar a visibilidade dos acervos e aprimorar a recuperação das informações. O objetivo geral da iniciativa é a constituição de um grafo cultural ou grafo de conhecimentos culturais da França, estimulando, assim, a capacitação dos profissionais do patrimônio documental cultural, com vistas à eficácia na estruturação e integração dos dados.

Segundo Leroi (2021), os arquivos têm um papel decisivo na estratégia semântica, “pela diversidade dos documentos que tratam, bem como pela natureza transversal deles” (Leroi, 2021, p. 22). A Figura 11 mostra um conjunto de triplas do Arquivo Nacional da França, que forma o grafo de dados da área de patrimônio e cultura daquele país.

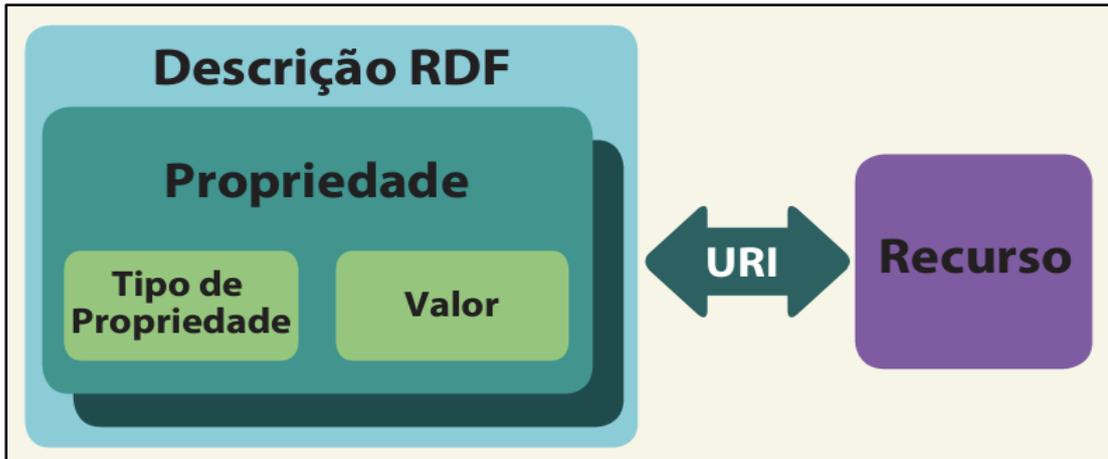
Figura 11: Exemplo de grafo de dados do Arquivo Nacional da França com fundos, documentos e autoridades relacionadas.



Fonte: ARCHIVES NATIONALES. Service Interministériel des Archives de France. Bibliothèque Nationale de France ([2017]).

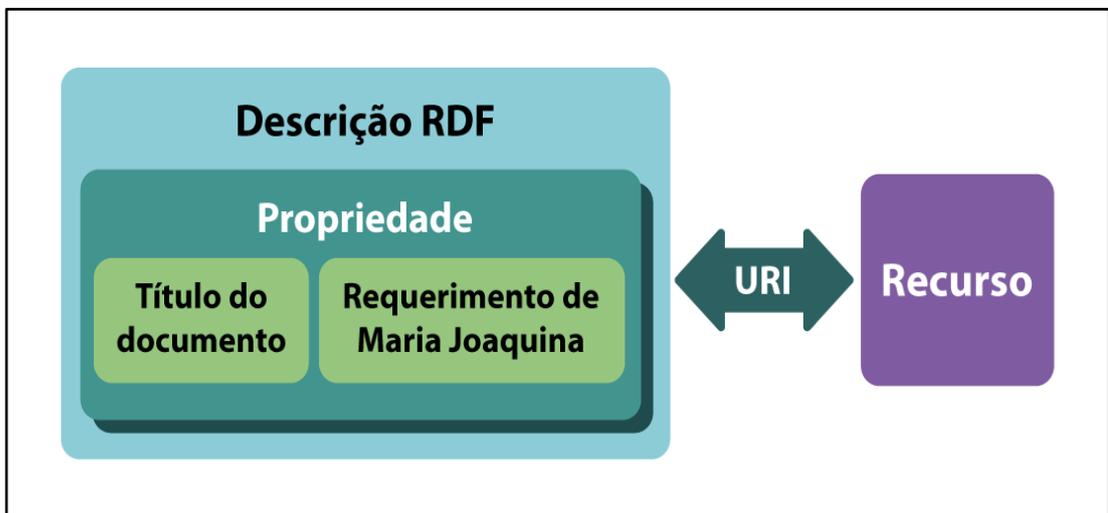
Não de outra forma, o modelo *RDF*, usado para a criação de grafos de dados, é empregado com o propósito de descrever recursos na *Web*. É imprescindível a exata identificação destes elementos. Para tal, adota-se o *URI*, identificador único e global. Conforme Ian Jacobs e Norma Walsh, membros da *W3C*, a arquitetura da *Web* fundamenta-se em três pilares: (1) o recurso único, garantido pelo *URI*, é composto de uniformidade, recurso e identificador; (2) a interação, caracterizada pela comunicação entre o servidor *Web* e o navegador cliente, mediada pelo protocolo *HTTP* e por linguagens, como o *HTML*; (3) os formatos de representação, que tornam disponíveis dados e metadados, exemplificados pelo *RDF*, *XHTML*, principalmente (Jacobs; Walsh, 2004). Portanto, no *RDF*, um recurso *Web* é descrito por um conjunto de atributos conhecidos como Descrição *RDF*, em que cada atributo incorpora um tipo de propriedade e um valor, além de cada propriedade ser individualizada por um *URI* (Figuras 12 e 13).

Figura 12: Representação abstrata da descrição em *RDF*.



Fonte: Isotani e Bittencourt (2015).

Figura 13: Representação concreta do *RDF* para o exemplo do requerimento de Maria Joaquina.



Fonte: Adaptada de Isotani e Bittencourt (2015).

O *RDF* é o mecanismo eficaz para estruturar informações para que os computadores trabalhem de forma eficiente com os dados, compreendendo seus detalhes individuais e as relações entre eles. Conforme Nigel Shadbolt, professor na Universidade de Oxford, Wendy Hall, professora na Universidade de Southampton e Tim Berners-Lee, o objetivo do *RDF* é prover uma representação minimalista do conhecimento publicado, ao mesmo tempo que uma

riqueza de expressividade¹⁶ na descrição das informações na *Web*. Para os autores, o *RDF* detém um papel-chave no aumento da interoperabilidade na *Web*, além de ser o padrão global para a descrição de informações, de modo a facilitar a compreensão e o uso eficaz dos dados por diferentes sistemas (Shadbolt *et al.*, 2006).

Pedra angular da *Web* semântica e do conceito de *Linked Open Data*, o *RDF* foi desenvolvido em 1997 e tornou-se recomendação do *W3C*, em 1999. A versão inicial estabeleceu as bases do modelo, introduzindo os conceitos de sujeito, predicado e objeto – as triplas. Em 2000, o avanço marcante se deu pela publicação do *RDF Schema*, que aprimorou a descrição dos recursos da *Web* e suas relações. A atualização de 2004 do modelo reduziu ambiguidades e melhorou a interoperabilidade, especificando o vocabulário de sintaxe *RDF* e *RDF-S*, este último para a representação e modelagem de conceitos com maior expressividade. Em 2013, foi lançada a *Protocol and RDF Query Language (SPARQL)*, linguagem específica de consulta em *RDF*, melhorando bastante as capacidades de consulta e atualização dos dados. Em 2014, a publicação do *RDF 1.1* simplificou as regras e padronizou a serialização dos dados, incluindo os formatos *RDFa* e *RDF/XML*, além de introduzir novos elementos, como o *RDF Dataset*, um conjunto de grafos *RDF* nomeados, que proporcionou aperfeiçoamento e flexibilidade na modelagem de dados. Atualmente, o *W3C* desenvolve a versão *RDF 1.2* (World Wide Web Consortium, 2023).

Assim, nos últimos anos, o *RDF* apresentou padrões de representações para codificar a estrutura dos dados a serem selecionados, conforme a necessidade do projeto de aplicação. Para a World Wide Web Consortium (2014), a serialização do *RDF* consiste em representar os dados de forma seriada, adequada para o armazenamento ou transmissão em diferentes plataformas. Existem vários formatos de serialização – *RDF*, *RDFa*, *RDF/XML*, *N-Triples*, *Turtle*, *JSON-LD*, *TriG* e *N-Quads* –, cada um com regras próprias para codificar a estrutura dos dados *RDF* em uma sequência de caracteres ou *bytes*. Estes formatos são projetados para diferentes casos, com objetivos que variam em termos de legibilidade humana e facilidade de processamento por máquinas, como mostra o Quadro 4.,

¹⁶ Em Lógica, define-se expressividade como “a capacidade de uma linguagem definir um sujeito formalmente” (Noronha, 2021, p. 25).

Quadro 4: Formatos de serialização do *RDF* e suas características de uso.

Serialização RDF	Tipo de código	Quando usar
RDFa	Código RDF embutido em HTML	SEO (Search Engine Optimization)
RDF/XML	Código RDF com estrutura em XML	Aplicações que usam estruturas em XML
JSON-LD	Código RDF com estrutura JSON	Aplicações que usam JSON
N-Triples	Código RDF com estrutura de triplas	Processamento e intercâmbio de Big Data em RDF
Turtle	Código RDF para facilitar a leitura humana	Processamento e intercâmbio de Big Data em RDF
TriG	Código com estrutura Turtle	Representação de múltiplos grafos
N-Quads	Código RDF com estrutura de triplas	Processamento e intercâmbio de grandes catálogos de dados

Fonte: Isotani e Bittencourt (2015).

Por fim, na lógica da Web semântica, o *RDF* detém um papel determinante na descrição precisa dos recursos *Web*, fornecendo a estrutura detalhada para os dados ou descrição sintática. Além disso, o *RDF* amplia o seu escopo para a elaboração de relações e significados entre os recursos, ao aprofundar a descrição semântica (Isotani e Bittencourt, 2015). Como será tratada na subseção subsequente, a dimensão semântica da *Web* é enriquecida quando o modelo *RDF* integra-se a modelos conceituais ou ontologias de áreas específicas do conhecimento, como o modelo *RiC*, no campo dos arquivos; o *FRBR*, na área de bibliotecas; e o *CIDOC-CRM*, no âmbito dos museus. Tal integração aprofunda a descrição das relações entre os dados, tanto para a compreensão humana quanto para o processamento por máquinas.

2.1.4 Ontologias

Na Web semântica, a estruturação de dados não se dá de maneira aleatória para não prejudicar o uso posterior das informações. Isotani e Bittencourt (2015, p. 93) frisam que “um dado pode ser descrito, representado e interpretado de diferentes maneiras”. Como visto no tópico anterior, é necessário usar padrões *RDF*, o *RDF-S* e suas respectivas serializações, recomendados pelo *W3C*, para uma estruturação de dados adequada. Esta subseção discorre acerca do aprofundamento do chamado enriquecimento semântico dos dados por meio da

criação e aplicação de ontologias e modelos conceituais de dados, representações de conhecimento em domínios específicos.

É fundamental compreender as nuances da representação do conhecimento, como a escolha de vocabulário para representar um conceito específico, dada a ampla variedade de termos disponíveis. Isotani e Bittencourt (2015) citam as diferentes denominações para uma pessoa que assiste às aulas em uma universidade – **aluno**, **estudante**, **aprendiz** ou **universitário** – e para alguém que leciona – **professor**, **instrutor** ou **mestre**. Uma instituição que deseja publicar seus dados abertos precisa escolher um conjunto de termos, um vocabulário, que descreva adequadamente os seus dados.

Além da escolha do vocabulário, é possível aumentar a expressividade dos dados no contexto em que foram produzidos, com a intenção de minimizar ambiguidades de interpretação. A representação envolve a explicação adicional dos dados depois da seleção do vocabulário. Exemplo: ao esclarecer que **alunos** e **professores** exercem **papéis** em uma universidade e que **pessoas** ocupam estas funções, ou que é necessário um título de **mestre** ou **doutor** para ser professor, os dados são enriquecidos e melhor representados. Dependendo de como são mostradas, as restrições podem auxiliar ou causar problemas nas futuras interpretações dos dados por humanos e máquinas. Para formalizar as relações de hierarquia e restrições entre conceitos, de modo que humanos e máquinas compreendam os conceitos que representam os dados disponíveis, além da criação de ontologias, são necessárias as linguagens de representação: *RDF*, *RDF-S* e *OWL* (Isotani; Bittencourt, 2015).

Na Web semântica, a fim de compreender os conceitos de representação e expressividade, antes de mais, é preciso entender as ontologias. Esta pesquisa identificou a obrigação de esclarecer as semelhanças e diferenças entre os conceitos de ontologia e de modelos conceituais de dados. Há proximidade entre os conceitos, mas o seu uso varia de acordo com a perspectiva da área acadêmica e o tipo de aplicação. Para Maurício Barcellos Almeida, doutor em Ciência da Informação, as ontologias são exploradas sob as óticas da Filosofia, da Ciência da Computação e da Ciência da Informação, como mostra o Quadro 5. A pesquisa ainda averiguou a aplicação dos dois conceitos, intrinsecamente relacionados, em Ciência da Computação e no uso feito pelo Conselho Internacional de Arquivos para o modelo *Records in Contexts*.

Quadro 5: Uso de ontologias por campos científicos.

Distinção	Campo	O que é?	Propósito	Exemplo
Ontologia como uma disciplina	Filosofia	Ontologia como um sistema de categorias	Entender a realidade, as coisas que existem e suas características	Sistemas de Aristóteles, Kant, Husserl
Ontologia como um artefato	Ciência da Computação	ontologia como uma teoria (baseada em lógica)	Entender um domínio e reduzi-lo à modelos	BFO, DOLCE (genéricas)
		ontologia como um artefato de software	Criar um vocabulário para representação em sistemas e para gerar inferências	OWL (linguagem de RC)
	Ciência da Informação	ontologia como uma teoria (informal)	Entender um domínio e classificar termos	Sistema de classificação de Ranganathan
		ontologia como um sistema conceitual informal	Criar vocabulários controlados para recuperação da informação a partir de documentos	um catálogo, um glossário, um tesaurus

Fonte: Almeida (2014).

Almeida (2014) revela que, na Ciência da Informação, os princípios ontológicos, concentrados no estabelecimento de regras e ideias para categorizar e organizar informações, são considerados valiosos. Para o teórico, as ontologias são comumente aplicadas na elaboração de sistemas de classificação e na descrição do conteúdo dos documentos. Entretanto, o uso se estende além desta função primária, englobando a compreensão aprofundada do contexto de interação de grupos específicos de usuários com informações e documentos. Tal aspecto adiciona complexidade à Ciência da Informação, pois permite uma análise mais detalhada e ajustada às necessidades e comportamentos dos usuários.

O uso do termo ontologia em Ciência da Computação remonta aos anos 1960, no subcampo da Representação do Conhecimento (RC), vertente da Inteligência Artificial (IA). A partir da década de 1990, o termo também é empregado na Web semântica. Para este domínio, Almeida (2014) define ontologia como uma teoria representativa dos fatos e regras que governam a realidade, com objetivos computacionais. Ele sustenta que, na Ciência da Computação, as ontologias são aplicadas, inicialmente, à modelagem de sistemas baseados em banco de dados e esclarece que “o uso do termo nesse contexto está alinhado com seu papel original na Filosofia, isto é, fornece uma descrição do que existe e caracterizar entidades nas atividades de modelagem” e, num segundo momento, elas são usadas em sistemas de representação do conhecimento, que consistem “em um conjunto de declarações expressas em uma linguagem de representação, que pode ser processada por mecanismos de inferência automatizados” (Almeida, 2014, p. 249). A perspectiva sublinha o valor das ontologias tanto

para a estruturação de bancos de dados quanto para a elaboração de sistemas capazes de realizar inferências a partir de conhecimentos representados. E conclui que:

O termo ontologia em Ciência da Computação é usado para se referir tanto a um vocabulário expresso em uma linguagem de RC, quanto a um tipo de teoria onde fenômenos são explicados a partir de fatos e regras. O primeiro uso corresponde a um *software*, um artefato computacional. O segundo uso mantém a noção filosófica, um inventário de coisas do mundo e relações entre elas em um domínio particular, baseado nos princípios da Ontologia como disciplina (Almeida, 2014, p. 253).

Ainda na Ciência da Computação, no âmbito da Web semântica, define-se ontologia como “um conjunto de conceitos fundamentais e suas relações, que capta como as pessoas entendem (ou interpretam) o domínio em questão e permite a representação de tal entendimento de maneira formal, compreensível por humanos e computadores”, segundo o cientista japonês Riichiro Mizoguchi (2004 *apud* Isotani; Bittencourt, 2015, p. 96).

Na Ciência da Computação, sobretudo na Inteligência Artificial, Isotani e Bittencourt (2015) elucidam que as ontologias são estruturadas por dois elementos. O primeiro consiste num conjunto de conceitos resultado do conhecimento básico de áreas como medicina, agronomia ou arquivologia, e que, para descrevê-los, emprega-se um vocabulário específico. O segundo elemento é um corpo de conhecimentos detalhado da área em questão, que inclui a hierarquização em classes e subclasses, ilustrando como os conceitos se relacionam, seja com o uso da expressão **é um** ou **faz parte de**. A atividade engloba a definição de regras e restrições entre os conceitos e as suas inter-relações. Os autores complementam que:

Muitas vezes, durante o desenvolvimento de uma ontologia, as pessoas acabam gastando muito tempo para discutir a terminologia a ser utilizada (isto é, vocabulário), em vez de discutir e compreender os conceitos essenciais do domínio. Gostaríamos de enfatizar que, do ponto de vista dos autores, uma ontologia não é um conjunto de termos interligados que formam um vocabulário [...] para criar uma ontologia, o enfoque do problema é a definição de um conceito; e, portanto, o rótulo dado a ele passa a ter menos importância. Resumindo, uma ontologia propõe-se a representar e expressar o significado de conceitos, e não de termos (Isotani; Bittencourt, 2015, p. 97).

Nicola Guarino, cientista da computação italiano, propõe a classificação de ontologias em várias categorias: ontologias de informação, terminológicas, de modelagem de conhecimento, de aplicação, genéricas e de representação. De maneira complementar, há outras classificações: ontologias pesadas e leves, bem como ontologias de domínio e de tarefa (Guarino, 1997; Isotani; Bittencourt, 2015).

As ontologias leves se caracterizam por não definir detalhadamente cada conceito, apenas apresentam a taxonomia e a relação hierárquica entre conceitos, amplamente usadas na *Web* para categorizar grandes volumes de dados. As ontologias pesadas ultrapassam a elaboração de taxonomias simples, pois elas mostram uma representação semântica detalhada e rigorosa dos conceitos, definidos e organizados com base em princípios claros, além da definição formal das relações entre eles. Ontologias pesadas são essenciais à criação de bases de conhecimento a serem usadas e compartilhadas. As ontologias de domínio e de tarefa são fundamentais para o desenvolvimento de sistemas flexíveis e inteligentes, aplicáveis em diversos domínios. A ontologia de domínio concentra-se no conhecimento específico de um tópico (medicina ou automóvel), e a ontologia de tarefa, na aplicação desse conhecimento na resolução de problemas (diagnóstico ou venda) (Isotani; Bittencourt, 2015).

Na área dos arquivos, com base nas classificações, é possível considerar o modelo conceitual *RiC*, mais condizente com a categoria da ontologia de domínio, já que foi desenvolvido especificamente para definir e caracterizar os arquivos, utilizando, para tal, termos e conceitos, todos originários deste segmento do saber. O *RiC* é mais que uma mera taxonomia, por apresentar a descrição abrangente e detalhada dos conhecimentos e práticas do mundo dos arquivos. O *RiC* também trata da semântica das complexas relações entre os elementos, aproximando-o das características de uma ontologia pesada.

É possível atribuir ao modelo *RiC* a categoria de ontologia de alto nível, classificação reservada para ontologias que, conforme Maria Luiza Campos, da Universidade Federal Fluminense (UFF), “abordam conceitos extremamente gerais, como espaço, tempo, matéria, objeto, evento, ação, que são independentes de um problema específico ou domínio” (Campos, 2010, p. 224). Compreende-se que, embora o *RiC* se dedique ao domínio específico dos arquivos, ele não se aprofunda nos conceitos específicos relacionados a tipos distintos de arquivos especializados, como arquivos legislativos, universitários, judiciários, municipais, audiovisuais. A observação converge com as definições do próprio Conselho Internacional de Arquivos:

O *RiC-CM* é um modelo conceitual de alto nível, que se concentra na identificação e descrição intelectual dos documentos; das pessoas que os criaram e utilizaram; e das atividades realizadas por essas pessoas às quais os documentos dão suporte ao mesmo tempo em que as registram. Como modelo de alto nível, o *RiC-CM* é constituído por um amplo arcabouço conceitual. Ele não abrange a totalidade das atividades necessárias para gestão de documentos, nem detalha completamente todos os aspectos que podem ser exigidos em todos possíveis contextos de aplicação (International Council on Archives, 2023, p. 1, tradução nossa).

Oportuno realçar que a iniciativa do Conselho Internacional de Arquivos resultou na publicação de dois textos importantes: o modelo conceitual *Records in Contexts* e a correspondente ontologia *Records in Contexts-Ontology*. O primeiro documento, o modelo *RiC*, constitui um esquema conceitual sobre os arquivos, claramente definido e acessível ao entendimento humano. Já o segundo, o *RiC-O*, formaliza o modelo para a linguagem computacional, possibilitando a sua interpretação e processamento pelas máquinas. É imprescindível a compreensão integrada das duas ferramentas, apesar das diferenças conceituais entre modelos conceituais de dados e ontologias. Os termos não são sinônimos, mas ambos são usados no âmbito da organização do conhecimento e informação e em Ciência da Computação.

Segundo os professores especialistas em Tecnologia da Informação (TI) Jugurta Lisboa, da Universidade Federal de Viçosa (UFV), e Cirano Iochpe, da Universidade Federal do Rio Grande do Sul (UFRGS), os modelos conceituais de dados têm a função primordial de mapear a estrutura de organização, armazenamento e relacionamento dos dados em sistemas de bancos de dados, aderindo a regras consagradas, como o modelo Entidade-relacionamento (ER ou MER)¹⁷ ou o modelo Orientação a Objetos (OO) (Lisboa Filho; Iochpe, 1999). Em contraste com os modelos conceituais, as ontologias são empregadas sobretudo para representar conhecimentos de um domínio específico, incluindo regras e restrições para a representação dos conceitos e suas inter-relações. Caracterizadas pela abstração e flexibilidade, as ontologias são utilizadas para favorecer a interoperabilidade e o entendimento mútuo entre diferentes sistemas e contextos, logo elas não são elaboradas para bancos de dados.

Os modelos conceituais são elaborados com o objetivo de aprimorar o desempenho dos bancos de dados e assegurar a sua integridade, de acordo com padrões específicos de modelagem de dados, como confirmado por Lisboa Filho e Iochpe (1999):

O resultado do processo de modelagem, denominado esquema conceitual, é apresentado através de uma linguagem formal de descrição que pode estar expressa através de uma sintaxe e/ou uma notação gráfica. Para cada formalismo conceitual podem existir diversas linguagens de descrição de esquema que são compatíveis com o formalismo. O formalismo provê um conjunto de conceitos, elementos e regras que são usados no processo de modelagem da realidade, enquanto que a linguagem de descrição fornece uma gramática para a apresentação do esquema conceitual resultante da modelagem (Lisboa Filho; Iochpe, 1999, p. 70).

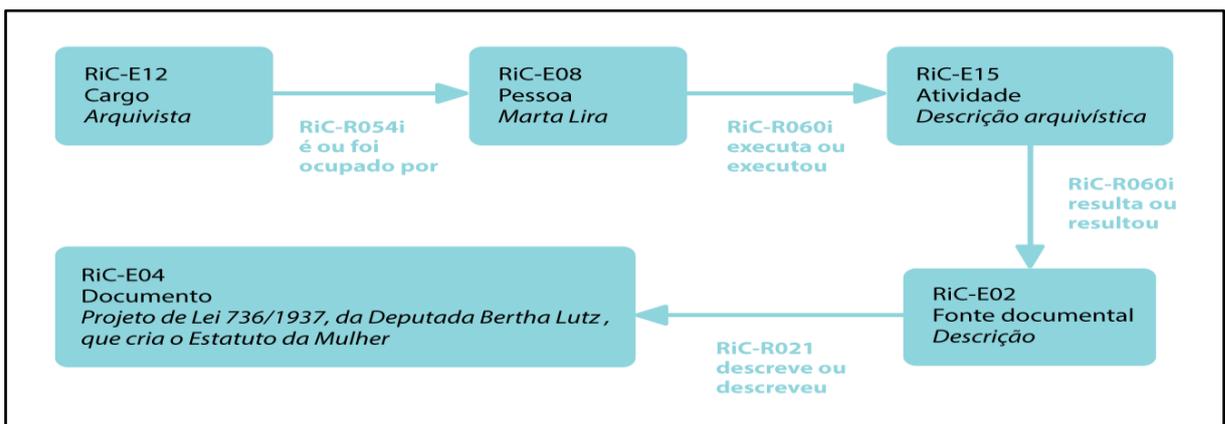
¹⁷ O modelo Entidade-relacionamento, apresentado por Peter Chen, em 1976, é considerado um marco na história da modelagem e de projeto de banco de dados e, como consequência, é importante em projetos de sistemas de informação. Ele se baseia em conceitos simples e possui um fácil entendimento, fornecendo uma perspectiva vigorosa para a modelagem de dados (Maia; Alvarenga, 2013).

Os professores sublinham que modelos conceituais de dados incorporam sempre um formalismo e uma linguagem de descrição. Ela pode assumir duas formas: a linguagem léxica, que permite o processamento computacional do esquema conceitual e a linguagem gráfica, preferível para facilitar a compreensão e a comunicação entre os seres humanos (Lisboa Filho; Iochpe, 1999). No campo dos arquivos, a linguagem léxica corresponde à ontologia *RiC-O*, pois garante o processamento computacional do modelo conceitual *RiC-CM*.

Atualmente, as linguagens léxicas e gráficas são compreendidas de duas formas na representação de ontologias. A representação formal é essencial para que os computadores interpretem as ontologias, enquanto a representação gráfica facilita o entendimento humano por grafos, linguagem de modelagem unificada (UML), estrutura de árvore e afins (Isotani e Bittencourt, 2015).

Para melhor compreensão de uma representação gráfica, podemos analisar, de modo hipotético, a operação na qual um arquivista descreve um documento de arquivo, apoiado no modelo conceitual e na ontologia do *RiC* (Figura 14). Observa-se o seguinte fluxo: inicialmente, a entidade **cargo**, neste contexto representado pelo valor **arquivista**. Este cargo **é ou foi ocupado** por uma **pessoa**, outra entidade relevante no *RiC*, que, neste exemplo, é identificada como **Marta Lira**. Essa pessoa, por sua vez, **executa ou executou** uma **atividade**, que, neste caso, está associada à **descrição arquivística**. Tal atividade **resulta ou resultou** em uma **fonte documental**, identificada, aqui, como **descrição**, podendo ser uma ficha de descrição em papel, um registro em uma base de dados ou no site do arquivo. Por fim, o registro documental **descreve ou descreveu** um **documento**, outra entidade do modelo *RiC*, que, neste exemplo, poderia ser um documento específico dentro do acervo, tendo como valor o **Projeto de Lei 736/1937, da deputada Bertha Lutz, que cria o Estatuto da Mulher**.

Figura 14: Representação gráfica da descrição de um documento de arquivo, segundo o modelo *RiC*.



Fonte: Adaptada do ICA (2023).

Segundo os autores Isotani e Bittencourt (2015), na esfera da representação formal para computadores, inúmeras linguagens podem ser empregadas, inclusive a lógica de predicados, a lógica descritiva e as linguagens baseadas em quadros (*frames*). Entre as mais expressivas para descrever ontologias, constam o *RDF*, o *RDF-S* e a *OWL*. Ao se utilizar o exemplo da operação de descrição de um documento de arquivo por um arquivista, a que se vê abaixo é a representação formal dos mesmos dados, em linguagem *OWL*, para que o computador compreenda:

```
<rdf:RDF
  xmlns:rico="https://www.ica.org/standards/RiC/ontology#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">

  <!-- Classe Position -->
  <owl:Class
rdf:about="https://www.ica.org/standards/RiC/ontology#Position">
    <rdfs:label xml:lang="en">Position</rdfs:label>
    <rdfs:comment>A functional role of a Person within a
Group.</rdfs:comment>
  </owl:Class>

  <!-- Indivíduo Arquivista da classe Position -->
  <rico:Position rdf:about="#Arquivista">
    <rdfs:label>Arquivista</rdfs:label>
  </rico:Position>

  <!-- Propriedade isOrWasOccupiedBy -->
  <owl:ObjectProperty
rdf:about="https://www.ica.org/standards/RiC/ontology#isOrWasOccupie
dBy">
    <rdfs:label xml:lang="en">is or was occupied by</rdfs:label>
    <rdfs:comment>Inverse of 'occupies or occupied' object
property.</rdfs:comment>
    <rdfs:domain
rdf:resource="https://www.ica.org/standards/RiC/ontology#Position"/>
    <rdfs:range
rdf:resource="https://www.ica.org/standards/RiC/ontology#Person"/>
  </owl:ObjectProperty>
```

```

    <!-- Classe Person -->
    <owl:Class
rdf:about="https://www.ica.org/standards/RiC/ontology#Person">
        <rdfs:label xml:lang="en">Person</rdfs:label>
        <rdfs:comment>A human being with a social identity or
persona.</rdfs:comment>
    </owl:Class>

    <!-- Indivíduo Marta Lira da classe Person -->
    <rico:Person rdf:about="#MartaLira">
        <rdfs:label>Marta Lira</rdfs:label>
    </rico:Person>

    <!-- Relacionamento entre Marta Lira e a posição Arquivista -
->
    <rico:Position rdf:about="#Arquivista">
        <rico:isOrWasOccupiedBy rdf:resource="#MartaLira"/>
    </rico:Position>
</rdf:RDF>

```

Na representação formal em *OWL*, são apresentadas a classe **Position** e a classe **Person**, com um indivíduo **Arquivista** da classe **Position** e um indivíduo **Marta Lira** da classe **Person**. Além disso, define-se a propriedade **isOrWasOccupiedBy** para estabelecer uma relação entre a posição e a pessoa.

Como vista na subseção anterior, a especificação *RDF*, proposta pelo *W3C*, utiliza triplas compostas por um **nó** sujeito, um predicado e um **nó** objeto (sujeito, predicado, objeto). Esta forma de representação, considerada simples e eficaz, permite que os computadores compreendam as relações entre os dados e a semântica neles contida, como mostrado no caso hipotético da petição de Maria Joaquina. O *RDF Schema (RDF-S)* representa o amadurecimento do modelo *RDF*, permitindo expressar a semântica das triplas por meio da definição de diversas *tags*, além da especificação de classes, propriedades e relacionamentos (Isotani; Bittencourt, 2015). Contudo, mesmo o *RDF-S* sendo um avanço dos primeiros modelos, ele contém várias limitações para o raciocínio computacional dos dados na *Web*. Por tal razão, com o suporte do *W3C*, a comunidade da *Web* semântica desenvolveu uma linguagem ainda mais expressiva.

A *Web Ontology Language* (OWL) é uma linguagem da Web Semântica projetada para representar, de forma enriquecida e complexa, o conhecimento sobre coisas, grupos de coisas e relações entre coisas. OWL é uma linguagem baseada em lógica computacional, de modo que o conhecimento expresso em OWL pode ser explorado por programas de computador, por exemplo, para verificar a consistência do conhecimento ou explicitar conhecimentos implícitos (World Wide Web Consortium, 2013).

A OWL incorpora características do *RDF*, como a estrutura de triplas e a descrição de recursos por meio de *URIs*, além da semântica do *RDF Schema*, mas não se confunde nem com uma linguagem de programação, nem com uma linguagem de esquema para conformidade sintática e, tampouco, é um banco de dados (Isotani; Bittencourt, 2015). A documentação publicada pelo W3C torna a OWL compreensível a diferentes públicos e não exige especialistas em tecnologia para a sua assimilação, como aponta o pesquisador Guilherme Noronha, da Universidade Federal de Minas Gerais (UFMG).

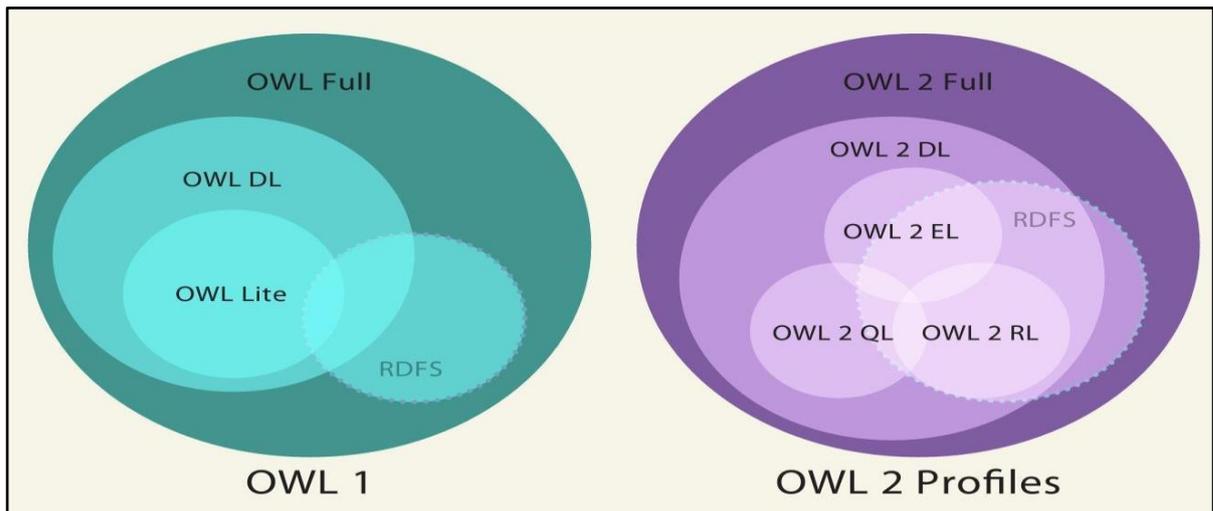
Segundo o pesquisador (2021), um dos maiores recursos semânticos da linguagem OWL é o poder de inferência, operado por construtores da lógica de descrição, que possibilita novos conhecimentos. Assim, a OWL permite fabricar ontologias e colocá-las à disposição para o aproveitamento em sistemas de organização do conhecimento, e, com a lógica de descrição acoplada, beneficia-se da capacidade de processamento computacional, com representação e expressão semântica.

A OWL possui duas dimensões fundamentais: a sintaxe e a semântica. A sintaxe obrigatória da OWL é baseada em *RDF/XML* e das suas alternativas de serialização – *OWL/XML*, *Funcional*, *Manchester* e *Turtle* – escolhidas de acordo com a necessidade do projeto. Semelhante ao *RDF*, a linguagem OWL também inclui as classes **owl:Class**, as propriedades **owl:ObjectProperty** e **owl:DataProperty**, além das instâncias ou indivíduos **owl:Individual**. Entretanto, a OWL se diferencia pela sua maior expressividade e capacidade de especificação, o que amplia o potencial de inferência e novas derivações, em virtude de recursos avançados, como axiomas de equivalências e entidades complexas (Isotani; Bittencourt, 2015).

Segundo a World Wide Web Consortium (2013), a linguagem OWL, desenvolvida entre 2002 e 2004, tem as suas raízes nas linguagens de ontologia *DARPA Agent Markup Language* (DAML) e *Ontology Inference Layer* (OIL) e foi aprimorada com o endosso do W3C. Com a versão OWL 1.0, ela se estabelece como padrão para a modelagem de ontologias, oferecendo alternativas como *OWL Lite*, *OWL DL* e *OWL Full*, cada uma atendendo a diferentes necessidades de expressividade e complexidade.

Em 2009, o W3C publica a OWL 2, uma atualização relevante por realçar a expressividade e a eficiência computacional da linguagem. Esta versão introduz os perfis OWL 2 EL, OWL 2 QL e OWL 2 RL, que atende a uma gama mais ampla de requisitos de desempenho e complexidade. Desde então, a OWL sempre é atualizada pelo W3C, consolidando-se como instrumento essencial no desenvolvimento de ontologias no ambiente digital, como mostra a Figura 15.

Figura 15: Perfis de expressividade das versões OWL 1 e OWL 2.



Fonte: Isotani e Bittencourt (2015).

A criação de ontologias para uso na Web semântica serve-se de metodologias da Engenharia de Ontologias, surgida da Engenharia do Conhecimento. Isotani e Bittencourt (2015) afirmam que as ontologias são usadas com diferentes propósitos, aplicadas em distintos domínios e reaproveitadas por inúmeras comunidades. A isso se somam as variações nas linguagens de formalização e nas metodologias de elaboração, com uma amplitude de atividades, a exemplo da análise conceitual e da modelagem de domínio.

Como mostra a Figura 16, da percepção de determinado fenômeno à elaboração de uma ontologia, há grandes perdas conceituais por causa das limitações das linguagens utilizadas, assim como ocorrem perdas conceituais da observação do mundo real até a sua conceitualização, em especial pelas limitações dos observadores na compreensão dos fenômenos, uma vez que:

O grande desafio na construção de ontologias é fazer com que haja um estreitamento e uma aproximação da ontologia com o mundo real. Quanto mais distante uma ontologia estiver do modelo que se espera, pior será a qualidade da ontologia (Isotani; Bittencourt, 2015).

Figura 16: Perda desde a percepção do mundo até a representação do conhecimento.



Fonte: Isotani e Bittencourt (2015).

Com o avanço das criações e aplicações de ontologias em áreas específicas do conhecimento, mais as oportunidades de inovação tecnológica da Web semântica, o Conselho Internacional de Arquivos tem dedicado esforços para o desenvolvimento do modelo conceitual *RiC* e da sua respectiva ontologia formalizada *RiC-O*.

O trajeto tecnológico na área de arquivos transcende os últimos dez anos de trabalho do ICA na elaboração do modelo *RiC*, como explicado pelos especialistas em descrição do EGAD Gretchen Gueguen, Vitor Manoel Marques da Fonseca, Daniel Pitti e Claire Sibille-de Grimouïard. Ao revisitar as tecnologias dominantes das últimas décadas, os membros do grupo frisam que, desde 1980, houve um deslocamento expressivo em direção à adoção de dois padrões tecnológicos: o *Structured Query Language (SQL)* e o *Standard Generalized Markup Language (SGML)*, este último precursor do *XML* (Gueguen *et al.*, 2013a). Eles salientam diferenças essenciais entre os padrões e evidenciam o dinamismo das práticas arquivísticas em resposta às novas tecnologias.

Exemplificadas pelo *SGML/XML*, as tecnologias de marcação são categorizadas como centradas em documentos, característica da *Web 1.0*. Como explicam os profissionais do ICA, elas foram primordialmente concebidas para documentos textuais tradicionais, tais como ensaios, artigos e livros. De maneira geral, tais documentos trazem características em comum: a presença de subpartes em número variável, a importância da sequência ou ordenação, uma

estrutura semirregular ou hierárquica, uma combinação diversificada e marcada de tipos de dados, além da multiplicidade de relações entre eles (Gueguen *et al.*, 2013).

Por outro lado, as tecnologias voltadas para bases de dados, a exemplo do *SQL*, são frequentemente descritas como orientadas a dados, distinguindo-se pela segmentação em campos interconectados, de maneira consistente, e por proporcionarem ampla versatilidade de configurações aos usuários do sistema, pois adaptam-se aos seus variados usos. Como detalhado pelos membros do ICA, as tecnologias em *SQL* apresentam campos em quantidade regular, em que a sequência não é de suma importância. Cada campo é preenchido com dados específicos, e campos inter-relacionados seguem uma hierarquia pré-definida e simples, além do conteúdo de cada campo obedecer a critérios rígidos quanto à sua forma e à sua estruturação, como elucidam os especialistas:

As duas tecnologias, base de dados e marcação, assumem diferentes abordagens na representação dos dados e cada uma delas é bem adequada para usos específicos. Os usuários, portanto, têm necessitado analisar seus dados para verificar suas características, ao mesmo tempo em que, tendo em mente como querem usá-los, decidem qual das duas tecnologias aplicar (Gueguen *et al.*, 2013a, p. 106).

Ao tratar do uso das duas tecnologias nos arquivos, não se observou uma adesão total e exclusiva a uma determinada tecnologia, não prevalecendo, portanto, uma em detrimento da outra. Em vez disto, como apontam os especialistas, evidenciou-se uma convergência entre ambas. As tecnologias baseadas em banco de dados, tal como o *SQL*, foram predominantemente adotadas em sistemas de registro das descrições e suas atualizações. Em contrapartida, as tecnologias fixadas em linguagens de marcação, a exemplo do *XML*, mostraram-se mais relevantes para a comunicação e transmissão de dados, seja entre máquinas ou entre máquinas e usuários finais (Gueguen *et al.*, 2013b).

Ao aderir às ferramentas da Web semântica e ao *Linked Open Data*, com a concepção do modelo conceitual *RiC*, o ICA recomenda, aos arquivos e aos arquivistas, um terceiro padrão para a representação de informações e dados em sistemas, baseado nas tecnologias de grafos. Diferente dos padrões anteriores, como o *XML*, alicerçado em estruturas em árvore ou hierárquicas, o *RDF* emprega grafos em configurações de teias, o que permite mais expressividade e inovação para representar a realidade, percebendo-a como uma extensa e dinâmica rede de indivíduos e objetos interligados, contextualizados no tempo e no espaço (Gueguen *et al.*, 2013b, p. 572). Assim, segundo os especialistas do ICA, o modelo *RiC* se

adapta às tecnologias semânticas, aspirando à criação de uma rede coesa de informações interconectadas, articuladas de maneira lógica e com clareza semântica nos arquivos.

Da análise do modelo *RiC* pelo ponto de vista tecnológico, entende-se que o ICA compartilha a ideia de que a Web semântica, em sua essência, contempla procedimentos e padrões que buscam transformar a *Web* em uma ampla rede semântica de declarações interligadas sobre o mundo e suas intrincadas relações. Gueguen *et al.* (2013b) constata a crescente tendência em diversos domínios da vida humana em adotar a perspectiva de dados abertos conectados. Como já observado, o *Linked Open Data* permite que as máquinas explorem a rede, propiciando a emergência de novos conhecimentos e padrões em que a informação coletiva transcende a mera junção de seus componentes individuais, além de promover a reutilização destes dados em sistemas distintos da sua origem.

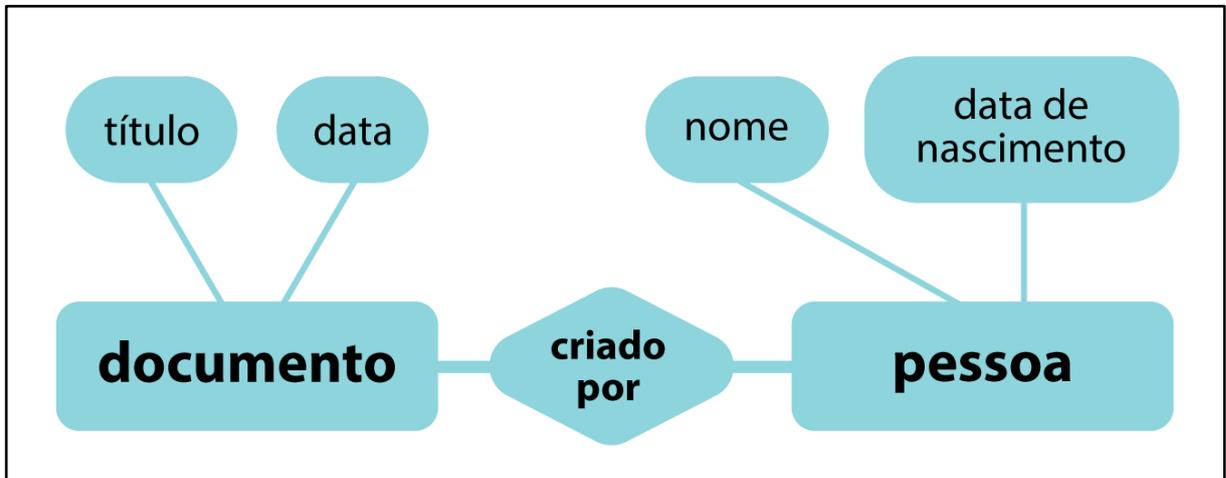
No âmbito das tecnologias semânticas, em teoria geral sobre modelos conceituais, sobressai o modelo Entidade-relacionamento, escolhido pelo Conselho Internacional de Arquivos como referência do modelo *RiC*. O surgimento do modelo ER ou MER, oriundo da Engenharia de *Software*, remonta ao artigo *The Entity Relationship Model: Toward a Unified View of Data*, de Peter Chen, publicado em 1976. O modelo é vastamente reconhecido na teoria e prática de criação de bancos de dados e introduz os conceitos fundamentais da tríade entidades, atributos e relacionamentos, como declaram os profissionais,

Com a emergência de bases de dados objeto-relacionais, a abordagem ER foi adaptada para refletir métodos. Com o surgimento das tecnologias semânticas baseadas em grafos, mais métodos de modelagens e ferramentas foram desenvolvidos para apoiá-los (Gueguen *et al.*, 2013a, p. 109).

As pesquisadoras Regina Maria Maia, da Universidade do Sul da Bahia (UFSB), e Lídia Alvarenga, da Universidade Federal de Minas Gerais (UFMG), oferecem uma interpretação detalhada sobre o modelo Entidade-relacionamento. Segundo as duas, o modelo MER conceitua o mundo como um agrupamento de objetos, denominados entidades, e as interações entre eles, conhecidas como relacionamentos. As entidades podem ainda exibir características ou propriedades chamadas atributos. O reconhecimento das entidades e dos seus relacionamentos leva à identificação de natureza específica dos objetos e, conforme Paulo Cougo, especialista em banco de dados (1997 *apud* Maia; Alvarenga, 2013, p. 6), inclui “as regras de existência dos objetos, associações possíveis e proibidas, além do número de participantes em cada associação”. Adicionalmente, Maia e Alvarenga (2013) esclarecem a simbologia utilizada no MER, onde as entidades são representadas por retângulos (por exemplo, Documento, Pessoa);

os atributos, por ovais (como título, data, nome, data de nascimento); e os relacionamentos, que conectam as entidades, por losangos, conforme ilustrado abaixo.

Figura 17: Modelo Entidade-relacionamento (ER) aplicado aos arquivos no modelo *RiC*.



Fonte: Adaptada do ICA (2023).

Em apresentação no Congresso Internacional de Arquivos na Coreia do Sul, em 2016, Daniel Pitti, Bill Stocking e Florence Clavaud, membros do grupo EGAD, encarregado da elaboração do *RiC*, apontaram que, para efetivar a representação do conhecimento em domínios específicos – arquivos, bibliotecas e museus –, tornam-se imprescindíveis ontologias próprias destes segmentos, a exemplo do próprio modelo *RiC* para os arquivos. Contudo, os profissionais acreditam que, para as ontologias atenderem aos propósitos da estruturação semântica dos dados e permitirem o seu processamento computacional nas instituições de arquivo – ou entre elas, bibliotecas e museus –, o obstáculo reside na transição das tecnologias consagradas. As tecnologias *XML* ou dos bancos de dados relacionais *SQL* – que não perdem totalmente a utilidade – devem migrar para *RDF*, *OWL*, bases de dados não-relacionais (NoSQL) e tecnologias em grafos, mais adequadas aos padrões da *Web 3.0* (Pitti *et al.*, 2016).

No caso dos arquivos, torna-se necessário que os profissionais tomem conhecimento do amplo ecossistema da Web semântica, no qual o modelo *RiC* e a sua respectiva ontologia *RiC-O* se inserem. O *RiC* é o dispositivo central na modelagem e no enriquecimento semântico dos dados de arquivos, atividade explicada por Claudiane Nazário e Célia Dias, pesquisadoras da Universidade Federal de Minas Gerais (UFMG) (*apud* Lira, 2021):

O enriquecimento semântico pode ser compreendido como um processo de atribuição de maior significado aos dados e metadados, através do uso da semântica atribuída por vocabulários pré-existentes, sinônimos e informações de proveniência. Esse processo tem por objetivo facilitar a compreensão, a integração e o processamento de dados por pessoas e máquinas (Lira, 2014 *apud* Nazário; Dias, 2021, p. 3).

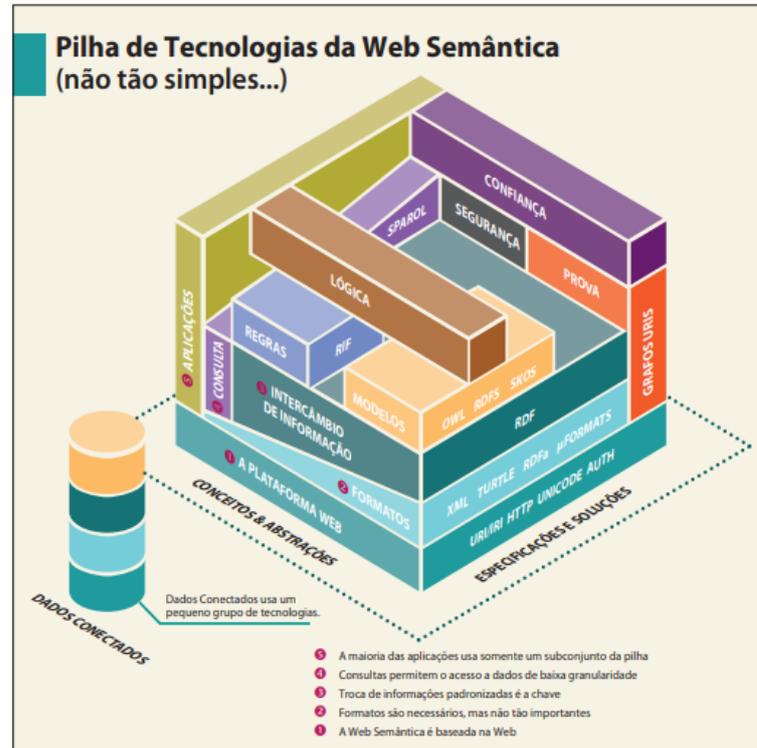
Apesar das vantagens da estruturação de dados e da sua semantização por ontologias e padrões tecnológicos, delineadas pela comunidade da Web semântica, existem questionamentos quanto à consolidação desta tecnologia. O diretor do sistema de informação do Instituto Nacional de Estatística e de Estudos Econômicos da França (Insee), Jean-Séverin Lair, argumenta sobre o aumento de custo na manutenção de várias máquinas para a manipulação dos grafos *RDF*, a exemplo do sistema SPAR, da Biblioteca Nacional da França. Nele, registrou-se a experiência de esperar horas, até mesmo dias, para obter respostas de consultas ao banco de dados completo, tendo que, posteriormente, se criar um banco de dados reduzido, restringindo diversos campos de metadados disponíveis. Na opinião do gestor, somam-se as limitações quanto ao tempo de sincronização dos servidores com os bancos de dados *NoSQL*, usados pela Web semântica, e a sua relação com os grandes volumes de dados (Lair, 2021).

Na sequência, analisando precisamente a aplicabilidade da Web semântica em arquivos, o profissional francês tece comentários críticos, sugerindo que a comunidade arquivística deveria fazer testes do modelo ou provas de conceito para avaliar a sua eficiência e que, talvez, as tecnologias da Web semântica serviriam melhor aos arquivos que possuem “repositórios de metadados de volume reduzido, facilmente gerenciáveis e que possam ser atualizados pelos humanos” (Lair, 2021, p. 29, tradução nossa). Lair (2021) relembra a obsessão dos profissionais de arquivo quando do lançamento do formato *XML*, usado para transmissão e intercâmbio de dados entre os arquivos, mas que hoje é considerado “catastrófico em matéria de eficiência e solidez”, por dificultar vários aspectos na gestão de documentos digitais e também por ter sido substituído pelo formato *JSON*. Assim, alertando que o formato *RDF* usado pela Web semântica possa ter vida útil breve como o *XML*, sugere que os arquivistas “fiquem mais centrados nas exigências e necessidades funcionais, nos metadados, na sua natureza e coleta”.

Portanto, dada a complexidade da Web semântica, como mostra a Figura 18, restam dúvidas quanto ao custo-benefício de sua implantação. De qualquer modo, com o crescente uso dos dados abertos conectados, o *W3C* publicou 35 boas práticas¹⁸, com o objetivo de estabelecer princípios de aplicação e uso de dados.

¹⁸ Disponível em: <https://www.w3.org/Translations/DWBP-pt-BR/>.

Figura 18: Pilha de tecnologias da Web semântica.



Fonte: Isotani e Bittencourt (2015).

Mesmo no momento inicial do modelo *Records in Contexts* e diante de tanta inovação tecnológica, o ICA recomenda aos arquivistas estruturarem os dados descritivos dos arquivos de acordo com as propostas da Web semântica e do *Linked Open Data*, pelo intermédio do modelo *RiC*. Apesar da sofisticação das tecnologias semânticas, elas são adotadas em bibliotecas, museus, setores comerciais, industriais e científicos, confirmando as vantagens da semantização, da expressividade e da flexibilidade no manejo dos dados na *Web* (Pitti *et al.*, 2016).

2.2 Revisão teórica e prática de descrição arquivística

Esta subseção trata de aspectos relativos à revisão teórica e à prática da descrição arquivística.

2.2.1 Revisão das normas internacionais de descrição arquivística

As inovações tecnológicas da Web semântica propostas no *RiC* somam-se às revisões das normas internacionais de descrição consolidadas pelo ICA, a partir de 2008.

Em retrospecto, as normas internacionais de descrição de arquivos, publicadas desde os anos de 1990 no quadro do ICA, têm a comunidade arquivística do Canadá como uma das principais referências. O arquivista Terry Cook do Arquivo Nacional do Canadá explorou a experiência do país na padronização da descrição, elucidando que, anteriormente à publicação da norma *Rules for Archival Description (RAD)*, em 1990, os arquivistas dedicaram-se ao entendimento sobre as inúmeras concepções de fundo, oriundas tanto da tradição francesa quanto da anglo-saxônica. Cook (2017) explica que, dado o propósito do comitê nacional em estabelecer um padrão de descrição em arquivos, era imperativo que conceitos basilares e práticas associadas – como a definição de fundo – fossem compreendidos de forma precisa e aceitos consensualmente pela comunidade arquivística, pelo menos no âmbito canadense. Com forte influência da *RAD* de seu país, inicia-se a separação de elementos de descrição dos arquivos, tal como a descrição de fundos e documentos da descrição de autoridades produtoras.

Conforme Gueguen *et al.* (2013b), a separação dos elementos de descrição em arquivos foi introduzida na formulação de uma norma internacional de descrição arquivística por membros do ICA. A separação se deu com a proposta de criação de pontos de acesso relacionados às autoridades vinculadas aos documentos. A princípio, a prática de separação das descrições das autoridades dos demais elementos de descrição foi também influenciada pela adoção de sistemas computacionais da década de 1980. Inspirado por tais perspectivas e almejando economia e flexibilidade na descrição dos arquivos, bem como novas maneiras de explorá-la em sistemas de acesso, o ICA adotou a segmentação dos componentes na elaboração das normas internacionais.

De 1994 a 2008, o ICA colocou à disposição quatro normas internacionais para o uso da comunidade arquivística: Norma Geral Internacional de Descrição Arquivística (ISAD-G); Norma Internacional de Registro de Autoridade Arquivística para Entidades Coletivas, Pessoas e Famílias (ISAAR-CPF); Norma Internacional para Descrição de Funções (ISDF); Norma Internacional para Descrição de Instituições com Acervos Arquivísticos (ISDIAH).

Quadro 6: Cronologia do desenvolvimento das normas internacionais de descrição arquivística.

Norma	Edição	Data de elaboração	Data de publicação
Declaração de princípios		(1988) 1989-1992	1992
ISAD	1 ^a	1990-1993	1994
ISAAR	1 ^a	1993-1995	1996
ISAD	2 ^a	1996-2000	1999
ISAAR	2 ^a	2000-2004	2004
ISDF	1 ^a	2005-2007	2007
ISDIAH	1 ^a	2005-2008	2008

Fonte: Guegen *et al.* (2013).

Segundo o ICA, a ISAD(G) abrange os princípios fundamentais da descrição arquivística, como as regras de descrição multinível. Ela tem a finalidade de padronizar a descrição de fundos, grupos, séries, processos, dossiês e documentos, ou seja, de determinar os elementos para descrever os arquivos propriamente ditos, em termos de contexto e de conteúdo. A ISAAR(CPF) traz diretrizes para descrever contextualmente as autoridades relacionadas aos arquivos, sejam instituições, famílias ou pessoas. A ISDF complementa os contextos não explorados na ISAAR(CPF) e traz diretrizes para a descrição de funções e de atividades que geram os documentos de arquivo. Por fim, a ISDIAH normatiza a descrição das instituições custodiadoras de acervos arquivísticos. O ICA sempre esperou que as normas fossem aplicadas de forma relacionada, posto que “são conectadas como um compêndio de descrição” (ICA, 2012, p. 5).

Em 2008, no XVI Congresso Internacional de Arquivos, realizado em Kuala Lumpur, Malásia, o Comitê de Boas Práticas e Normas (CPBS) do ICA – fundamentado em relatos de experiências de sua aplicação na comunidade arquivística internacional – deu início à revisão das normas arquivísticas. Conforme relatado pelo comitê (2012), embora o resultado dessa revisão não estivesse claramente definido na época, recomendou-se, para o congresso subsequente, a elaboração de um relatório-compêndio, abrangendo as quatro normas. Tal

relatório deveria ter objetivos específicos, incluindo fornecer explicações sobre as inter-relações das normas; melhorar a consistência e a clareza das descrições e dos relacionamentos dos seus elementos; desenvolver sistemas integrados de descrição entre instituições arquivísticas; aprimorar a estrutura, o conteúdo, o uso e os formatos das normas; e promover a comunicação entre os profissionais da área. No entanto, durante o Congresso de 2008, não foi confirmado se o relatório-compêndio resultaria na fusão das quatro normas em um único documento.

Em 2012, durante o XVII Congresso Internacional de Arquivos, realizado em Brisbane, Austrália, o CPBS divulgou o relatório propondo um capítulo específico sobre as inter-relações dos elementos de descrição das quatro normas. O relatório, que almejava a harmonização e a padronização delas, sugeriu, inicialmente, alterações em suas estruturas. Propôs-se a criação de um prefácio único para todas as normas, e as informações sobre a elaboração de cada norma seriam detalhadas separadamente em Notas Históricas. Ademais, recomendou-se a padronização da Introdução das normas ISAD(G) e ISDIAH, além do Escopo e objetivo das normas ISAAR e ISDF, sugerindo uma Introdução Geral para as quatro normas e uma seção de Escopo e Objetivo para cada uma delas. No relatório, a segunda proposta de aprimoramento procurou evitar redundâncias, sugerindo a unificação da seção Normas e Diretrizes relacionadas. Esta seção, que enumera as normas ISO, mencionadas nas normas ISAAR(CPF), ISDF e ISDIAH, seria agrupada para maior clareza e eficiência.

A terceira proposta consistiu na criação de um glossário comum para as quatro normas, considerando a existência de termos com diferentes aplicações e definições nos glossários individuais de cada norma, o que resultava em situações de ambiguidade e inconsistência. O quarto aspecto tratado na revisão propõe a inclusão de uma seção denominada Estrutura e Uso da norma na ISAD(G), seguindo o padrão visto nas demais normas. Essa sugestão também aventou integrar as seções Descrição Multinível e Regras para Descrição Multinível (ICA, 2012).

A quinta alteração visa à padronização dos títulos dos elementos de descrição entre as diferentes normas. Por exemplo, os termos tipo de entidade da ISAAR(CPF), tipo da ISDF e tipo de instituição custodiadora da ISDIAH foram unificados sob a denominação Tipo. Outro exemplo é a harmonização do elemento Forma de nome autorizada da ISAAR(CPF) e Forma autorizada de nome da ISDF e da ISDIAH para Forma autorizada de nome. A sexta recomendação foi a criação de uma área de controle comum as quatro normas, com os mesmos elementos de descrição, a saber: identificador da descrição; identificador da instituição; regras

e/ou convenções; status; datas da criação, revisão e eliminação; nível de detalhamento; idioma e sistema(s) de escrita; fontes e notas de manutenção (ICA, 2012).

A última e, possivelmente, a mais importante das recomendações, considerada um elemento central para o próximo mandato do grupo de especialistas do ICA na elaboração do modelo *RiC*, é a harmonização dos registros de relacionamentos. Esta seção está na ISAAR(CPF), ISDF e ISDIAH, com exceção da ISAD(G). No caso da ISAAR(CPF), especifica-se o registro das relações entre entidades coletivas, pessoas e famílias, e suas respectivas fontes de arquivo. Na ISDF, concentra-se nas relações entre funções e descrições arquivísticas, autoridades e demais fontes. Por último, a ISDIAH realça o registro das relações entre entidades custodiadoras e as descrições de fundos e documentos. O CBPS redigiu um esboço de um capítulo comum sobre o registro de relacionamentos entre as entidades da descrição arquivística, incluindo fundos, documentos, autoridades, funções e entidades custodiadoras. Este capítulo contempla três elementos de descrição: o identificador de entidades e de fontes relacionadas, a natureza do relacionamento e as datas do relacionamento. Além disso, foram criados diagramas para representar visualmente os relacionamentos possíveis entre as Normas do Conselho Internacional de Arquivos (ICA, 2012).

A análise do relatório progressivo de revisão e harmonização das Normas de Descrição do ICA, elaborado em 2012, revela outros aspectos. Na proposta da introdução geral, ao estabelecer os objetivos e princípios da descrição arquivística, o ICA reafirma a relevância dos arquivistas descreverem tanto o contexto quanto o conteúdo dos arquivos e documentos. Adicionalmente, ressalta dois elementos-chaves: a noção de entidades e os relacionamentos entre elas. Pondera a importância da descrição arquivística se concentrar nos documentos de arquivo, mas considerar as demais entidades pertinentes para o contexto arquivístico: as instituições, pessoas e famílias que produzem os arquivos; as funções desempenhadas e atividades executadas durante a produção dos documentos; e as instituições responsáveis pela custódia dos arquivos. No relatório, o termo entidade é empregado em 34 ocasiões, alternando-se com o termo elementos de descrição, o que sinaliza o início da ideia de se elaborar um modelo conceitual para os arquivos.

A ênfase nos relacionamentos entre as entidades sublinha o valor dos sistemas informatizados de descrição de arquivos, nos quais as quatro normas devem ser “criadas e mantidas de forma independente, mas conectadas de forma visível ao usuário” (ICA, 2012, p. 5). Ao lado disso, o relatório indica as vantagens da descrição arquivística padronizada, como o aumento da consistência da descrição e a possibilidade de troca de informações entre sistemas,

assim como a participação em portais corporativos entre arquivos ou instituições do patrimônio documental e cultural (ICA, 2012, p. 5).

As prioridades para o mandato do Comitê de 2012 a 2016 são delineadas no documento, tendo como objetivo principal o desenvolvimento de um modelo conceitual ou esquema de entidades que beneficiem os arquivos. Conclui-se que é a única maneira de eliminar a redundância do conjunto de normas de descrição (ICA, 2012). O modelo conceitual deveria abranger os diferentes tipos de entidades e documentar os relacionamentos entre elas. E, também, buscaria resolver a confusão entre identidade e designação, esclarecer conceitos fundamentais como fundo, atribuição, função, nível de descrição e expandir os conceitos para além das normas do ICA, facilitando o intercâmbio entre arquivos, bibliotecas e museus. Por fim, o grupo manifesta elogios ao modelo do sistema *Access to Memory (ICA-AtoM)* e à entidade denominada evento, que se aproxima do modelo australiano de metadados (ICA, 2012, p. 15).

Na avaliação retrospectiva feita sobre a aplicabilidade dos padrões internacionais de descrição, há concordância entre os especialistas do ICA quanto à predominância da ISAD(G) em relação às demais normas. Tal fato reforça a tradição na descrição de fundos e documentos em arquivos. Em consequência, as normas ISAAR(CPF) (autoridades/produtores), ISDF (funções/atividades) e ISDIAH (entidade custodiadora) – que têm seus elementos derivados da ISAD(G), e que deveriam ser exploradas autonomamente em sistemas específicos de forma interconectada, com vistas a novas compreensões e usos dos documentos de arquivo – não atingiram tal propósito (Gueguen *et al.*, 2013a, p. 101).

Um dos principais fatores que explicam a não assimilação e o uso das normas internacionais pelas instituições arquivísticas, segundo os profissionais da EGAD (ICA, 2023), é a carência de manuais práticos com diretrizes claras sobre como as normas poderiam ser conjuntamente empregadas. A esse aspecto, adicionam-se as limitações funcionais dos sistemas e suas ferramentas, fazendo com que os componentes do sistema de descrição seguissem direções distintas.

A centralidade da ISAD(G) é um dos aspectos mais criticados pelo grupo EGAD no processo revisional da descrição arquivística. Entende-se que a prioridade dada à descrição de fundos e documentos, em detrimento de outras entidades que poderiam ser descritas, é prejudicial no complexo contexto atual – o universo digital – em que os documentos são produzidos, usados e preservados. As pessoas, entidades coletivas, funções, atividades e entidades custodiadoras relacionadas aos arquivos, apesar de possuírem normas próprias de

descrição – ISAAR(CPF), ISDF e ISDIAH –, não tiveram suas descrições elaboradas na maioria das instituições de arquivo, tornando-se, quando muito, apenas pontos de acesso para recuperar fundos e documentos do acervo (ICA, 2023). Somam-se, ainda, as entidades – localidade, assunto e data –, que são também usadas como pontos de acesso, não possuem normas específicas nem metadados próprios, e sua descrição não é regularmente feita pelos arquivistas. Na visão do ICA (2023), a intenção é que, por intermédio do modelo conceitual *RiC*, os arquivistas descrevam também estes elementos, considerados importantes para a compreensão da ampla rede de entidades em que os arquivos se relacionam.

Em vez de continuar tentando revisar a ISAD(G) e suas normas complementares, uma a uma, para tratar dessa questão, o ICA instruiu o EGAD a desenvolver um esquema abrangente para a descrição arquivística baseado em um modelo conceitual que reconheça as realidades relacionais e contextuais dos arquivos e que considere a entidade documento como sendo apenas uma de uma série de entidades e relações que devem ser descritas para que os sistemas de controle arquivístico reflitam as realidades das diversas formas como os arquivos são feitos, mantidos e utilizados (ICA, 2023, p. 9, tradução nossa).

Nas justificativas da adoção para o modelo conceitual *RiC*, constatam-se inúmeras críticas adicionais às normas de descrição. Devido às normas do ICA terem sofrido forte influência em termos práticos e tecnológicos na época em que foram criadas – anos 1990 –, uma das críticas decorre do fato de o registro de entrada da descrição elaborada por arquivistas ser idêntica, na maioria dos casos, à versão de saída da descrição à disposição dos usuários, mesmo que o padrão tenha sido rejeitado explicitamente, como consta na introdução da ISAD(G), “a norma não define formatos de saída ou modos nos quais esses elementos são apresentados, por exemplo em inventários, catálogos, listas” (ICA, 2004, p. 11). Como resultado, caracterizada por sua natureza linear, a recuperação das descrições por usuários, em inúmeros sistemas de busca desenvolvidos com base nas normas, se assemelha às buscas nos tradicionais instrumentos de pesquisa impressos, lidos página a página, com alguns recursos adicionais, como índices, remissivas ou pontos de acesso.

Não de outro modo, tal como sugere o ICA (2023), a relação entrada-saída de dados é uma das principais inovações do *RiC*. Trata-se de um modelo referencial apenas para a entrada de dados nos sistemas de registro – dados separados e conectados em ampla rede de relações contextuais – com alta aderência às tecnologias atuais. Quanto à saída dos dados, à sua apresentação e à manipulação pelos usuários, a proposta do *RiC* é que não haja formato definido, sendo mais aberto e com possibilidades infinitas de uso conforme o interesse dos

pesquisadores. Coincidindo com a proposta original da ISAD(G), que, no entanto, não foi aplicada de forma adequada pela comunidade arquivística.

Acrescentam-se questões relativas à noção das unidades de descrição, já que os elementos de descrição eram os mesmos, independentemente do nível a ser descrito: fundo, grupo, série, dossiê ou item. A ISAD(G) contempla 26 elementos de descrição, enquanto a Nobrade, 28 elementos. Na norma brasileira, destaca-se a obrigatoriedade de sete deles: código de referência, título, data(s), nível de descrição, dimensão e suporte, nome(s) do(s) produtor(es) e condições de acesso, este último aplicável aos níveis da entidade custodiadora e do fundo. Importante ressaltar que, embora a maior parte dos elementos de descrição não ser prescritivo para cada nível, isto é, poder ser usados caso a caso, não havia individualização dos metadados a serem descritos por nível de descrição. Assim, na opinião dos especialistas do EGAD, supõe-se certa equivalência pouco real entre fundos e itens, por exemplo. No modelo *RiC*, apesar de vários atributos compartilhados, há metadados individualizados por entidades, posto que “tratá-los cada um como algo distinto, produzidos em momentos distintos e com propósitos diferentes, possibilita uma descrição inequívoca e clara” (ICA, 2023, p. 10, tradução nossa).

Um dos aspectos mais inovadores do modelo *RiC* é a descrição dos arquivos para além da lógica da descrição multinível em direção à descrição multidimensional. A descrição multinível é consagrada na área de arquivos e teve a sua divulgação ampliada pela ISAD(G) e, no Brasil, pela Norma Brasileira de Descrição Arquivística (NOBRADE), elaborada pelo Conselho Nacional de Arquivos (CONARQ), com base na primeira.

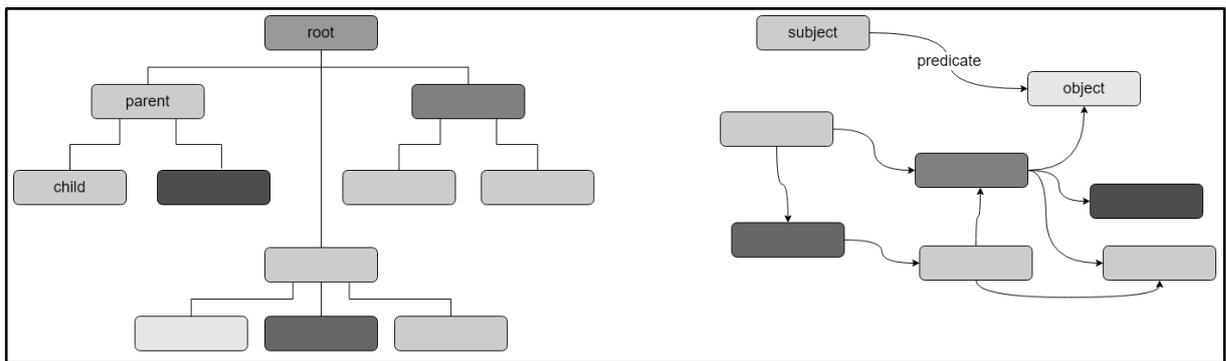
O conceito de descrição multinível consiste em descrever o fundo no geral, posteriormente, cada parte, consideradas as divisões hierárquicas nos quadros de arranjo da documentação. Conforme consta na ISAD(G), para a confecção da descrição multinível, existem regras básicas a serem cumpridas pelos arquivistas: descrição do geral para o particular, que corresponde a “representar o contexto e a estrutura hierárquica do fundo e suas partes componentes”; a informação relevante para o nível de descrição, que consiste em “representar com rigor o contexto e o conteúdo da unidade de descrição”; a relação entre as descrições, que “torna explícita a posição da unidade da descrição na hierarquia”; e, por fim, a não repetição da informação, que objetiva “evitar a redundância de informação em descrições hierarquicamente relacionadas” (ICA, 2004, p. 17).

Segundo o ICA (2023), a descrição multidimensional proposta no modelo *RiC* aproxima-se mais de uma representação de dados dos arquivos em redes ou grafos, conforme se observa na Figura 19, não importando a hierarquia da descrição multinível, o que torna

possível uma descrição “mais factível para documentos com proveniência complexa e diversificada dentro de um fundo”, e complementa:

Na perspectiva da descrição multidimensional, os documentos e conjunto(s) de documentos, suas inter-relações, suas inter-relações com pessoas, grupos e atividades, bem como as inter-relações destes elementos são representados como uma rede na qual um fundo específico pode ser identificado. O contexto imediato de cada fundo é estabelecido, embora as suas fronteiras sejam adaptáveis, pois ele existe dentro de uma rede de registros inter-relacionados e de conjunto(s) de documentos, indivíduos, grupos e atividades (ICA, 2023, p. 11, tradução nossa).

Figura 19: Representação comparativa de dados em estrutura hierárquica, como *XML* (à esquerda) e estrutura em triplas, como grafos (à direita).



Fonte: ICA (2023).

Todas as críticas às normas de descrição do ICA e aos seus conceitos relacionados somam-se às discussões mais amplas da teoria arquivística. Dessa forma, ao elaborar o modelo *RiC*, o ICA não só observou os ganhos, as perdas e o aperfeiçoamento das normas de descrição, mas considerou as discussões teóricas sobre o princípio da proveniência que vem ocorrendo há décadas na comunidade arquivística, as quais retomaremos pontos centrais para melhor compreensão das propostas do modelo *RiC*.

2.2.2 Princípio da proveniência ampliado e princípio do respeito à ordem original

Em sua obra, Cook (2017) discorre sobre múltiplos aspectos inerentes ao conceito de fundo, dos quais, destaca-se a sua natureza intrínseca. E como tal noção visa a garantir o caráter de prova dos arquivos, desde a sua origem, configura-se ainda como o princípio que distingue os arquivos de outras instâncias do patrimônio documental, tais como bibliotecas, museus e centros documentais. O autor também ressalta a origem francesa do princípio de respeito aos

fundos, as suas dimensões internas e externas, a sua correlação com o princípio do respeito à ordem original e com a proveniência germânica, além da assimilação desse conceito no panorama anglo-saxão por meio da noção de *records group*, um tópico de notável controvérsia, sobre o qual ele defende o abandono do referido conceito.

Das inúmeras considerações teóricas salientadas por Cook (2017), ganham destaque as perspectivas custodial e pós-custodial do conceito de fundo, bem como diferentes visões sobre a descrição dos arquivos. Por um lado, pratica-se na comunidade arquivística a descrição com base no entendimento do fundo como uma categoria conceitual, delineada nas classes mais gerais do quadro de arranjo, centrado nos produtores e respectivos contextos jurídico-administrativos. Em contrapartida, há a prática de descrição dos níveis mais concretos – processos, dossiês e itens documentais – onde são localizados os documentos em sua essência (Cook, 2017, p. 61). Tal dicotomia é ratificada pelos especialistas do ICA:

O modelo descritivo dominante no período era o de usar uma única descrição para descrever um fundo arquivístico, começando com a descrição mais geral e, em crescente especificidade, as partes do fundo, as partes das partes, e assim por diante. Todos os atributos dos documentos em um fundo, e o contexto da produção e uso dos documentos, eram combinados em uma descrição, com os diferentes elementos descritivos entrelaçados para formar uma descrição compreensiva e completa. A descrição cobria, assim, os documentos como tais, bem como aquilo que com o tempo viemos a entender como o contexto do documento: o produtor, funções e atividades, e entidade custodiadora. (Gueguen *et al.*, 2013a, p. 103).

Para Cook (2017), a concepção clássica da área, que interpreta o fundo como um conjunto físico de documentos oriundos de um único produtor, é categorizada como custodial ou curatorial.

Na perspectiva pós-custodial, a compreensão de fundo assume uma conotação mais abstrata, visto que reconhece que os documentos contemporâneos são caracterizados pela produção e acumulação coletiva, envolvendo múltiplos produtores. Tal dinâmica torna inexequível a associação exclusiva entre uma série documental e um único produtor, como frequentemente observado em instituições do século XIX. Na proposta de Cook (2017), embora o conceito ainda não esteja plenamente consolidado na comunidade arquivística, diversos fatores justificam a natureza de multiprodução da documentação, sobretudo a atual dinâmica governamental, a intrincada burocracia, a presença de estruturas produtivas concomitantes e as atividades de grupos de trabalho. A estas considerações, adicionam-se as implicações das inovações tecnológicas emergentes nas últimas décadas do século XX, como na produção de documentos em sistemas eletrônicos ou digitais. As séries, os processos, os dossiês e os itens

documentais muitas vezes possuem, simultaneamente, múltiplos produtores, resultando em documentos com multiproveniência. Assim, Cook (2017) se põe a favor de uma libertação da perspectiva custodial no âmbito arquivístico (Cook, 2017, p. 13).

Outro pesquisador, o arquivista australiano Peter Scott, um dos mais proeminentes teóricos a tratar da natureza da multiproveniência das séries documentais, fundamentado em sua vasta experiência, aparece como um crítico contundente da perspectiva custodial do fundo. Ele postula a série documental como o núcleo da descrição arquivística. Nesta ótica, Scott (1980 *apud* Cook, 2017, p. 23) sustenta que uma série pode ser gerada por diversos produtores, seja de forma sucessiva ou concomitante.

Para exemplificar os conceitos no âmbito dos arquivos legislativos, ao tomarmos a tramitação ordinária de um projeto de lei que circula por diversas comissões em uma das casas legislativas (Câmara dos Deputados ou Senado Federal) e, depois, encaminhado à casa revisora, para a análise de outras comissões. A depender da interpretação dos arquivistas, aplica-se a visão mais maximalista à noção de fundo – os produtores da série projetos de lei seriam a Câmara dos Deputados e o Senado Federal, e, portanto, as comissões seriam apenas autoras. Já a visão minimalista, todas as comissões, por onde o projeto de lei tramitou, podem ser reconhecidas como as suas entidades produtoras. Neste último caso, a série de projetos de lei estaria descrita e associada a todos os fundos dos respectivos produtores (inclusive as comissões), sendo representada em seus quadros de arranjo e descrita nos instrumentos de pesquisa pertinentes desses fundos. Em tal perspectiva, teóricos como Peter Scott defendem que os procedimentos de arranjo e descrição devem possuir uma maleabilidade superior àquela preconizada pela visão custodial dos arquivos, na qual uma série estaria vinculada a um único produtor, inserida em um único quadro de arranjo e registrada exclusivamente em um instrumento de pesquisa, a exemplo do inventário de um determinado fundo, como explicado por Cook (2017):

À medida que empresas e órgãos públicos nessas circunstâncias adotam novos modelos de informação baseados em planejamentos de dados corporativos e administração de recursos de dados, a ideia de um documento pertencente fisicamente a um lugar ou até mesmo a um sistema está desmoronando frente a novos paradigmas conceituais, em que a “produção” é um processo fluido de manuseio de informação de várias fontes em uma enorme quantidade de formas, ao invés de algo que conduz a um produto estático, fixo e físico. Para os arquivistas, isso sinaliza que a era custodial está dando passagem à era pós-custodial, quando a curadoria de objetos físicos definirá muito menos a profissão que uma compreensão das inter-relações conceituais entre estruturas produtoras, suas funções vitais e os documentos então resultantes (Cook, 2017, p. 47).

Na prática arquivística pós-custodial, Peter Scott (1980 *apud* Cook, 2017, p. 33) evidencia que um dos principais desafios aos arquivistas reside em determinar a que produtor ou fundo uma dada série pertence. Como solução, Scott sugere o uso de remissivas nos instrumentos de pesquisa e a elaboração de diversas listagens específicas que estabeleçam referências cruzadas entre os produtores e as séries documentais inter-relacionadas, a exemplo das listas de controle de autoridades. De acordo com Cook (2017), o ponto de vista de Peter Scott, que contempla a concepção de “múltiplas entradas e a distinção entre as descrições dos contextos administrativos e dos documentos físicos, constitui-se em uma inovação verdadeiramente revolucionária” (Cook, 2017, p. 54).

Nas discussões da comunidade arquivística, durante a elaboração da primeira versão de consulta do modelo *RiC*, entre 2012 e 2016, os especialistas do ICA não apenas levantaram críticas em relação às normas, mas analisaram conceitos mais abrangentes que têm impacto direto na descrição. Nelas, questionam-se os entendimentos tradicionais dos princípios da proveniência e do respeito à ordem original.

Concernente às críticas lançadas pelo ICA, encontram-se aquelas à visão clássica e restrita do princípio da proveniência¹⁹, prejudicial à descrição dos arquivos, na perspectiva do EGAD. Mesmo que a proposta inicial da ISAAR(CPF) incluísse o registro de inúmeras pessoas e instituições relacionadas aos documentos, o modelo *RiC* afirma que, na prática, a descrição atual concentrou-se preponderantemente em registrar e preservar a relação dos documentos com o produtor ou acumulador dos fundos, em detrimento de outras pessoas, instituições e contextos relacionados aos arquivos, em seus momentos de produção, acumulação e uso. Dessa forma, o modelo *RiC* sugere a visão do que seria o princípio da proveniência ampliado, direcionando a descrição a outros contextos negligenciados, nos quais os arquivos se relacionam, além da mera descrição da proveniência ou da acumulação. O modelo *RiC* ambiciona estabelecer as relações destes documentos na intrincada rede de contextos, interações e interdependências, de forma a espelhar mais eficientemente a complexidade social e dos registros documentais (ICA, 2023). O Conselho Internacional argumenta que:

¹⁹ O princípio da proveniência, também chamado de princípio do respeito aos fundos, segundo o Dicionário Brasileiro de Terminologia Arquivística, Dibrate (Arquivo Nacional, 2005): “princípio básico da arquivologia segundo o qual o arquivo produzido por uma entidade coletiva, pessoa ou família não deve ser misturado aos de outras entidades produtoras”.

[...] todas estas observações levam à conclusão de que a origem dos documentos é muito mais complexa do que o entendimento de fundo documental há muito estabelecido. Os documentos em um fundo não existem apenas em relação a ele, nem o fundo documental existe isoladamente. Tanto o fundo quanto os documentos nele contidos existem em contextos interconectados do passado, do presente e do futuro (ICA, 2023, p. 7, tradução nossa).

O segundo aspecto relacionado às novas propostas teóricas e práticas de descrição relaciona-se com o princípio do respeito à ordem original²⁰, que, em termos gerais, objetiva a manutenção da ordem originária dos arquivos para fins de preservação do inter-relacionamento dos documentos no fundo documental.

De acordo com o ICA (2023), o princípio sempre foi controverso na comunidade arquivística, tanto no entendimento de original quanto de ordem. Em relação à ordem original, há inúmeras interrogações, como a verificação da ordem original dos arquivos numa realidade em que os documentos são acumulados de modo dinâmico, fluido e mutável, ocasionando, na maioria das vezes, reordenações do conjunto documental nestas fases. Ademais, a qual ordem o princípio se refere: a ordem intelectual entre os documentos ou a disposição física original? Somam-se, ainda, os muitos casos de documentos que entram nos arquivos sem uma ordem tangível – nem física, nem intelectual –, ocasião em que arquivistas refazem ou estabelecem a ordenação.

Observa-se que, na perspectiva do ICA, a concepção ampliada do princípio da proveniência não descarta, em momento nenhum, a conexão primordial dos documentos com os respectivos fundos de origem. Tal vínculo representa uma garantia fundamental para a preservação do contexto original de produção e acumulação dos documentos, sendo, de resto, um princípio identitário dos arquivos.

No entanto, adota-se uma visão mais inclusiva, em que se incorporam outros contextos que não são tradicionalmente incluídos nos sistemas de registro e de descrição de arquivos. Conforme o entendimento do autor desta pesquisa sobre a descrição de outros contextos segundo a proposta do modelo *RiC*, sugere-se que, para além do contexto de origem, descreva-se o contexto de uso dos documentos, a exemplo das interações de pesquisadores com os documentos de arquivo, que, muitas vezes, são registradas em bancos de dados específicos ou em fichas de pesquisa. Tais registros contêm uma variedade de dados relacionados às pesquisas,

²⁰ Princípio do respeito à ordem original: princípio segundo o qual o arquivo deveria conservar o arranjo dado pela entidade coletiva, pessoa ou família que o produziu (Arquivo Nacional, 2005).

incluindo, sobretudo, o nome do pesquisador, o nome da instituição de origem, o tema da pesquisa e os documentos selecionados. É importante notar que, depois da conclusão da pesquisa, quando os documentos selecionados são utilizados em teses, dissertações, artigos científicos, reportagens, processos judiciais, geralmente não existe a prática de registro unificado no sistema de descrição do acervo. Vale a pena mencionar que o elemento nota de publicação desempenha esta função na ISAD(G). No entanto, deve-se considerar que, além da dificuldade prática, os sistemas de descrição que contém os metadados sobre documentos e arquivos, estão reiteradamente isolados dentro da instituição arquivística, seguindo a lógica de silos de dados.

A mesma observação pode ser aplicada aos dados relacionados às ações de preservação do acervo. Em alguns casos, quando documentos analógicos passam por intervenções nos laboratórios de restauração, higienização, desinfestação ou reacondicionamento, as informações geradas por estes procedimentos – embora mantenham referências de controle sobre os documentos – são mantidas separadas em bases de dados e sistemas específicos, também seguindo a lógica de silos de dados nos arquivos.

O tom crítico é reforçado por representantes do ICA, salientando que a descrição arquivística é frequentemente percebida como autocontida e desvinculada do contexto externo dos arquivos (Pitti *et al.*, 2016). Diante deste quadro, o modelo *RiC* sugere a transição da descrição hierarquizada multinível atual, nos moldes das normas internacionais, para um modelo conceitual de dados, que se baseia em descrição multidimensional, considerada mais versátil e expressiva.

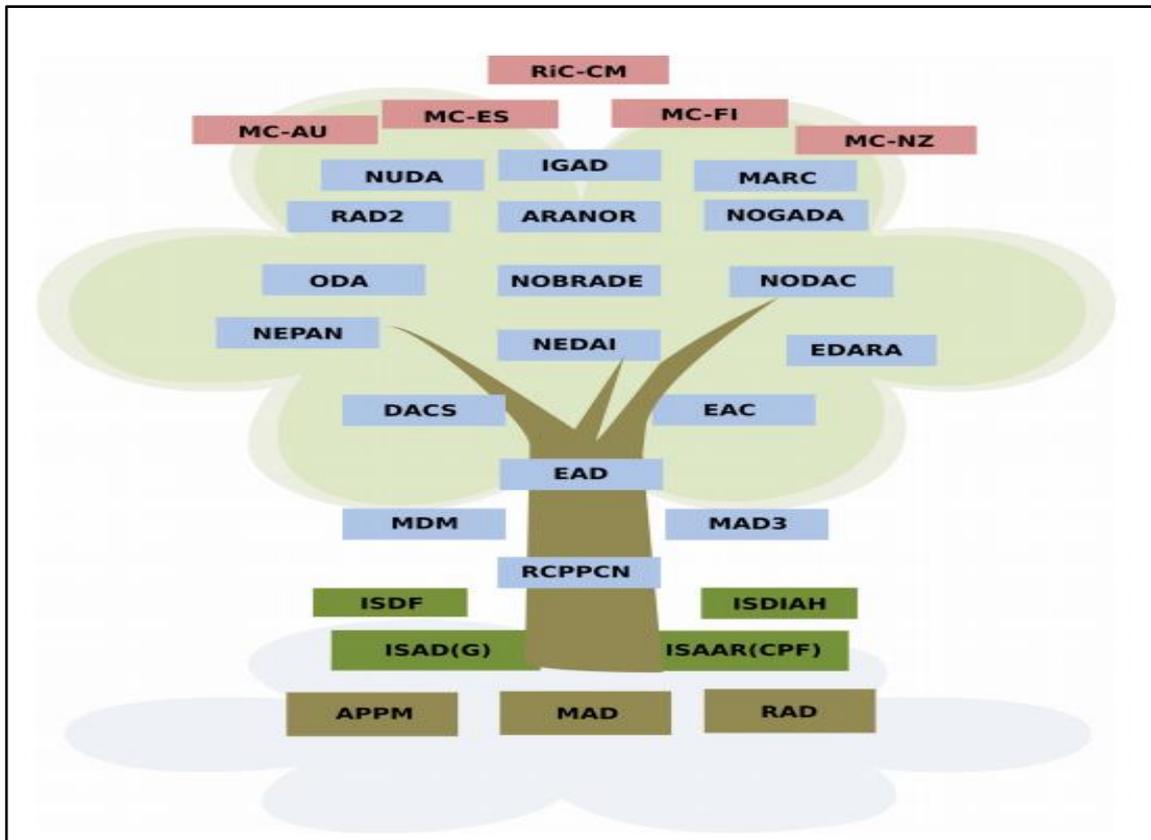
Assim, o modelo *RiC* sugere uma nova compreensão de conceitos arquivísticos como o princípio da proveniência (ampliado) e a integração das quatro normas vigentes, que não ocorreu na prática por meio dos sistemas. Este objetivo é anterior à concepção das normas internacionais, quando a teoria em questão era revisitada à luz das emergentes tecnologias e da consolidada perspectiva pós-custodial dos arquivos. Conforme exemplifica Cook (2017),

Usando automação para combinar ou relacioná-las, qualquer combinação de produtores e séries é possível: múltiplos produtores a uma única série, múltiplas séries de um único produtor, múltiplos produtores com séries múltiplas, mas relacionadas, séries com séries, e produtores com produtores, seja para cima ou para baixo na tradicional cadeia mono-hierárquica vertical, ou para os lados, nas novas estruturas organizacionais horizontais (Cook, 2017, p. 55).

A análise crítica da teoria e da prática em descrição arquivística, particularmente o resultado do uso das normas internacionais de descrição do Conselho Internacional de

Arquivos, e o questionamento da visão tradicional do princípio da proveniência destacam-se como os pilares de renovação proposta no *RiC*. Por fim, o modelo junta-se a uma série de padrões na área de descrição de arquivos, de acordo com a Figura 20.

Figura 20: Evolução das normas e modelos conceituais em descrição arquivística.



Fonte: Llanes-Padrón e Pastor-Sánchez (2017).

2.2.3 Modelos conceituais de arquivos, bibliotecas e museus em descrição arquivística

A ideia do acesso integrado aos dados compartilhados entre instituições de patrimônio cultural e documental representa um dos alicerces do modelo *RiC*. Pitti *et al.* (2016) esclarecem que, embora o pensamento de uma comunidade integrada entre arquivos, bibliotecas e museus tenha surgido nos meados do século XIX, pautada na noção de patrimônio, essa interconexão ganhou impulso nos tempos atuais. O renovado interesse deve-se, em grande parte, às inovações tecnológicas disponíveis para as instituições, destacando-se a Web semântica. Os especialistas do ICA argumentam ser possível identificar conceitos de descrição arquivística que se assemelham aos empregados nas comunidades de bibliotecas e museus. Tal similaridade conceitual abre caminho para que os usuários tenham acesso integrado ao patrimônio

documental e cultural da humanidade, ampliando o escopo e a eficiência dos serviços prestados por estas instituições.

Constata-se a existência de diversas iniciativas pioneiras e relevantes de integração no campo do patrimônio documental e cultural. Entre os projetos mais notáveis está a Europeia, com um modelo conceitual inovador baseado em ontologia para integrar o acesso a arquivos, bibliotecas e museus europeus. O modelo foi adotado como referência pelo projeto da Digital Public Library of America (DPLA). Adicionalmente, a Biblioteca do Congresso dos Estados Unidos já pôs à disposição uma parte de seus vocabulários controlados e descrições bibliográficas, seguindo o padrão *Linked Open Data*. Projetos como Wikipedia-VIAF-DBpedia, OCLC-WorldCat, Biblioteca Nacional da França e Biblioteca Nacional Britânica também seguem essa tendência (Gueguen *et al.*, 2013a).

Entretanto, é perceptível que bibliotecas e museus iniciaram a adoção de tecnologias semânticas antes dos arquivos. Diversas razões podem ser levantadas para justificar a adesão tardia dos arquivos aos modelos conceituais e à Web semântica:

[...] suas práticas descritivas são menos normalizadas que as das outras comunidades do patrimônio, [...] (sendo a hierarquia e as listas ordenadas um desafio particular) e a falta de uma estrutura conceitual clara e abrangente para a descrição arquivística. Com a orientação fornecida pelas normas do ICA, em particular a ISAD(G), a comunidade arquivística tem feito grandes avanços na padronização da prática descritiva (Gueguen *et al.*, 2013a, p. 108).

Como referência teórica, o EGAD reconheceu e adotou uma variedade de modelos conceituais de arquivos, desenvolvidos em projetos nacionais, bem como modelos conceituais de bibliotecas e museus. No que tange aos modelos já utilizados em arquivos, constata-se que os projetos desenvolvidos na Espanha e na Finlândia são identificados como modelos conceituais de descrição arquivística. Por outro lado, os projetos da Austrália e da Nova Zelândia são classificados como modelos conceituais voltados para a gestão de documentos, no que se refere especificamente a metadados e requisitos funcionais de sistemas.

O modelo mais antigo origina-se da colaboração entre Austrália e Nova Zelândia, resultante da fusão do *Australian Government Recordkeeping Metadata Standard Version 2.0 (AGRkMS)*, de 2008, com o *Technical Specifications for the Electronic Recordkeeping Metadata Standard Version 1.0*, também de 2008, elaborado pelo Arquivo Nacional da Nova Zelândia. Os modelos evoluíram a partir da revisão do *Recordkeeping Metadata Standard for Commonwealth Agencies Version 1.0*, da Austrália, inicialmente desenvolvido em 1999. O modelo australiano-neozelandês é baseado na descrição de entidades múltiplas, distanciando-

se da concepção tradicional em arquivologia, concentrada na descrição de uma entidade única. A visão de múltiplas entidades propõe a separação e a inter-relação de entidades, expandindo o escopo para além da perspectiva tradicional no fundo/documento. Em tal modelo, são reconhecidas cinco entidades principais: documento, agente, atividade, atribuição e relacionamento. Nota-se uma relação mais horizontal entre as entidades, em contraste com a tradição teórica e prática da descrição arquivística, que, historicamente, deu prioridade à entidade fundo/documento. Merece destaque especial a entidade relacionamento, por registrar as atividades humanas e eventos que resultam na criação de documentos de arquivo, servindo como elo entre as outras quatro entidades do modelo. Ademais, cabe mencionar que o modelo australiano-neozelandês é influenciado por duas normas *ISO*: a *Australian Standard on Records Management (AS ISO 15489)* e a *Metadata for Records (AS ISO 23081)*.

Um segundo modelo conceitual expressivo na área de arquivologia, que exerceu influência sobre o modelo *RiC*, tem suas raízes na Espanha. O *Modelo Conceptual de Descripción Archivística y Requisitos de Datos Básicos de las Descripciones de Documentos de Archivo, Agentes y Funciones* foi desenvolvido entre 2007 e 2012 pela Comissão de Normas Espanholas de Descrição Arquivística (CNEDA). O modelo espanhol integra diversas referências, inclusive o modelo Entidade-relacionamento (ER), as quatro normas do ICA, o sistema ICA-AtoM, as normas *ISO 15489 e 23081*, além do modelo australiano-neozelandês AGRkMS. O modelo espanhol CNEDA, ao se assemelhar ao modelo AGRkMS, sublinha uma fase de transição nos arquivos, que passam da prática da descrição unidimensional, concentrada na representação de fundos e documentos, para uma descrição multidimensional, que enfatiza a representação de várias entidades. O modelo CNEDA delinea seis entidades distintas: agentes, atividades, atribuições, conceitos, documentos, localidades e objetos (eventos). Embora o relacionamento não seja categorizado como entidade separada no modelo espanhol, ele desempenha um papel determinante, facilitando a conexão entre as entidades.

Uma terceira vertente de influência para o modelo *RiC* é o *Finnish Conceptual Model for Archival Description*, desenvolvido pelo Arquivo Nacional da Finlândia *Arkistolaitos*. Publicado em 2013, o modelo finlandês – concebido sob a perspectiva de acesso integrado ao patrimônio documental e cultural, abrange arquivos, bibliotecas e museus – adota o padrão *Resource Description and Access (RDA)*. Originário das bibliotecas, o *RDA* baseia-se nos modelos *FRBR*, *CRM* e *Functional Requirements for Bibliographic Records Object Oriented (FRBRoo)*. O modelo finlandês levou em consideração as especificidades das normas de descrição do ICA e o princípio da proveniência. Similarmente aos modelos AGRkMS e

CNEDA, reforça a separação e a inter-relação das entidades descritivas. No modelo, incluem-se as entidades: agente, assunto, atribuição, evento do ciclo de vida (notavelmente, dedicada à descrição das atividades relacionadas à gestão de documentos), evento temporal, fonte de informação, função e localidade.

Dois outros projetos de modelos conceituais na área arquivística também influenciaram o EGAD. O projeto *Archives Hub Linked Data (LOCAH)*, do Reino Unido, visa a integrar descrições arquivísticas em grandes conjuntos de dados na Internet, utilizando o padrão *LOD*, com especial atenção na conversão de descrições pertencentes ao padrão *EAD*. As entidades definidas no modelo *LOCAH* incluem: agente, assunto, documento, evento, localidade e tempo. O projeto sofreu influência dos serviços de *LOD* da Biblioteca do Congresso dos Estados Unidos e do *Bibliographic Framework (BIBFRAME)*. Por sua vez, o projeto ICA-AtoM, desenvolvido pela Artefactual Systems, fundamenta-se nas normas de descrição arquivística do ICA e incorpora ontologias amplamente utilizadas, que não são exclusivamente arquivísticas.

No que se refere aos modelos conceituais das comunidades aliadas que tiveram impacto no desenvolvimento do modelo *RiC*, já na década de 1990, a Federação Internacional das Associações de Bibliotecas estabeleceu um grupo de estudo com o objetivo de identificar requisitos funcionais para fontes bibliográficas, mirando a criação de um modelo conceitual para a descrição de acervos em bibliotecas. Entre 1992 e 1995, o grupo da IFLA empregou as técnicas do modelo Entidade-relacionamento (ER) no desenvolvimento do modelo para as bibliotecas. Paralelamente, em 1996, o Conselho Internacional de Museus (ICOM), junto ao Comitê Internacional de Documentação (CIDOC), deu início à elaboração de um modelo conceitual destinado à descrição de objetos museológicos, denominado modelo de referência conceitual (CRM). Embora inicialmente voltado para os museus, o escopo do modelo foi posteriormente ampliado, transformando-o em um modelo de referência com o fim de facilitar o acesso integrado aos dados do patrimônio cultural provenientes de museus, bibliotecas e arquivos.

Desde 2003, um grupo de trabalho foi estabelecido para harmonizar os modelos conceituais das bibliotecas (*FRBR*) e dos museus (*CRM*). Este grupo se dedicou a atividades de mapeamento conceitual entre os modelos, assim como ao seu aprimoramento e refinamento, quando necessário. Como resultado, o modelo *CRM* evoluiu para se tornar o mais completo e abrangente modelo conceitual semântico, orientado para bancos de dados baseados em objetos, o que culminou na criação do *FRBRoo*, publicado em 2009. No relatório de revisão das normas do ICA, os especialistas do EGAD afirmam ter havido interesse por parte dos arquivistas em

participar do projeto *FRBRoo* e incorporar a descrição arquivística no modelo. No entanto, a participação foi limitada, dada a inexistência de um modelo conceitual específico e estabelecido para a área arquivística naquela época (Gueguen *et al.*, 2013b).

De acordo com o ICA, a promoção de uma colaboração mais intensa entre arquivos, bibliotecas, museus e instituições culturais é vista de maneira positiva, especialmente ao considerar a possibilidade de acesso integrado aos dados do patrimônio que tais instituições custodiam. A perspectiva de colaboração tem sido, cada vez mais, reconhecida e valorizada não apenas pelos profissionais desses campos, mas, também, pelos gestores políticos, agências de financiamento e pelos próprios usuários (ICA, 2023).

Embora reconheça as essenciais diferenças que caracterizam cada uma das áreas envolvidas com o patrimônio documental e cultural, o ICA sugere a adoção de uma visão mais compartilhada e colaborativa entre as áreas. Segundo a organização, isso não implicaria na diminuição das perspectivas particulares de cada campo, mas promoveria o acesso integrado aos dados oriundos de arquivos, bibliotecas, museus e instituições culturais. De acordo com o documento apresentado pelo ICA (2023),

Diferentes comunidades de patrimônio cultural têm, fundamentalmente, entendimentos distintos sobre a natureza dos objetos os quais custodiam. Esses entendimentos e práticas estão bem estabelecidos e adaptados (e se adaptando) às dificuldades particulares de cada uma dessas comunidades e às especificidades dos objetos com que lidam. Enquanto muitos dos esforços para alcançar o objetivo do acesso integrado concentram-se em desenvolver normas de descrição compartilhadas, diminuir as diferentes perspectivas [destas áreas do patrimônio] sobre as práticas de descrição é intelectual e politicamente desafiador (ICA, 2023, p. 3, tradução nossa).

De acordo com o ICA, como exposto em seu modelo, a visão de acesso a dados conectados oriundos das áreas do patrimônio cultural não implica limitações ou reducionismos às atividades desses setores. Isto se deve ao fato de que arquivos, bibliotecas e museus requerem apenas uma forma específica de colaboração: a adoção de conceitos e práticas de descrição amplamente compartilhados entre eles. Tal tratamento facilita o acesso mútuo aos dados do patrimônio cultural, sempre respeitando os princípios e as necessidades distintas de cada área (ICA, 2023).

Em linhas gerais, as instituições responsáveis pela custódia do patrimônio documental, apesar de aderirem à Web semântica de modo assimétrico, têm progressivamente modernizado os seus sistemas de busca e de acesso a fundos e coleções – a maior parte em forma de sites na *Web*: “a publicação de acervos de arquivos, bibliotecas e museus diretamente na Web usando

as tecnologias da web semântica e de dados abertos interligados permite a estas instituições tirarem partido das sinergias múltiplas” (Marcondes, 2016, p. 78).

Esta integração procura repercutir positivamente os pesquisadores de dados do patrimônio documental e cultural. O ICA projeta uma transformação na maneira como os arquivos são utilizados, sobretudo por um grupo específico de usuários: os pesquisadores científicos. Assim, o ICA assegura que o *RiC*, além de elevar a experiência desses usuários, irá também aprofundar a sua compreensão sobre os documentos arquivísticos. Isso porque, como já afirmado, o modelo facilita “uma representação mais extensa dos contextos nos quais os documentos foram criados e utilizados ao longo do tempo” (ICA, 2023, p. 3, tradução nossa).

Pesquisadores com interesse no uso de dados provenientes da descrição de documentos arquivísticos terão a oportunidade de acessá-los sob a ótica de suas respectivas disciplinas acadêmicas, já que o novo modelo não define um tipo específico de produto final. Em vez disso, o *RiC* concentra-se em colocar à disposição os dados dos arquivos, de acordo com os princípios do modelo conceitual, de maneira aberta e conectada.

2.2.4 O modelo conceitual *Records in Contexts*

Depois de uma década de dedicação dos especialistas do EGAD, o ICA publicou três versões de consulta do *RiC*, entre 2016 e 2021, período em que a comunidade arquivística pôde avaliar o modelo. Em novembro de 2023, a versão oficial deste modelo foi divulgada. Como visto nas subseções desta pesquisa, o *RiC* inova ao incorporar discussões teóricas e práticas recentes, como: a expansão do princípio da proveniência; as críticas às normas de descrição estabelecidas pelo ICA e o modo como as normas foram adotadas pelas instituições de arquivos. Além do que, sugere a atualização tecnológica.

Em relação ao seu público destinatário, o *RiC* o considera mais abrangente do que o das normas de descrição do ICA. Apesar de ainda voltado aos arquivos permanentes, como consta em seu texto, o modelo inclui os arquivistas dedicados à gestão documental nas fases corrente e intermediária. A meta é aproximar os padrões das práticas de descrição desde a produção dos documentos, dada à complexidade própria do ambiente arquivístico digital. Adicionalmente, o *RiC* fornece condições para se pensar em possíveis relações com os profissionais de áreas correlatas ao patrimônio documental – biblioteconomia e museologia –, pelo incentivo a projetos de integração de dados destas instituições. Destina-se, ainda, aos desenvolvedores de sistemas e profissionais da TI, tendo em vista a intrincada natureza tecnológica do modelo *RiC*

e os problemas de sua aplicação. Por fim, contempla os pesquisadores dos arquivos, pois permite a utilização e a reutilização dos dados sob variados pontos de vista acadêmicos (ICA, 2023).

Como anteriormente mencionado, o modelo *RiC* é fundamentado no modelo Entidade-relacionamento e na descrição multidimensional, abraçando a definição de entidades, atributos e relacionamentos. As entidades representam os elementos centrais de interesse para profissionais da gestão de documentos e dos arquivos permanentes, uma vez que são elas que “fornecem o contexto intelectual que atende às necessidades de gerenciamento físico, da preservação, da localização, da utilização e da compreensão dos documentos ao longo de sua trajetória histórica” (ICA, 2023, p. 17, tradução nossa). No *RiC*, foram estabelecidas 19 entidades²¹, organizadas em quatro níveis hierárquicos, apresentado no Quadro 7.

Quadro 7: Entidades do modelo *RiC* em quatro níveis de hierarquia.

Hierarquia de Entidades do RiC			
1º Nível	2º Nível	3º Nível	4º Nível
RiC-E01 Coisa ou Entidade	RiC-E02 Fonte documental	RiC-E03 Conjunto documental	
		RiC-E04 Documento	
		RiC-E05 Parte de documento	
	RiC-E06 Registro documental (manifestação documental ou inscrição)		
	RiC-E07 Agente	RiC-E08 Pessoa	
		RiC-E09 Grupo	RiC-E10 Família
			RiC-E11 Entidade coletiva

²¹ Na versão 1.0 do *RiC*, lançada em 2023, observa-se a revisão na estrutura das entidades, onde o número total foi reduzido de 22 para 19. A mudança exclui as subclasses de data, consolidada em uma só entidade, identificada como *RiC-E18*. Apesar das alterações, a entidade localidade manteve a identificação original *RiC-E22*, escolhendo-se não renumerá-la para *RiC-E19*.

		RiC-E12 Cargo	
		RiC-E13 Mecanismo	
	RiC-E14 Evento	RiC-E15 Atividade	
	RiC-E16 Regra	RiC-E17 Atribuição	
	RiC-E18 Data		
	RiC-E22 Localidade		

Fonte: ICA (2023, tradução nossa).

As entidades foram especificadas em 19 quadros que contêm informações como ID, nome, definição, nota de escopo, exemplos e comentários (quando necessários), conforme exemplo da entidade documento no Quadro 8.

Quadro 8: Exemplo de quadro explicativo de entidade no modelo *RiC*.

ID	RiC-E04
Nome	Documento
Definição	Informação registrada por pelo menos um método em um suporte durável e recuperável, por um agente, no curso de sua vida ou atividades.
Nota de escopo	<p>Documento é um tipo de fonte documental (RiC-E02).</p> <p>Um documento pode ser, ou tem que ter, ao menos um registro. Um documento pode ter mais de um registro.</p> <p>Um registro originado de outro registro de documento pode ser considerado um registro do mesmo documento ou um registro de novo documento, a depender do contexto.</p> <p>Um documento pode servir para várias finalidades, apesar de serem sempre documentos ou provas de uma atividade.</p>
Exemplos	<p>Ato de nomeação de John Bambridge, xerife de Leicestershire, com o 3º Grande Selo do Charles I anexado.</p> <p>Esboço do mapa da Península do Qatar.</p> <p>Mensagem de e-mail relativo ao acordo de participação no Congresso do ICA em Seul (2016), contendo dois anexos e um sinal digital.</p>

	Registro de entrada do trabalhador alemão Jacob Schwarz na Hospedaria de Imigrantes da Ilha das Flores, vindo da Antuérpia no navio Graf Bismark e de sua saída para Porto Alegre.
Comentários	–

Fonte: ICA (2023, tradução nossa).

Cada entidade detém características distintas, denominadas de atributos, que definem a identidade da própria entidade. Indicada nas diretrizes do ICA (2023), a descrição das entidades contém a exibição de suas características intrínsecas. No entanto, ao contrário das normas tradicionais do ICA, que possuíam conjuntos uniformes de elementos de descrição independentemente do tipo de entidade a ser descrita no modelo *RiC* – fundo, grupo, série, dossiê, item (na ISAD-G); uma entidade coletiva, pessoa ou família (na ISAAR-CPF); ou ainda uma função, atividade ou tarefa (na ISDF) –, os atributos para cada entidade são pré-definidos e individualizados por entidade. Embora alguns atributos sejam compartilhados por várias entidades, existem atributos exclusivos de entidades específicas. O ICA (2023) reconhece que a distinção entre o que deve ser considerado uma entidade e o que deve ser classificado como um atributo não é uma tarefa trivial. No entanto, o EGAD escolheu designar entidades apenas aqueles fenômenos fundamentais no âmbito da gestão de documentos e dos arquivos permanentes.

Os atributos no modelo *RiC* são acompanhados de um esquema de valores, que não é prescritivo e nem proibitivo, mas indicativo. Em outras palavras, cada entidade está vinculada a um atributo por meio de uma afirmação, como exemplificado a seguir: o documento (entidade) **A** possui o idioma (atributo) **B**, em que **B** representa o valor do idioma. O esquema de valores dos atributos no modelo *RiC* pode ser categorizado em quatro tipos distintos: texto livre, texto baseado em modelo, valor controlado e valor baseado em regra (ICA, 2023).

O EGAD definiu o total de 41 atributos, que abrangem uma variedade de características, como tipo de atividade, nota de autenticidade, dimensões, tipo de suporte, classificação, condições de acesso, condições de uso, formato, tipo de entidade coletiva, entre outros. Os atributos estão associados às 19 entidades definidas no modelo. Para facilitar a sua compreensão, o EGAD criou quadros padronizados para cada um deles (ver Quadro 9), contendo informações essenciais, como número de identificação, nome, definição, entidades às quais se aplicam, especificações, possibilidade de extensões, repetibilidade, escopo e exemplos ilustrativos.

Quadro 9: Exemplo de quadro explicativo do atributo classificação.

ID	RiC-A07
Nome	Classificação.
Definição	Um termo, número ou sequência alfanumérica, geralmente proveniente de um vocabulário ou esquema de classificação externo, que qualifica o registro ou parte dele.
Domínio de entidade	Conjunto de documentos; documento; parte do documento.
Especificações	–
Extensibilidade	Pode ser estendida com qualquer número de atributos específicos.
Repetibilidade	Repetível
Esquema de valor	Valor controlado; texto baseado em modelo.
Escopo	Pode ser utilizado na descrição de um conjunto de registros, quando o valor do atributo é compartilhado por alguns ou todos os membros do conjunto de registros. O atributo não deve ser confundido com o atributo identificador, embora, em alguns casos, as informações possam ser as mesmas.
Exemplos	BUD-01-F002 [exemplo do número de classificação de um plano de classificação]. Gestão de recursos humanos Registro de estudantes Assuntos financeiros Itens digitalizados

Fonte: ICA (2023, tradução nossa).

Depois de fornecer as definições dos 41 atributos por meio de quadros explicativos, o modelo *RiC* coloca à disposição a lista das entidades acompanhadas de seus atributos correspondentes (ver Quadro 10). A coloração atribuída aos atributos indica aqueles compartilhados com entidades do nível hierárquico superior, em conformidade com a estrutura hierárquica das 19 entidades. Os atributos sem codificação de cores são exclusivos da entidade em questão.

Quadro 10: Exemplo de quadro de atributos da entidade documento.

RiC-E04	Documento
ID do atributo	Nome do atributo
RiC-A16	Nota descritiva
RiC-A22	Identificador
RiC-A28	Nome
RiC-A21	Histórico
RiC-A35	Dimensão
RiC-A38	Âmbito e conteúdo
RiC-A40	Estrutura
RiC-A03	Nota de autenticidade
RiC-A07	Classificação
RiC-A08	Condições de acesso
RiC-A09	Condições de uso
RiC-A10	Gênero documental
RiC-A17	Espécie documental
RiC-A24	Integridade
RiC-A25	Idioma
RiC-A26	Status legal
RiC-A39	Status (tradição documental)

Fonte: ICA (2023, tradução nossa).

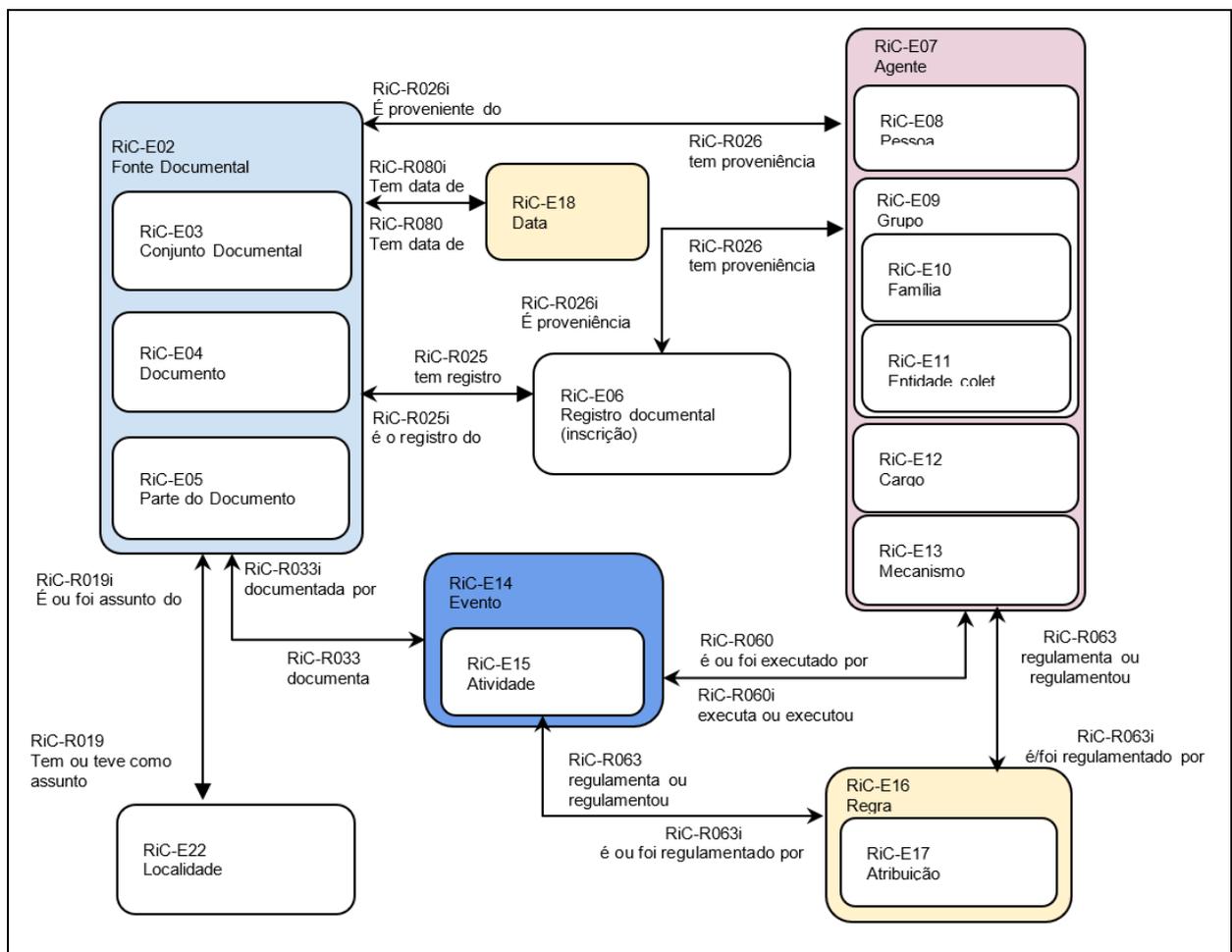
Por último, o modelo *RiC* exhibe o conjunto de relacionamentos, cujo objetivo é conectar as entidades. Os relacionamentos têm função central no modelo *RiC*, pois descrevem todos os contextos nos quais os documentos são produzidos, mantidos e utilizados – que é uma das principais metas do modelo. O ICA (2023) reforça que os relacionamentos são considerados simples e flexíveis, pois utilizam o modelo de grafos e o padrão *RDF* da Web semântica.

Publicado em novembro de 2023, o modelo *RiC* apresenta o total de 78 relacionamentos e, para cada um, a sua relação invertida. Eles estão agrupados em 13 categorias, e, de acordo com as orientações do ICA (2023), compreender o tipo de relacionamento esclarece o lugar das

entidades no sistema de descrição. As categorias incluem os tipos de relacionamentos possíveis: todo-parte; sequencial; por assunto; entre uma fonte documental e seu registro; proveniência; entre tipos de registro; de gestão; entre agentes; de eventos; de regras; de datas e espaciais. Assim, os relacionamentos podem variar de mais genéricos a mais específicos.

Os relacionamentos seguem uma hierarquia que vai do geral ao específico. O relacionamento mais amplo, ou mais geral, é "está relacionado com", que pode conectar qualquer entidade RiC-CM e não faz declaração específica sobre como ou por que essas entidades estão conectadas. Conforme se avança na hierarquia, cada relacionamento se torna mais específico, herdando as propriedades dos relacionamentos imediatamente superiores na hierarquia (ICA, 2023, p. 73, tradução nossa).

Figura 21: Relacionamentos entre entidades do modelo *RiC*.



Fonte: ICA (2023, tradução nossa).

Na aplicação do modelo *RiC* em sistemas, a ontologia *RiC-O* é fundamental. Florence Clavaud, do Arquivo Nacional da França e membro do EGAD do ICA, depois de testes feitos na França, sublinhou as vantagens da adoção do *RiC-O*. Segundo Clavaud (2021), ele oferece

duas possibilidades para a representação da informação: uma mais simplificada e outra mais precisa e complexa. Isso confere aos arquivistas a liberdade de escolher a alternativa mais adequada, considerando o nível de estruturação dos metadados descritivos do acervo em questão. Clavaud também destaca as vantagens na visualização e consulta dos dados semantizados no padrão *RDF*, incluindo a capacidade de conexão desses dados com outros conjuntos de dados *RDF*, tanto dentro da mesma instituição quanto fora.

Posteriormente, Clavaud (2021) salienta a possibilidade de integrar o modelo *RiC* com outros modelos conceituais por meio de mapeamentos, como a ontologia *Preservation Metadata: Implementation Strategies (PREMIS)* para metadados de preservação. Ela observa que as conversões para o padrão *RDF/RiC-O*, em curso no Arquivo Nacional da França e aplicadas a metadados de instrumentos de pesquisa, registros de autoridades e vocabulários controlados/tesauros, não alteram os arquivos-fonte das descrições originárias.

Por fim, Clavaud (2021) mostra a necessidade de várias ações para a implantação do modelo *RiC* no Arquivo Nacional da França. A exemplo da nova versão de um vocabulário de referência que contempla relacionamentos; os ajustes no vocabulário de funções e de localidades; a publicação de um vocabulário com licença-livre para padrão *RDF*, baseado no *RiC-O* e no *Simple Knowledge Organization System (SKOS)*²²; a atualização do *RiC-O Converter*²³, conforme as novas versões do *RiC-O*; os entendimentos sobre a aplicação do *RiC* a documentos nato-digitais; o desenvolvimento de um componente em *JavaScript Sparnatural*²⁴ para consultas em páginas *Web* sem conhecimento prévio de SPARQL ou ontologias; a aplicação do protótipo *Sparnatural* em um amplo conjunto de dados; e, finalmente, o desenvolvimento de um *software* livre personalizável para extração, classificação e uso de entidades no Wikidata.

²² Segundo o W3C (2024), o *Simple Knowledge Organization System (SKOS)* é uma recomendação do W3C projetada para representação de tesauros, esquemas de classificação, taxonomias, sistemas de controle de autoridade ou qualquer outro tipo de vocabulário controlado estruturado.

²³ O *RiC-O Converter* é uma ferramenta para converter dados nos padrões EAC-CPF e EAD 2002 XML para conjuntos de dados em *RDF*, conforme a *Records in Contexts Ontology (RiC-O)*. Disponível em github.com/ArchivesNationalesFR/rico-converter.

²⁴ *Sparnatural* é um componente em *Typescript*, subconjunto de *JavaScript*, que permite aos usuários explorar, de forma intuitiva, um grafo de conhecimento em *RDF* por consultas SPARQL. É possível configurá-lo para se adaptar à ontologia da sua área de conhecimento. Disponível em: <https://sparnatural.eu/>.

3 ANÁLISE QUALITATIVA

Esta seção se dedica à análise qualitativa dos dados obtidos nesta pesquisa.

3.1 Análise de conteúdo

Para atingir o objetivo de sistematizar as propostas do modelo *RiC* voltada à descrição arquivística, adotou-se a metodologia de pesquisa qualitativa e a análise de conteúdo. Laurence Bardin, renomada pesquisadora francesa, desenvolveu um sistema para a análise qualitativa de dados, que contempla a extração objetiva de significados e padrões contidos nos textos, possibilitando aos pesquisadores realizarem inferências. Segundo Bardin (2021), a análise de conteúdo é:

Um conjunto de técnicas de análise das comunicações visando obter [,] por procedimentos sistemáticos e objetivos de descrição de conteúdo das mensagens [,] indicadores (quantitativos ou não) que permitam a inferência de conhecimentos relativos às condições de produção/recepção (variáveis infinitas) dessas mensagens (Bardin, 2021, p. 44).

Em sua proposta, Bardin (2021) divide a análise em etapas de trabalho. A de pré-análise é, em geral, composta de três missões: a organização geral – onde se definem os documentos a serem analisados; a formulação das hipóteses e dos objetivos; e, por fim, a elaboração de indicadores para interpretação final.

Nesta pesquisa, a seleção de documentos para a análise de conteúdo se deu em várias etapas. De início, procedeu-se à análise geral do referencial teórico do modelo *RiC* e, a partir dela, a primeira redução de fontes de referência. Posteriormente, houve uma segunda diminuição pela exclusão de itens que tratam apenas de áreas específicas da arquivologia, biblioteconomia, museologia ou Web semântica. Por fim, definiu-se o conjunto de documentos a ser analisado, limitando-se aos textos do próprio modelo *RiC* e àqueles de seu referencial teórico que discutem a associação entre arquivos e a Web semântica – critério para definição do universo selecionado. Este procedimento detalhado, explicado a seguir, assegurou a seleção de um corpus pertinente e representativo para a análise de conteúdo subsequente.

Diante do referencial teórico completo que embasou o *RiC*, como consta na bibliografia do modelo, versão 0.1 (ICA, 2016), os itens vistos na seção padrões e modelos foram selecionadas para a leitura inicial. Optou-se por uma primeira redução de fontes, com a exclusão dos itens contidos na segunda seção da bibliografia, intitulada literatura profissional em

arquivos, por ser composta de inúmeros textos teóricos, cuja análise detalhada excederia o escopo temporal desta pesquisa. Contudo, é bom ressaltar que, dentre os textos teóricos excluídos da leitura, o Conselho Internacional de Arquivos havia destacado a relevância de oito deles em seu site oficial (marcados em cinza no Quadro 11), os quais optamos por incluir na leitura inicial. Assim, procedeu-se à leitura flutuante, que, como explica Bardin (2021), caracteriza-se por ser livre e não estruturada, em que o analista se familiariza com o material de modo aberto e intuitivo. Durante esta fase, o objetivo é obter uma visão geral e compreender o contexto, o estilo, a estrutura e os temas predominantes do material. Depois da leitura flutuante, os textos foram classificados em amplas áreas temáticas: arquivos; arquivos e Web semântica; bibliotecas, museus, patrimônio e Web semântica. As divisões constantes na referência bibliográfica do modelo *RiC* (versão 0.1) foram utilizadas como inspiração para a criação das subáreas dentro de cada uma dessas categorias principais, com aprimoramentos do autor desta pesquisa, com o objetivo de facilitar a compreensão das fontes de referência.

Quadro 11: Lista geral de fontes de referência do modelo *RiC* em que foi feita a leitura flutuante, na etapa de seleção de documentos para a análise de conteúdo.

Referências do modelo RiC	Área especializada	Área geral
ISAD(G): Norma geral internacional de descrição arquivística	Normas de descrição do ICA	Arquivos
ISAAR (CPF): Norma internacional de registro de autoridade arquivística para entidades coletivas, pessoas e famílias		
ISDF: Norma internacional para descrição de funções		
ISDIAH: Norma internacional para descrição de instituições com acervo arquivístico		
Rules for Archival Description (RAD), 2008	Normas de descrição e de conteúdo ²⁵	
Describing Archives: A Content Standard (DACS), 2007		
Encoded Archival Description (EAD), versão 2002	Normas de interoperabilidade em arquivos	
Encoded Archival Context-Corporate Bodies, Persons, and Families (EAC-CPF), 2010		

²⁵ A *RAD* e a *DACS* são normas de descrição de conteúdo porque, além da descrição, prescrevem regras de como o conteúdo da descrição deve ser apresentado.

Australian Government Recordkeeping Metadata Schema, 2015	Normas e modelos em gestão de documentos e de metadados arquivísticos	
Australian Recordkeeping Metadata Schema, 2000		
e-GMS-e-Government Metadata Standard, versão 3.0, 2004		
ISO 23081:2006 - Information and documentation - Records management - Metadata for records		
ISO 15489:2016 - Information and documentation - Records management - Part 1: Concepts and principles		
Principles and Functional Requirements for Records in Electronic Office Environments, ISO 16175: 2010/2011		
AS/NZS ISO 5478 Recordkeeping Metadata Property Reference Set (RMPRS), 2015		
Technical Specifications for Electronic Recordkeeping Metadata Standard, 2016		
MoReq2010: Modular Requirements for Records Systems, 2016		
Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives, 2013	Relatórios de revisão da descrição do ICA	Arquivos e Web semântica
Progress report for revising and harmonizing ICA descriptive standards, 2012		
Relationship in archival descriptive systems, 2012		
CNEDA. Modelo conceptual de descripción archivística y requisitos de datos básicos de las descripciones de documentos de archivo, agentes y funciones, 2012	Modelos conceituais em arquivos	
Report on the work of CNEDA (2007-2012): toward a conceptual model for archival description in Spain, 2012		
ICA AtoM Entity Model, 2012		
Towards the new era of archival description - the Finnish Approach, 2012		

Archives Hub Linked Data (LOCAH) project, 2011		
ANGJELI, A. Archives & Linked Open Data: are our tools ready to 'complete the picture'?, 2012	Arquivos e <i>Linked Open Data</i>	
HARPER, C. A. Linked Data in Archives: Publish, Enrich, Refine, Reconcile, Relate, 2012		
LEE, C. A. A framework for contextual information in digital collections, 2011		
MCLUHAN-MYERS, M. The Archival Web: Contextual Authority Files and the Representation of Institutional Textual Documents in Online Description, 2012		
CIDOC Conceptual Reference Model, version 6.2.2., 2016		
ISO 21127:2006. Information and documentation. A reference ontology for the interchange of cultural heritage information	Modelos conceituais em bibliotecas, museus e áreas do patrimônio	Bibliotecas, museus, áreas do patrimônio e Web semântica
Europeana Data Model and Primer, 2016		
IFLA. Functional Requirements for Bibliographic Records, 2008		
IFLA. Working Group on FRBR/CRM. Draft definition of FRBRoo, 2008		
IFLA. FRBR/CRM Harmonisation Meetings Consolidated Minutes, 2007		
NOY, N. F.; MCGUINNESS, D. L. Ontology Development 101: A Guide to Creating Your First Ontology, 2001	Engenharia de ontologia	
LODE: An ontology for Linking Open Descriptions of Events	Padrões e modelos para Web semântica.	Web semântica
W3C Prov [provenance]		
SMITH, M.; WELTY, C.; MCGUINNESS, D. OWL Web Ontology Language Guide. W3C Recommendation, 2014		
GUHA, R.V.; BRICKLEY, Dan. RDF Vocabulary Description Language 1.0: RDF Schema, 2014		

RUBINSTEIN, A. An RDF vocabulary for describing archival collections and the names associated with them, 2011		
---	--	--

Fonte: Adaptado do ICA (2016).

Ainda na seleção de documentos, na segunda etapa, foram excluídas as fontes que não tratam diretamente da intersecção entre a descrição de arquivos e a Web semântica. Descartados foram os textos sobre aspectos tecnológicos da Web semântica, unicamente, como metodologias de desenvolvimento de ontologias, linguagens *RDF/OWL* e vocabulários utilizados, bem como retirados os textos sobre modelos conceituais específicos das áreas de bibliotecas (*FRBR*), dos museus (*CIDOC*) e do patrimônio cultural. Em relação às fontes sobre arquivos, foram excluídas as subcategorias anteriores às noções de modelos conceituais em arquivos, incluindo as próprias normas de descrição do ICA, as normas voltadas para a descrição de conteúdo (*DACS, RAD*), as normas de interoperabilidade em arquivos (*EAD, EAC-CPF*), bem como as normas e modelos relacionados à gestão de documentos e de metadados em sistemas de arquivamento (MoReq2010, normas ISO, esquemas de metadados da Austrália e Nova Zelândia).

Todos os textos descartados com temas exclusivamente da área dos arquivos, embora exerçam influência direta e específica na estruturação do modelo *RiC* – como inspiração para definição de entidades, atributos e relacionamentos –, não tratam de discutir ou revelar, de maneira explícita, as ponderações sobre a descrição arquivística na perspectiva da Web semântica. No entanto, foram mantidos os textos oriundos do grupo de trabalho do ICA que discutem a revisão das normas e a sua transição para o modelo *RiC*.

Na última etapa da seleção de documentos, definiu-se a lista de fontes selecionadas ou corpus, segundo Bardin (2021), para a análise de conteúdo, incluindo os três principais documentos do modelo *RiC* mais as fontes selecionadas do referencial teórico, como mostra o Quadro 12:

- Compêndio do modelo conceitual do *RiC*:
 - *Records in Contexts (RiC-CM)*, versão 1.0., 2023;
 - *Records in Contexts Ontology (RiC-O)*, versão 0.2, 2023;
 - *Records in Contexts: Foundations on Archival Description (RiC-FAD)*, versão 1.0, 2023.

Quadro 12: Lista final de fontes selecionadas para a análise de conteúdo.

Referências do modelo RiC	Área especializada	Área geral
Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives, 2013	Relatórios de revisão da descrição do ICA	Arquivos e Web semântica
Progress report for revising and harmonizing ICA descriptive standards, 2012		
Relationship in archival descriptive systems, 2012		
CNEDA. Modelo conceptual de descripción archivística y requisitos de datos básicos de las descripciones de documentos de archivo, agentes y funciones, 2012	Modelos conceituais em arquivos	
Report on the work of CNEDA (2007-2012): toward a conceptual model for archival description in Spain, 2012		
ICA AtoM Entity Model, 2012		
Towards the new era of archival description - the Finnish Approach, 2012		
Archives Hub Linked Data (LOCAH) project, 2011		
ANGJELI, A. Archives & Linked Open Data: are our tools ready to 'complete the picture'?, 2012	Arquivos e <i>Linked Open Data</i>	
HARPER, C. A. Linked Data in Archives: Publish, Enrich, Refine, Reconcile, Relate, 2012		
LEE, C. A. A framework for contextual information in digital collections, 2011		
MCLUHAN-MYERS, M. The Archival Web: Contextual Authority Files and the Representation of Institutional Textual Documents in Online Description, 2012		

Fonte: Elaborado pelo autor (2023).

Em resumo, a seleção de documentos seguiu três etapas, como visto no Quadro 13.

Quadro 13: Resumo das etapas da seleção de documentos para a análise de conteúdo.

	Ação	Reduções
Lista 1	Análise geral da bibliografia do modelo RiC.	Exclusão da seção literatura profissional em arquivos.
Lista 2	Leitura flutuante de oito textos-teóricos (destacados pelo EGAD) e dos textos da seção padrões e modelos.	Exclusão de textos que não tratam da relação arquivos-Web semântica.
Lista 3 (final)	Definição de corpus dos documentos para análise de conteúdo.	

Fonte: Elaborado pelo autor (2023).

Bardin (2021) estabelece diretrizes para a seleção de documentos em pesquisas. A regra da exaustividade estipula que, uma vez definido o tipo de documento a ser analisado, não deve haver uma seleção adicional. Portanto, ao definir as categorias diretamente relacionadas ao objetivo de estudar as propostas do *RiC* em relação à descrição arquivística, foram mantidos todos os itens do próprio compêndio do modelo – *RiC-CM*, *RiC-O* e *RiC-FAD* –, bem como todos os itens das categorias relatórios de revisão das normas do ICA; modelos conceituais em arquivos; e arquivos e *Linked Open Data* – única categoria que versa a respeito da Web semântica em arquivos. Em contrapartida, a regra da representatividade, que permite a amostragem com a condição de que os resultados sejam extensíveis a todo o universo estudado, não foi aplicada, visto que o escopo da análise abrangeu a totalidade dos documentos selecionados. Por sua vez, em todos os documentos foram utilizados o mesmo critério de escolha (regra da homogeneidade), pois todos tratam das novas perspectivas da descrição arquivística sob a ótica da Web semântica, garantindo, assim, a regra da pertinência: os documentos são adequados ao objetivo da análise.

A fase subsequente do estudo abarca a formulação de hipóteses e objetivos. Conforme orientado por Bardin (2021), é fundamental estabelecer uma suposição preliminar a ser investigada, isto é, a hipótese, bem como definir claramente a finalidade da análise, ou seja, seu objetivo. Por conseguinte, a presente pesquisa tem, como hipótese, que as fontes selecionadas oferecem recentes percepções para a descrição arquivística sob a ótica da Web semântica e do modelo *RiC*. O objetivo da análise é identificar, nas fontes documentais, as alterações propostas pelo modelo *RiC*, no âmbito da descrição arquivística para fins de sistematização, destacando inovações e potenciais mudanças neste campo.

Depois, critérios para a análise e interpretação dos dados coletados são estabelecidos. Segundo Bardin (2021, p. 126), “um índice pode ser a menção explícita de um tema numa mensagem”. Os índices são identificados por meio de fragmentos específicos do texto. O objetivo desta pesquisa é identificar os principais questionamentos nos materiais selecionados, críticas e considerações acerca da teoria e prática da descrição arquivística, bem como as inovações, soluções e benefícios oferecidos pelo modelo *RiC* no âmbito arquivístico.

3.2 Análise e sistematização dos resultados

Finalizada a pré-análise, é posta em marcha a exploração do material. Observa-se que os textos selecionados apresentam variações consideráveis quanto à sua organização e natureza. Alguns deles possuem um caráter predominantemente teórico, como os relatórios de revisão de normas do Conselho Internacional de Arquivos, enquanto outros são mais técnicos, exemplificados pelo modelo *RiC* e sua ontologia *RiC-O*. Nestes últimos, seções específicas – introdução, objetivos, escopo – foram escolhidas para análise, uma vez que a discussão teórica não é o pilar de tais documentos.

Como propõe Bardin (2021), realizou-se a identificação de temas, assim como a codificação e a categorização dos dados. Os fragmentos textuais extraídos de todos os documentos analisados foram compilados e organizados em três categorias: críticas à descrição e à teoria arquivística; críticas às tecnologias empregadas nos arquivos atualmente; e soluções sugeridas pelo modelo *RiC*, conforme ilustrado no Quadro 14.

Quadro 14: Exemplo de trecho de análise e identificação de temas.

Documento	Críticas à descrição e à teoria arquivística	Críticas às tecnologias atuais dos arquivos	Soluções propostas pelo RiC
<i>Records in Contexts</i> (RiC-CM), versão 1.0., 2023	Visão tradicional do princípio da proveniência (restrita aos produtores e/ou acumuladores).		Visão ampla do princípio da proveniência (outros contextos além da produção e/ou acumulação).
	Descrição multinível é fechada – estabelece poucas relações com contexto externo ao fundo.		Descrição multidimensional – padrão se adapta às complexas relações do fundo com o contexto.

	Inconsistência e pouca integração entre as normas de descrição.		Registro de relacionamentos contextuais entre as entidades.
		Similaridade da entrada e saída dos dados no sistema advinda do uso das normas pelas instituições de arquivos.	Padronização das entradas de dados no sistema. Inúmeras possibilidades de saída de dados e interfaces aos usuários (inovação e experimentação).

Fonte: Elaborado pelo autor (2023).

Na etapa subsequente, procedeu-se ao agrupamento – em cada uma das três classes criadas – dos trechos extraídos de todos os documentos analisados. Depois, foram ainda juntados por similaridades, resultando na criação de subcategorias que refletem a natureza específica de cada conjunto de trechos, aproximando-se da codificação, que “corresponde a uma transformação – efetuada segundo regras precisas – dos dados em bruto do texto, transformação esta que, por recorte, agregação e enumeração, permite atingir uma representação do conteúdo, ou de sua expressão” (Bardin, 2021, p. 129).

Neste momento, depois de sistematizados os resultados, procedeu-se ao seu tratamento e interpretação. Observou-se que as críticas direcionadas à descrição arquivística abrangem a sua teoria; as normas internacionais estabelecidas pelo ICA, de forma particular; e o uso a elas dado quando de sua aplicação pelas instituições arquivísticas. A descrição arquivística é frequentemente percebida como um fator que contribui para o isolamento da descrição dos arquivos, pois – além do registro prioritário do contexto de produção/acumulação e a opacidade de outros contextos conectados aos arquivos – a atual descrição não incentiva parcerias entre arquivos, bibliotecas e museus. De modo igual, ela contém dificuldades de integração entre as quatro normas do ICA e a respectiva aplicação em sistemas variados; inflexibilidade de alterações específicas por parte das instituições arquivísticas; problemas de padronização, como a ambiguidade de conceitos entre as quatro normas; e, finalmente, dificuldades de adaptação aos documentos digitais e às novas tecnologias.

Quadro 15: Sistematização das críticas à descrição e à teoria arquivística.

Críticas à descrição arquivística e à teoria arquivística	Subcategoria
Visão tradicional do princípio da proveniência (restrita ao acumulador).	Isolamento da descrição dos arquivos
A descrição multinível é fechada – estabelece poucas relações com contextos externos ao fundo.	
As normas de descrição de arquivos possibilitam pouca interoperabilidade com bibliotecas e museus.	
A descrição apenas dos fundos não atende adequadamente à realidade arquivística.	
Falta de consistência e de coerência entre as normas de descrição, pois foram elaboradas de maneira separada.	Problemas de integração das normas
Há pouca interoperabilidade das normas com outros sistemas, como a aplicação em sistemas de gestão de documentos. As normas são voltadas para os arquivos permanentes.	
As normas falharam em articular um quadro conceitual abrangente para serem trabalhadas em conjunto.	
A preponderância da ISAD e da descrição apenas dos documentos não beneficia os arquivos.	
As normas apresentam redundância em alguns pontos e necessitam de revisão e harmonização.	
Há problemas no registro de relacionamentos entre a descrição arquivística (ISAD) e as autoridades (ISAAR-CPF), pois eles são registrados em eventos/acontecimentos de forma separada.	
As normas de descrição não oferecem orientação precisa sobre o seu uso unificado em sistemas.	
Há poucos relacionamentos para as funções da ISDF (eventos/acontecimentos).	
Existem os mesmos elementos de descrição, independente da unidade de descrição (fundo, dossiê, item).	Inflexibilidade
Descrição hierarquizada, do geral para o particular.	
Os elementos de descrição padronizados nas normas têm pouca flexibilidade de criação ou modificação em contextos arquivísticos específicos.	
A descrição em arquivos é menos padronizada do que nas demais comunidades do patrimônio (bibliotecas e museus).	Problemas de padronização

As normas não apresentam uma estrutura conceitual ou uma ontologia para a área dos arquivos.	
Há problemas e ambiguidades nos conceitos de fundo, atribuição, função e níveis de descrição.	
As normas não forneceram orientação adequada para o desenvolvimento da próxima geração de sistemas de descrição arquivística.	Problemas de adaptação às inovações tecnológicas
As normas não acompanharam as possibilidades tecnológicas atuais e emergentes.	
A ISAD(G) é mais voltada para a descrição de documentos analógicos. Ela não abrange adequadamente a descrição da documentação digital.	
A descrição baseada na ISAD(G) não considera metadados descritivos de documentos digitais.	
A descrição não inclui certos tipos de documentos digitais, como bancos de dados.	
Na atual descrição, não existem regras específicas de descrição para itens digitalizados.	
A atual descrição não considera a possibilidade de descrição feita por usuários e folksonomias.	
Os instrumentos de pesquisa on-line se assemelham aos instrumentos de pesquisa analógicos (lineares).	

Fonte: Elaborado pelo autor (2023).

Como mostra o Quadro 16, as críticas às tecnologias atualmente usadas nos arquivos referem-se, principalmente, às limitações das tecnologias de banco de dados *SQL* e das linguagens de marcação *XML*. Ademais, destacam-se as críticas ao uso de modelos hierárquicos e de estruturas rígidas, como os padrões *EAD* e *EAC-CPF*. Outro ponto posto em questão é a adoção de instrumentos de busca on-line que se assemelham aos analógicos, caracterizados por sua linearidade e por apresentarem um limitado manejo dos dados pelos usuários.

Quadro 16: Sistematização das críticas às tecnologias em uso nos arquivos.

Críticas às tecnologias atualmente em uso nos arquivos	Categorias
As bases de dados e as tecnologias de marcação não representam a realidade complexa dos arquivos.	Limitações das tecnologias de banco de dados e das linguagens de marcação
As tecnologias de base de dados e de marcação foram bastante usadas na transição para os computadores, mas são limitadas.	
A informação do mundo dos arquivos não está bem representada em SQL e XML.	
A linguagem de marcação XML é pouco flexível para representação das complexas relações da descrição dos arquivos.	
Os modelos hierárquicos em árvores são limitados.	Modelos hierárquicos com estruturas rígidas
Há o uso de estruturas rígidas, como EAD e EAC-CPF.	
Os instrumentos de pesquisa on-line se assemelham aos instrumentos de pesquisa analógicos e lineares.	Instrumentos de busca on-line semelhantes aos analógicos
A entrada e a saída dos dados são a mesma nos sistemas.	

Fonte: Elaborado pelo autor (2023).

Como observado na análise de conteúdo (Quadro 17), o modelo *RiC* introduz inovações no patamar da descrição arquivística. Inicialmente, por ser a ontologia da área de arquivos, ele reduz as ambiguidades e redundâncias vistas nas quatro normas de descrição do ICA. Ao lado do que, a descrição dos relacionamentos entre as entidades do *RiC* valoriza o registro dos inúmeros contextos arquivísticos – princípio fundamental para os arquivos enaltecido no modelo.

O *RiC* oferece, ainda, maior integração entre as entidades descritas em comparação às normas do ICA, com intenção de promover a interoperabilidade das descrições entre os sistemas da gestão de documentos e dos arquivos permanentes. Compreende-se que o *RiC* não é um referencial de aplicação direta no desenvolvimento de sistemas da gestão de documentos, porém, como salientado no texto do modelo, prevê-se que ele terá um papel importante na integração e alinhamento das práticas de descrição em todas as fases do ciclo vital. A maior capacidade de integração do *RiC* facilita igualmente as parcerias das instituições arquivísticas e, entre elas, bibliotecas e museus. Isso ocorre porque o *RiC* – fundamentado na lógica da Web

semântica, semelhante ao *FRBR*, utilizado em bibliotecas, e ao *CIDOC*, empregado em museus – compartilha com estes modelos a estrutura dos padrões *RDF* e *OWL*, facilitando a integração de dados com instituições de patrimônio cultural que já processam informações para a Web semântica.

O modelo distingue-se também por sua flexibilidade e adaptabilidade, pois permite alterações e inclusões na resolução de casos específicos nas instituições arquivísticas. Por fim, o *RiC* revoluciona o acesso e a recuperação de dados em arquivos, fundamentalmente por alterar a saída de dados nos sistemas e ter como um de seus objetivos atingir o uso de dados dos arquivos pelos pesquisadores acadêmicos.

Quadro 17: Sistematização das soluções do modelo *RiC* para a descrição arquivística.

Soluções do modelo <i>RiC</i> para a descrição arquivística	Subcategorias
Os atributos são individualizados por entidades no <i>RiC</i> .	Descrição arquivística sem ambiguidades e redundâncias
O <i>RiC</i> aprimora a descrição em termos de economia e precisão.	
O <i>RiC</i> usa novos métodos de representação e comunicação das informações de arquivo.	
O <i>RiC</i> é uma ontologia de domínio na área de arquivos para a Web 3.0. Não existia esquema conceitual na área de arquivos antes do <i>RiC</i> .	
As tecnologias semânticas oferecem formas mais expressivas de representação dos dados dos arquivos.	
O <i>RiC</i> reposiciona a descrição arquivística para melhor identificar, garantir autenticidade e gerenciar os documentos por meio da descrição de seu contexto, conteúdo e estrutura.	
As tecnologias semânticas criam redes de informação interrelacionadas, com estrutura mais precisa e lógica do que as atuais tecnologias em uso nos arquivos.	
O <i>RiC</i> é único e elimina redundância que existia entre as quatro normas de descrição.	
O <i>RiC</i> resolve problemas de definição de entidades e seus possíveis relacionamentos.	
O <i>RiC</i> resolve o problema dos conceitos de identidade e de designação das normas.	
O <i>RiC</i> esclarece os conceitos-chave de fundo, atribuição, função e níveis de descrição.	

O RiC promove a ampla visão do princípio da proveniência (inclui outros contextos além da produção/acumulação).	Valorização do contexto arquivístico
O RiC valoriza os relacionamentos contextuais entre as entidades.	
O RiC melhora o acesso e a compreensão das fontes arquivísticas descritas em diversos contextos.	
No RiC, a entidade evento tem mais recursos na criação de relacionamentos, como o registro da data inicial e da final.	
No RiC, o registro do contexto dos eventos/acontecimentos é bem explorado.	
No RiC, o registro dos relacionamentos entre as entidades coopera com a descrição dos arquivos.	
O RiC protege de forma mais sólida a base conceitual da descrição arquivística em relação à descrição em outros domínios (bibliotecas, museus).	
O RiC é flexível e pode ser testado em sistemas na gestão de documentos (arquivística integrada).	Integração dos arquivos e interoperabilidade
O RiC apoia a parceria de projetos com instituições de patrimônio, como bibliotecas e museus.	
Com o RiC, há a possibilidade de trabalhar cooperativamente dentro e fora da comunidade arquivística.	
No RiC, as tecnologias semânticas, como o RDF, conectam dados de forma a superar a separação entre diferentes sistemas descritivos e entre as entidades que eles descrevem.	
O RiC permite a conexão de dados criados em diferentes sistemas.	
O RiC revisa os relacionamentos entre as normas e facilita o intercâmbio entre informações arquivísticas, bibliográficas e museológicas.	
O RiC inclui a visão de ciclo vital <i>continuum</i> com visão pós-custodial.	
No RiC, a descrição multidimensional é mais aberta e se adapta às complexas relações com os contextos externos ao fundo.	Flexibilidade
No RiC, os atributos das entidades podem ser ampliados, pode-se usar vocabulários específicos, além de aplicações em contextos arquivísticos específicos.	
No RiC, as linguagens RDF e OWL possuem qualidades de representação e flexibilidade mais sofisticadas que os padrões anteriores.	

O RiC é genérico e pode ser adaptado às necessidades específicas das instituições arquivísticas.	
No RiC, o padrão OWL permite o uso de ferramentas de desenvolvimento e testes para assegurar a correção formal do modelo.	
No RiC, há a padronização das entradas, porém inúmeras possibilidades de saída de dados e de criação de interfaces aos usuários.	Fortalecimento do acesso e recuperação dos dados dos arquivos
No RiC, as tecnologias semânticas e de grafos em redes representam a complexidade do mundo dos arquivos.	
No RiC, há a possibilidade de uso de grafos e do RDF, com a capacidade de inferências pelos computadores.	
No RiC, prevê-se melhor recuperação, localização e compreensão dos documentos pelos usuários.	
O RiC oferece múltiplos caminhos de acesso aos recursos arquivísticos na Web.	
No RiC, as tecnologias semânticas podem ser usadas para gerar novos conhecimentos e oferecer um universo de informação mais rico que a soma de suas partes.	
No RiC, o RDF e os grafos oferecem ilimitadas representações dos dados dos arquivos.	

Fonte: Elaborado pelo autor (2023).

Na análise de conteúdo, verificou-se que a hipótese inicial se confirma: as fontes selecionadas introduzem novos pontos de vista na descrição arquivística. A Web semântica adotada no modelo *RiC* propõe avanços para a descrição em arquivos. As inovações incluem a eliminação de ambiguidades e redundâncias constantes nas quatro normas de descrição do ICA, a valorização dos contextos arquivísticos, bem como a promoção da integração e da interoperabilidade entre diferentes sistemas e instituições. Mais que isso, destacam-se a flexibilidade e a contínua atualização como aspectos positivos, juntamente com o fortalecimento do acesso e da recuperação de dados pelos usuários de arquivos. O *RiC* é um divisor de águas nas práticas arquivísticas contemporâneas, mais favoráveis às exigências e às possibilidades da era digital.

4 ANÁLISE EMPÍRICA

Esta seção se dedica à análise empírica dos dados obtidos nesta pesquisa.

4.1 Aplicação dos questionários

Para alcançar o objetivo de analisar a aplicação do modelo *RiC*, realizou-se pesquisa junto a profissionais de instituições arquivísticas engajados na execução de testes ou na gerência de projetos de uso do *RiC*. Elaborou-se um questionário em inglês, disponível, na íntegra, no **Apêndice A**. Com base no referencial teórico e na análise de conteúdo, foram criadas 16 perguntas – sete abertas e nove fechadas. Elas abarcam uma variedade de temas sobre o *RiC*, desde os dados mais gerais de identificação dos profissionais e dos projetos às principais motivações para o uso do modelo. O propósito do questionário foi a captura das opiniões dos inquiridos sobre o atual estado da descrição arquivística e das tecnologias em uso nos arquivos, bem como investigar, segundo as próprias experiências, os benefícios do modelo *RiC*.

Feito no *Google Forms*, o questionário foi enviado por e-mail para os profissionais que trabalham com o modelo *RiC*. A lista de instituições que testam ou usam o *RiC* foi identificada em diferentes fontes divulgadas pelo ICA, inclusive a página do grupo de especialistas do EGAD e a lista de participantes em eventos sobre o modelo. Não foi identificada a totalidade dos projetos que testam o modelo *RiC* em suas instituições e os 31 profissionais contatados, apenas nove responderam. Alguns profissionais, em especial os participantes do EGAD/ICA, justificaram não responder por causa do momento de consolidação do *RiC* junto ao ICA, publicado em novembro de 2023.

Por outro lado, o questionário foi preenchido pelos profissionais representantes das seguintes instituições:

- Arquivo da Cidade de Amsterdam (Gemeentearchief/Amsterdam City Archives), coordenador do projeto ACA²⁶;
- Biblioteca Estatal de Berlim (Staatsbibliothek zu Berlin/ Berlin State Library) e da Fundação do Patrimônio Cultural da Prússia (Stiftung Preußischer Kulturbesitz/Prussian

26

assets.amsterdam.nl/publish/pages/1002802/thesis_records_in_contexts_merel_geerlings_definitive_1.pdf

Cultural Heritage Foundation), à frente do projeto SoNAR²⁷, que inclui, ainda, o Departamento de Biblioteca e de Ciência da Informação da Universidade Humboldt de Berlim (Humboldt-Universität zu Berlin), o Centro Alemão de Pesquisa em Inteligência Artificial (DFKI), o Departamento de História, Filosofia e Ética da Medicina da Universidade Heinrich Heine de Düsseldorf (HHU) e o Laboratório de Complexidade Urbana (UCLAB) da Universidade de Ciências Aplicadas em Potsdam (FH;P);

- Docuteam AG²⁸, empresa suíça especializada em gestão de documentos e de informação, no projeto *Docuteam context* e *Docuteam box*;

- Arquivos cantonais de Vaud²⁹, na Suíça – Archives cantonales vaudoises (ACV) –, no projeto do Sistema de arquivamento histórico (*Système d'archivage historique*);

- Memoriav³⁰, associação suíça para a preservação do legado audiovisual da Suíça, no projeto Memobase;

- Arquivo Nacional da Noruega³¹ – Arkivverket/National Archives of Norway –, no projeto de catalogação de metadados denominado Mime;

- Arquivo Nacional da Suécia³² – Riksarkivet/Swedish National Archives –, no projeto Data Platform;

- Serviço Interministerial dos Arquivos da França – Service interministériel des Archives de France/Ministère de la Culture –, no projeto *FranceArchives*³³, em parceria com a empresa de desenvolvimento de software Logilab;

- Arquivo Nacional de Liechtenstein³⁴ – Liechtensteinisches Landesarchiv/Liechtenstein National Archives –, no projeto de aplicação do RiC na tese de bacharelado em Ciência da Informação do pesquisador Joy Walser.

²⁷ sonar.fh-potsdam.de/

²⁸ www.docuteam.ch/en/

²⁹ www.vd.ch/toutes-les-autorites/archives-cantonales-vaudoises-acv

³⁰ <https://memoriav.ch/en/>

³¹ <https://www.arkivverket.no/en>

³² <https://riksarkivet.se/startpage>

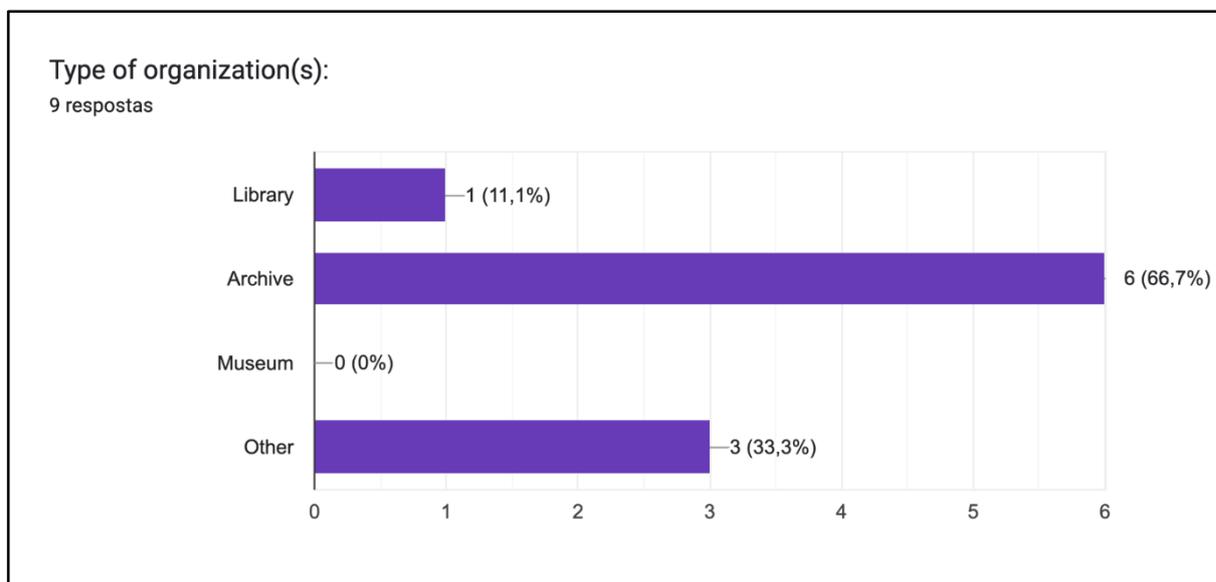
³³ <https://francearchives.gouv.fr/>

³⁴ <https://www.llv.li/de/landesverwaltung/amt-fuer-kultur/landesarchiv>

4.1.1 Análise e sistematização dos resultados

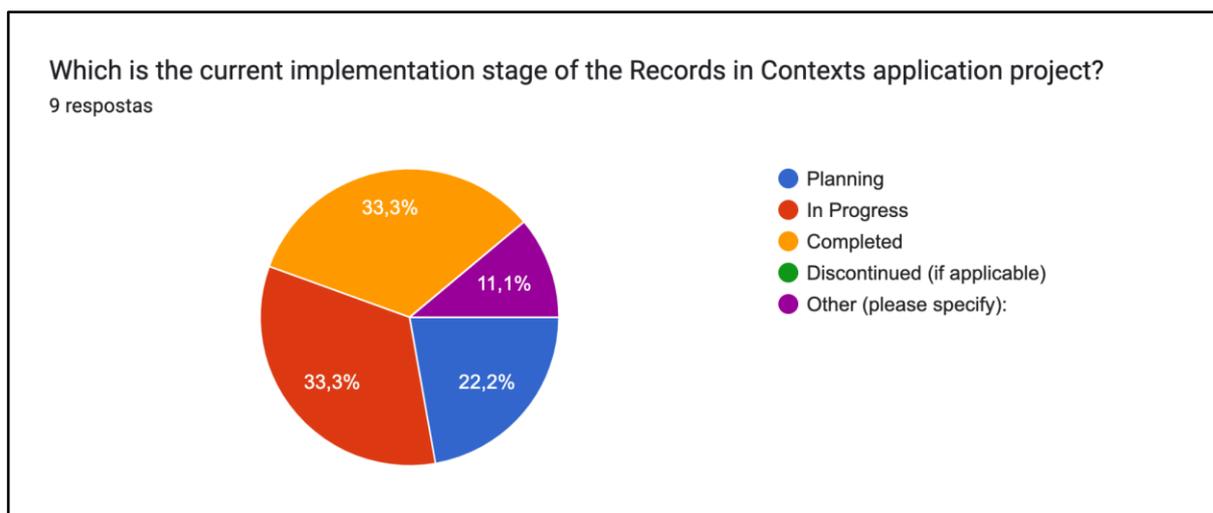
Na **questão 1** (fechada), a maior parte que respondeu o questionário está vinculada a instituições arquivísticas, contudo, constata-se também a participação de bibliotecas e outras instituições relacionadas, incluindo departamentos universitários, como ilustrado no Gráfico 1.

Gráfico 1: Tipos de instituições que aplicam o modelo *RiC*.



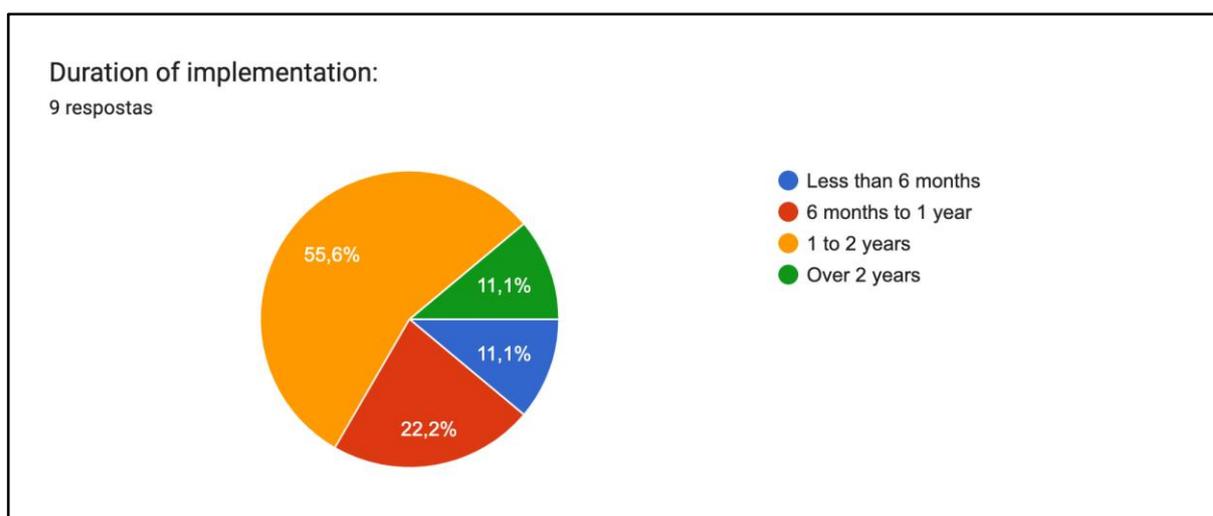
Fonte: Elaborada pelo autor (2023).

Na **questão 2** (fechada), que versa sobre o estágio em que se encontram os projetos de aplicação do modelo *RiC*, observou-se o equilíbrio entre projetos em fase de planejamento, em andamento e aqueles já concluídos. Exemplos de projetos finalizados incluem os da Docuteam AG e da Memoriav, ambos na Suíça, além do *FranceArchives*. Importante frisar que nenhum dos projetos foi descontinuado antes da conclusão.

Gráfico 2: Estágios de implementação dos projetos do modelo *RiC*.

Fonte: Elaborada pelo autor (2023).

Em termos de duração, pouco mais de 55% das respostas mostram que os projetos duram cerca de um a dois anos. Apenas o SoNAR, da Alemanha, existe há mais de dois anos. Ainda há os recentes, como o do Arquivo Nacional de Liechtenstein e do Arquivo Nacional da Suécia, com menos de um ano de iniciados.

Gráfico 3: Duração dos projetos de aplicação do modelo *RiC*.

Fonte: Elaborada pelo autor (2023).

Depois das perguntas de identificação e de contextualização dos projetos, iniciaram-se as indagações técnicas relacionadas à descrição arquivística, às tecnologias em uso nos arquivos e ao modelo *RiC*. A **questão 3** (aberta) investigou as razões de as instituições adotarem o *RiC*.

O Arquivo da Cidade de Amsterdam testou o modelo *RiC* em razão da insuficiência da norma ISAD(G) para descrever adequadamente os documentos nato-digitais. A Biblioteca Estatal de Berlim e a Fundação do Patrimônio Cultural da Prússia, no projeto SoNAR, adotaram o *RiC* com o fim de transformar os dados arquivísticos em *EAD* para o *RDF*, valorizando a capacidade deste padrão em inferir relações – especialmente entre agentes –, como integrar dados de bibliotecas e de museus, compreendendo metadados de autoridades e aplicação de ontologias específicas.

A Docuteam AG, empresa suíça, adotou o *RiC* por ser um padrão arquivístico recomendado. Os Arquivos Cantonais de Vaud, na Suíça, decidiram testar o *RiC* por várias razões: por ser o novo padrão do ICA; por ser recomendado pela Associação Suíça de Arquivistas; pela possibilidade de preservação a longo prazo de dados conectados; pela maior interoperabilidade com outros arquivos e sistemas; e pelo aprimoramento da descrição arquivística.

O Arquivo Nacional da Noruega aplicou o *RiC* por sua capacidade de interoperabilidade, por ser um modelo desenvolvido por especialistas de diversos países sob a égide do ICA e pela habilidade do modelo em expressar relações complexas entre os dados, sem falar das características de legibilidade e de autoexplicação do *RDF*. O Arquivo Nacional da Suécia procurou tornar os seus registros arquivísticos acessíveis no padrão de dados abertos conectados, optando pelo *RiC* por sua adequação superior em comparação ao *EAD*.

O Serviço Interministerial de Arquivos da França adotou o *RiC* com o intento de oferecer múltiplos pontos de acesso aos metadados arquivísticos. O projeto suíço *Memoriav* utilizou o *RiC* como recomendado por especialistas de bibliotecas, de arquivos e por desenvolvedores de sistemas. Por fim, o Arquivo Nacional de Liechtenstein escolheu o *RiC* para descrever uma coleção especial que não seria adequadamente descrita pela norma ISAD(G), visando a conectá-la ao contexto do material arquivístico.

Nas respostas, foram registrados diversos motivos para o uso do *RiC*:

- 4 projetos buscam a inovação e os benefícios das novas tecnologias, como a conversão do *EAD* para *RDF*, além das vantagens dos motores de inferência, pelo manejo de dados e pela flexibilidade no uso de vocabulários pré-existentes;

- 3 projetos ressaltaram a adoção de padrões mais modernos recomendados por especialistas da área;

- 2 projetos destacaram as limitações da ISAD(G) para a descrição dos arquivos digitais contemporâneos;

- 2 projetos referiram-se ao *Linked Open Data* como um potencial novo padrão para a codificação de dados dos arquivos, avaliado como mais apropriado que o *EAD* e menos limitado que os bancos de dados tradicionais;

- 2 projetos mencionaram a integração e a interoperabilidade como fatores motivadores, ressaltando a possibilidade de conexão entre diferentes sistemas de arquivos, incluindo aqueles relacionados à gestão documental, assim como sistemas de dados externos, como o Wikidata, além das parcerias para a integração de dados com bibliotecas e museus;

- 2 projetos mencionaram o aprimoramento no acesso e na flexibilidade de uso dos dados arquivísticos sob o padrão do modelo *RiC*;

- 1 projeto lançou críticas aos princípios da proveniência e do respeito à ordem original, considerados apropriados, mas insuficientes.

Na **questão 4** (aberta), solicitou-se aos participantes a identificação dos principais problemas – práticos e teóricos – da atual descrição de arquivos. O Arquivo da Cidade de Amsterdam identificou problemas na hierarquia da norma ISAD(G), observando que ela não se adequa a estruturas mais complexas contemporâneas, como as de sistemas do tipo *Document Management Systems (DMS)*, o que pode resultar em perda de registro de informação. A instituição ressalta que o modelo *RiC* permite registrar estruturas complexas de maneira mais eficiente na fase de gestão documental.

Já a Biblioteca Estatal de Berlim e a Fundação do Patrimônio Cultural da Prússia, do projeto SoNAR, constataram que a descrição arquivística atual é excessivamente centrada nas informações do contexto do produtor/acumulador dos documentos e se baseia em textos narrativos. Elas observam que diferentes usuários dos arquivos têm variados interesses, sugerindo a necessidade de migração para o modelo de dados conectados, por permitir múltiplas formas de recuperação e aumentar a visibilidade dos contextos dos dados dos arquivos.

A empresa suíça Docuteam AG registrou a dificuldade em vincular os dados de descrição de autoridades. Os Arquivos Cantonais de Vaud destacaram as limitações das normas ISAD(G) e ISAAR(CPF), especialmente a fragilidade dos relacionamentos entre as entidades e a falta de flexibilidade dos sistemas atuais, criando silos de dados internos. Eles reforçam que o *RiC-O* oferece relações mais flexíveis e uma riqueza na descrição que possibilita retratar os arquivos com maior precisão.

O Arquivo Nacional da Noruega notou que seus sistemas digitais de descrição, em sua maioria, são bancos de dados relacionais e, apesar de hierárquicos, não são completamente autocontidos, relacionando-se com outros sistemas. O Arquivo Nacional da Suécia grifou que

as descrições arquivísticas ainda são elaboradas de maneira semelhante à era pré-digital, com limitadas inovações. Mencionou o uso insuficiente de vocabulários controlados, que compromete a recuperação dos dados.

O Serviço Interministerial de Arquivos da França assinalou as semelhanças das ferramentas de busca on-line com instrumentos de pesquisa impressos, em termos de linearidade, além da falta de metadados estruturados, da heterogeneidade na indexação e da ausência de vocabulários compartilhados. O Arquivo Nacional de Liechtenstein mencionou as rígidas estruturas hierárquicas associadas à ISAD(G) e aos seus bancos de dados, exemplificadas no sistema *Scope Archive* – não muito adequado ao *RiC* – e também destacou a falta de conhecimento dos profissionais sobre o modelo *RiC*. Por fim, o projeto suíço *Memoriav* não se referiu à descrição atual, mas ao próprio *RiC*, e observou a dificuldade na aceitação institucional do modelo devido à sua complexidade e à falta de aplicativos de fácil acesso para uso do modelo.

Para a análise eficaz das respostas, utilizamos as críticas à descrição arquivística da análise de conteúdo:

- 4 projetos citaram a inflexibilidade da atual descrição hierárquica, lembrando a sua incompatibilidade com a estrutura de sistemas complexos, como os *Document Management Systems (DMS)*, frequentemente resultando em perda de informações ou em limitações na inclusão de dados de autoridades;

- 3 projetos mencionaram a dificuldade em adaptar a descrição arquivística às novas tecnologias, já que a descrição ainda é elaborada para ser consumida de maneira linear, semelhante ao formato tradicional em papel;

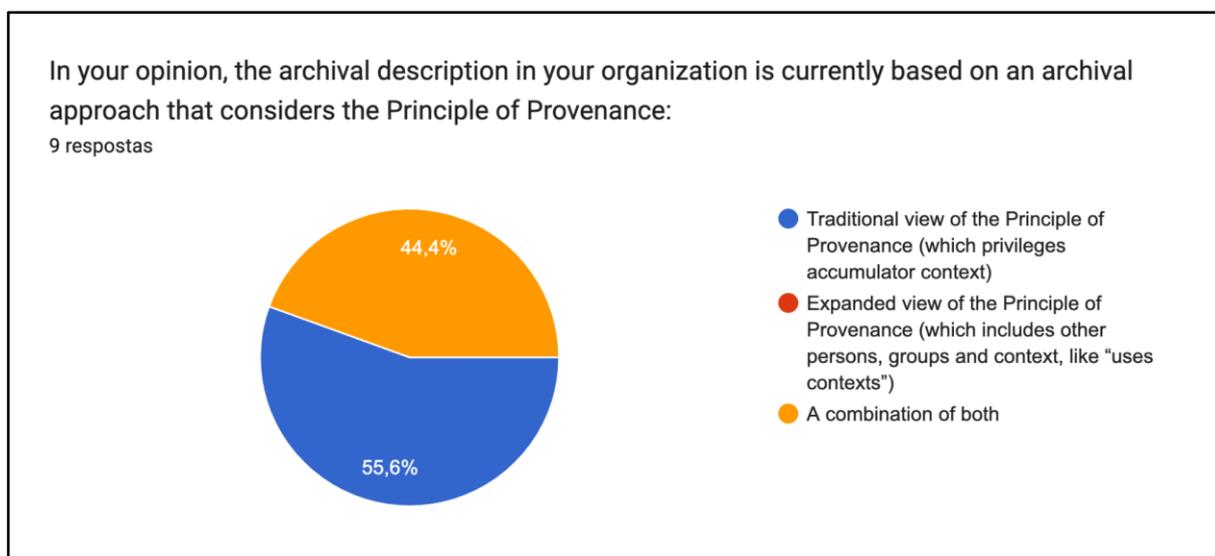
- 2 projetos referem-se ao isolamento na descrição dos arquivos, pois há um registro excessivo de informações sobre os produtores/acumuladores de arquivos a afetar as diversas perspectivas dos pesquisadores em relação a outros contextos dos documentos. Adicionalmente, as atuais normas de descrição apresentam limitações no registro dos relacionamentos temporais entre as entidades, restringindo a riqueza destas relações;

- 2 projetos indicaram que a falta de padronização também é um problema, pois há, atualmente, metadados desestruturados, heterogeneidade na indexação e uso limitado de vocabulários controlados compartilhados, o que prejudica a recuperação de informações.

A visão sobre o princípio da proveniência na descrição documental estruturada pelas instituições que aplicam o modelo *RiC* foi trazida na **questão 5** (aberta). Nas respostas, observa-se uma divisão: enquanto 55% das instituições seguem a aplicação tradicional do princípio –

centrado na descrição do contexto de produção e de acumulação de documentos –, 45% delas adotam uma perspectiva híbrida. A perspectiva combinada engloba tanto a visão tradicional quanto a forma mais inovadora, o princípio da proveniência ampliado – não limitado apenas à descrição do contexto de produção e acumulação, ele incorpora a identificação de outras pessoas, grupos e contextos relevantes, como o contexto de uso dos documentos.

Gráfico 4: Visão sobre o princípio da proveniência nas instituições que aplicam o modelo *RiC*.



Fonte: Elaborada pelo autor (2023).

A **questão 6** (aberta) tratou de quais problemas tecnológicos dos atuais sistemas de descrição fundamentaram a adoção do modelo *RiC* por parte das instituições arquivísticas. O Arquivo da Cidade de Amsterdam identificou a importância da transição da descrição multinível para uma multidimensional, revelando a rigidez do atual modelo das normas de descrição, que não conseguem adequadamente representar a complexidade dos arquivos. Paralelamente, a Biblioteca Estatal de Berlim e a Fundação do Patrimônio Cultural da Prússia, no projeto SoNAR, frisaram que as soluções tecnológicas adotadas devem ser holísticas e fundamentadas, atendendo a necessidades e requisitos específicos.

Os Arquivos Cantonais de Vaud, na Suíça, citaram a obsolescência de *softwares* baseados na ISAD(G)/EAD e na ISAAR(CPF)-EAC, como *AtoM*, *archivespace* e *scopeArchiv*, e reiteraram a urgência em migrar para tecnologias mais avançadas, como dados conectados e bancos de dados em grafos, visando melhor desempenho, flexibilidade e interoperabilidade. O Arquivo Nacional da Noruega reforçou a necessidade de sistemas de descrição mais duradouros e autoatualizáveis, considerando as vantagens da semântica de dados abertos em *RDF*.

Destacaram, ainda, a abrangência e as nuances do *RiC*, capazes de expressar relações complexas e uma ampla gama de conceitos. O Arquivo Nacional da Suécia declarou ter implementado o *RiC* de maneira específica, apenas para saída de dados, sem intenção de reestruturar completamente o seu sistema de descrição arquivística.

O projeto *Memoriav* enfrenta a grande heterogeneidade de dados audiovisuais conectados de diferentes instituições suíças. O Arquivo Nacional de Liechtenstein lembrou que os atuais sistemas não fornecem as funcionalidades necessárias à completa aplicação do *RiC*. O Serviço Interministerial de Arquivos da França considerou a questão não aplicável à sua situação, enquanto a Docuteam AG, empresa suíça, não respondeu à pergunta.

Para compreender as respostas, recorreu-se às categorias de análise de conteúdo:

- 2 projetos apontaram as limitações inerentes às atuais tecnologias de bancos de dados e às linguagens de marcação;

- 1 projeto ressaltou que os sistemas vigentes usam tecnologias ultrapassadas e, em alguns casos, obsoletas, porque os bancos de dados relacionais já não satisfazem as demandas contemporâneas dos arquivos, exigindo, assim, a adoção de tecnologias emergentes como os bancos de dados em grafos, que se destacam pela sua flexibilidade, desempenho superior e maior capacidade de interoperabilidade;

- 1 projeto destacou que a rigidez das estruturas e do modelo hierárquico da descrição multinível, replicados nesses sistemas, constitui uma importante limitação;

- 1 projeto salientou que os atuais sistemas não conseguem expressar adequadamente as complexas relações inerentes à descrição arquivística;

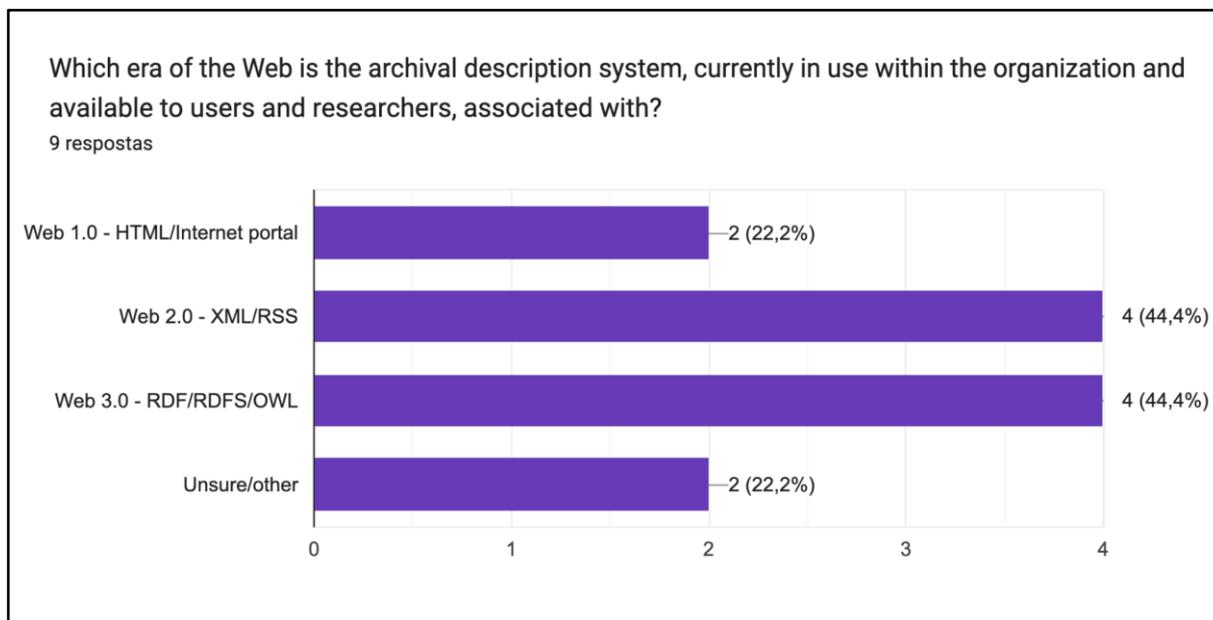
- 1 projeto apontou a heterogeneidade dos dados oriundos de diferentes instituições e coleções como um ponto preponderante;

- 1 projeto criticou os sistemas atuais por não fornecerem as funcionalidades necessárias para uma aplicação integral do *RiC*.

A análise da **questão 7** (fechada) mostrou dados relevantes sobre as tecnologias em uso pelas instituições que aplicam o modelo *RiC*. Observa-se que 44% das instituições já adotaram padrões associados à Web semântica (*Web 3.0*), como *RDF*, *RDFS* e *OWL*. Eles constituem uma mudança na gestão e no compartilhamento de dados. Paralelamente, outra parcela, equivalente a 44% das instituições, utiliza os formatos *XML/RSS*, característicos da *Web social* ou *Web 2.0*, com prioridade na colaboração e na interatividade on-line. Por fim, 22% das instituições ainda recorrem ao *HTML* e a portais *Web*, ferramentas típicas da *Web* de

documentos ou *Web 1.0*, indicando uma adesão a tecnologias mais tradicionais na apresentação e no acesso à informação.

Gráfico 5: Fases da Web nos sistemas de descrição das instituições que aplicam o modelo *RiC*.



Fonte: Elaborada pelo autor (2023).

Na **questão 8** (aberta), que tratava os critérios usados para a seleção de dados em projetos, as respostas das instituições demonstraram uma diversa gama de respostas. O Arquivo da Cidade de Amsterdam adotou a estratégia de incluir todos os metadados disponíveis na instituição. Em contrapartida, o projeto SoNAR da Alemanha optou por dados representativos nas áreas de patrimônio e cultura, incluindo bibliotecas, arquivos e museus, e se concentrou em registros de autoridades, além de explorar o conceito de *Big Data* em suas análises.

A Docuteam AG, da Suíça, adotou uma posição pragmática, sem estabelecer critérios específicos, preferindo a inclusão de todos os dados. Já o Arquivo Nacional da Noruega deu prioridade aos metadados de documentos nato-digitais para superar as tradicionais limitações dos sistemas descritivos. De forma similar, o Arquivo Nacional da Suécia baseou a sua escolha na abundância de determinados dados, como descrições de fundos, séries, volumes/dossiês, localidades e registros de autoridades.

O Serviço Interministerial de Arquivos da França selecionou dados com base em sua qualidade, privilegiando a precisão e a confiabilidade das informações. O projeto da Memoriav incluiu todos os dados do audiovisual suíço recebidos de instituições parceiras, aplicando ontologias específicas, como a *EBUCore*, para enriquecer as suas análises.

Por fim, o Arquivo Nacional de Liechtenstein concentrou-se em um conjunto específico de dados de documentos postais. As variadas estratégias de seleção de dados espelham diferentes necessidades, capacidades e objetivos das instituições que usam o *RiC*, ilustrando a diversidade de pontos de vista no campo da descrição arquivística.

Na **questão 9** (fechada), investigou-se as motivações que levaram diversas instituições a adotar o modelo *RiC* em seus projetos:

- 9 projetos (todos) endossaram a capacidade de interoperabilidade e de acesso integrado ao patrimônio de bibliotecas, arquivos, museus e monumentos culturais tombados. Esta escolha destaca o valor percebido no modelo *RiC* para conectar e tornar acessíveis diferentes legados culturais;

- 8 projetos apontaram a utilidade do modelo *RiC* na integração de dados fragmentados, ou seja, dados relacionados que estão isolados em diferentes sistemas de setores das instituições ou fora delas. Esta razão comprova a necessidade de superar a segmentação de informações em silos de dados para uma gestão mais eficiente e integrada;

- 7 projetos citaram três motivações: a descrição multidimensional e a representação dos dados em estruturas de grafos; as inovações das tecnologias da Web semântica e do *Linked Open Data*; e, por fim, o compartilhamento de dados dos arquivos para múltiplos usos na pesquisa acadêmica. As escolhas indicam um reconhecimento das capacidades avançadas do modelo *RiC* em termos de descrição, acesso e uso de dados;

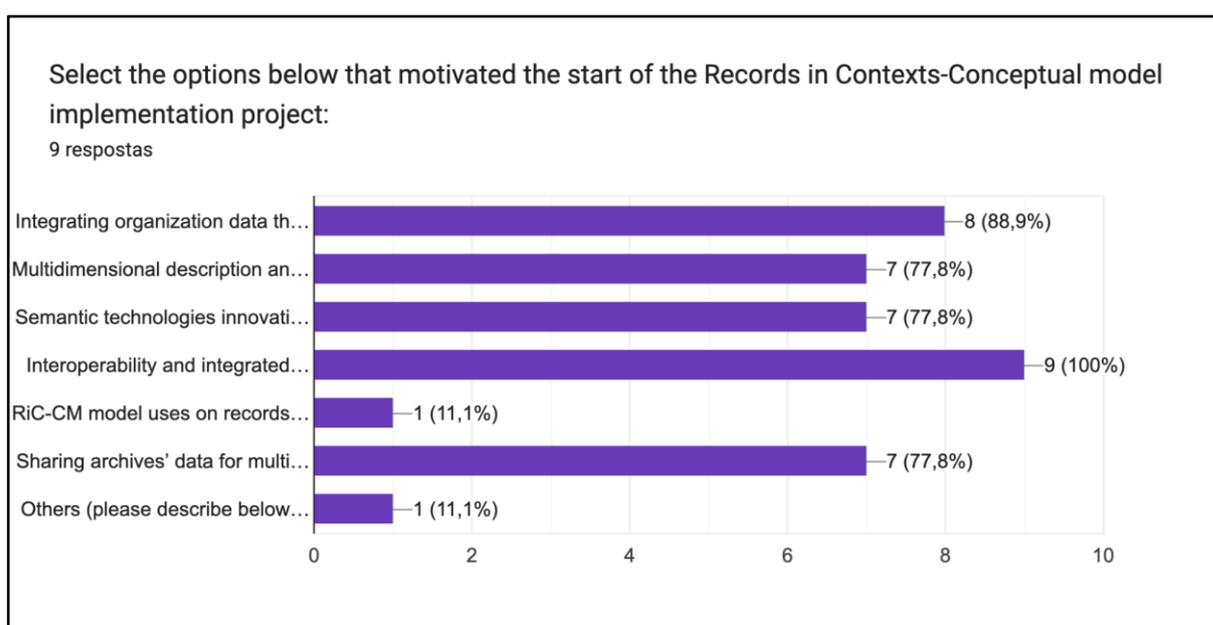
- 1 projeto foi motivado pela aplicação do modelo *RiC* na gestão de documentos, o que indica haver uma lacuna no uso da ontologia de arquivos em sistemas no início do ciclo de vida dos documentos, apesar da sugestão do ICA, prejudicando a arquivística integrada.

Juntos, os resultados mostram o interesse crescente na integração de dados culturais e patrimoniais em sistemas unificados, dando prioridade ao acesso e à interoperabilidade, com interesse crescente nas novas possibilidades de recuperação de dados em arquivos.

Na **questão 9** (tratado no formulário como **questão 10**), o representante do projeto SoNAR da Alemanha comentou que:

O RiC-O é um dos vários modelos de domínio expressos por uma ontologia; ele dá suporte à conversão do EAD tradicional em LOD (como GND-O, AgRelOn, ontologia RDA). Transformar dados em dados conectados com base nas ontologias de domínio de bibliotecas, arquivos e museus é uma pré-condição para o propósito do projeto SoNAR. Uma segunda pré-condição é que os dados originais devem ter números de identificação persistentes codificados de arquivos de autoridades nacionais/internacionais, para ambas as entidades identificadas com a descrição e as organizações de patrimônio cultural, particularmente o uso do Identificador Padrão Internacional para Bibliotecas e Organizações Relacionadas (Projeto SoNAR, 2023, tradução nossa).

Gráfico 6: Motivações das instituições para criação dos projetos de aplicação do modelo *RiC*.



Fonte: Elaborada pelo autor (2023).

A **questão 11** (fechada) avaliou se os benefícios esperados pelas instituições no uso do *RiC* se confirmaram depois da aplicação do modelo. Ao analisar as cinco motivações iniciais na escala de 1 (nada) a 5 (totalmente), verificam-se distintos graus de sucesso de cada uma.

No quesito da integração de dados organizacionais não conectados ou fragmentados em silos de dados, predominaram avaliações positivas em 7 projetos, com pontuações 5 (4 projetos) e 4 (3 projetos), mostrando o alto grau de satisfação. Isto é, a maior parte das instituições que lançaram mão do *RiC* obteve a integração de dados de modo bem-sucedido. Entretanto, percepções divergentes foram notadas, indicando que, em certos casos – 1 projeto com avaliação 3 e 1 projeto com avaliação 2 –, o modelo não atingiu a integração esperada, seja por características do próprio modelo ou por particularidades dos projetos.

Quanto às possibilidades da descrição multidimensional e da representação de dados em estruturas de grafos, não houve consenso. Alguns projetos relataram total satisfação, enquanto outros não obtiveram resultados positivos. A maioria das experiências foi favorável, com várias avaliações entre 4 (2 projetos) e 5 (1 projeto), mas também foram registradas avaliações moderadas e casos de baixa satisfação (3 projetos com avaliação 3, 1 projeto com avaliação 2 e 1 projeto com avaliação 1), possivelmente devido a dificuldades na aplicação ou às limitações do modelo *RiC*.

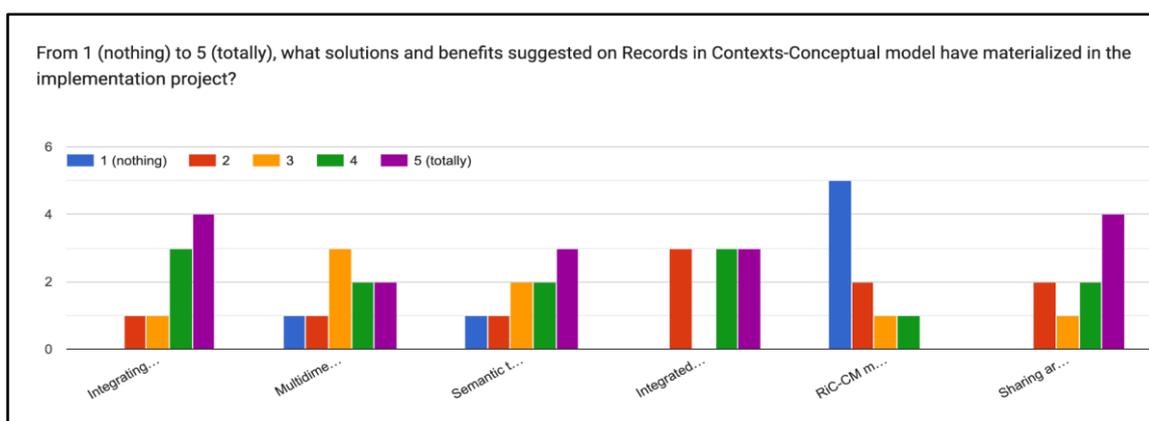
Em relação às inovações tecnológicas, como a Web semântica e o *Linked Open Data*, a tendência foi de alta satisfação, com a maioria das pontuações situando-se entre 4 e 5. No entanto, houve relatos de experiências moderadas e algumas avaliações mais baixas, retratando uma variedade nas experiências de aplicação do *RiC*.

A interoperabilidade e o acesso integrado aos dados do patrimônio cultural (bibliotecas, arquivos, museus e locais/monumentos culturais tombados) receberam avaliações majoritariamente positivas, com muitas pontuações máximas e notas 4, sugerindo a sua consolidação praticamente plena. Contudo, 3 projetos atribuíram nota 2, indicando que nem todos os participantes tiveram êxito neste tópico.

A aplicação do *RiC* na gestão de documentos teve uma recepção majoritariamente negativa, com pontuação 1 na maioria das respostas. Apenas 2 projetos atribuíram pontuação 2, e houve um caso isolado de sucesso, o projeto *Memoriav*, que conferiu pontuação 4.

No compartilhamento de dados de arquivos para múltiplos usos na pesquisa acadêmica, 4 instituições demonstraram alta satisfação, com várias atribuições da pontuação máxima; 3 projetos apresentaram notas 4 ou 3; e 2 projetos relataram um benefício limitado, com pontuação 2.

Gráfico 7: Benefícios obtidos pelos projetos de aplicação do modelo *RiC*.



Fonte: Elaborada pelo autor (2023).

Se calcularmos a média das respostas, obtém-se o *ranking* de soluções do modelo *RiC*, das mais às menos bem-sucedidas, de acordo com os projetos de implementação, como mostra o Quadro 18.

Quadro 18: Ranking de soluções do *RiC* bem-sucedidas nos projetos de aplicação do modelo.

Soluções do RiC	Média
Integração de dados organizacionais não conectados ou fragmentados (silos de dados).	4,11
Compartilhamento de dados de arquivos para múltiplos usos na pesquisa acadêmica.	3,89
Interoperabilidade e o acesso integrado ao patrimônio cultural (bibliotecas, arquivos, museus e locais/monumentos culturais tombados).	3,67
Inovações tecnológicas, como a Web semântica e o <i>Linked Open Data</i> .	3,56
Descrição multidimensional e representação de dados em estruturas de grafos.	3,33
Aplicação do RiC na gestão de documentos.	1,78

Fonte: Elaborado pelo autor (2023).

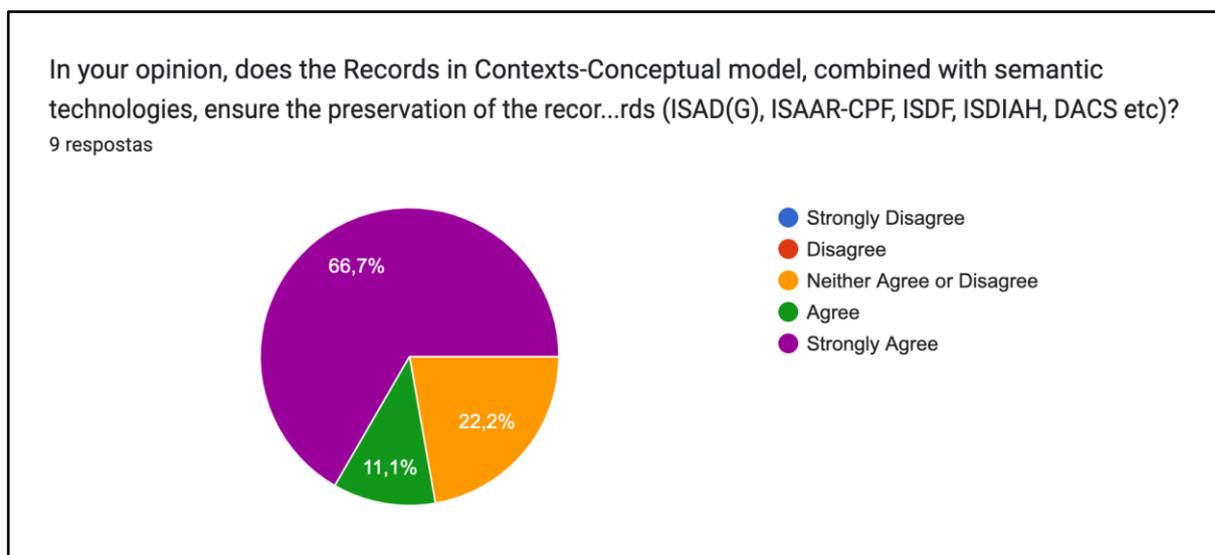
A **questão 12** (fechada) examina a capacidade do modelo *RiC*, aliado às tecnologias semânticas, de garantir a preservação dos contextos de produção, acumulação e uso dos documentos de arquivos de maneira mais precisa que as normas de descrição do ICA e as atuais tecnologias. As respostas coletadas revelam uma tendência amplamente positiva em relação ao modelo:

- 7 projetos expressaram alta concordância, portanto, nota-se a percepção das melhorias do *RiC* na descrição arquivística;

- 1 projeto indicou “concorda”, sinalizando que, embora haja um reconhecimento dos benefícios do *RiC*, ainda existem aspectos a serem debatidos ou aprimorados;

- 1 projeto registrou uma postura neutra – “não concorda, nem discorda” –, o que mostra incertezas ou a necessidade de mais informações para uma avaliação mais precisa sobre a eficácia do modelo *RiC* em comparação com as tradicionais normas de descrição.

Gráfico 8: Contexto arquivístico no modelo *RiC* mais representativo que nas normas de descrição do *ICA*.



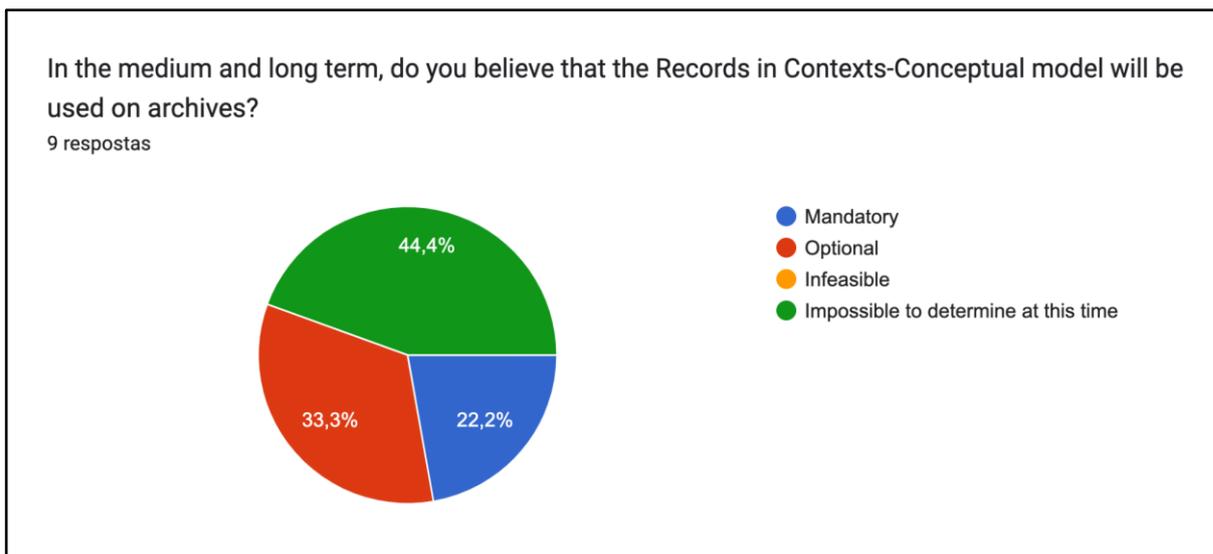
Fonte: Elaborada pelo autor (2023).

A **questão 13** (fechada) explora as opiniões das instituições sobre a futura consolidação do modelo *RiC* na comunidade arquivística, transparecendo uma clara divisão de opiniões:

- 4 projetos indicam incerteza, afirmando ser impossível determinar, no momento, se o *RiC* se consolidará como padrão na área;
- 3 projetos defendem que a aplicação do *RiC* deve ser opcional, conforme as necessidades específicas de cada organização;
- 2 projetos consideram que o *RiC* se tornará obrigatório, confiantes em seu potencial de inovação.

As respostas assinalam que, embora o valor do *RiC* seja reconhecido, há diferentes expectativas em relação à sua adoção na comunidade arquivística. É importante frisar que as dúvidas são de profissionais conhecedores do modelo em projetos de aplicação nas suas instituições.

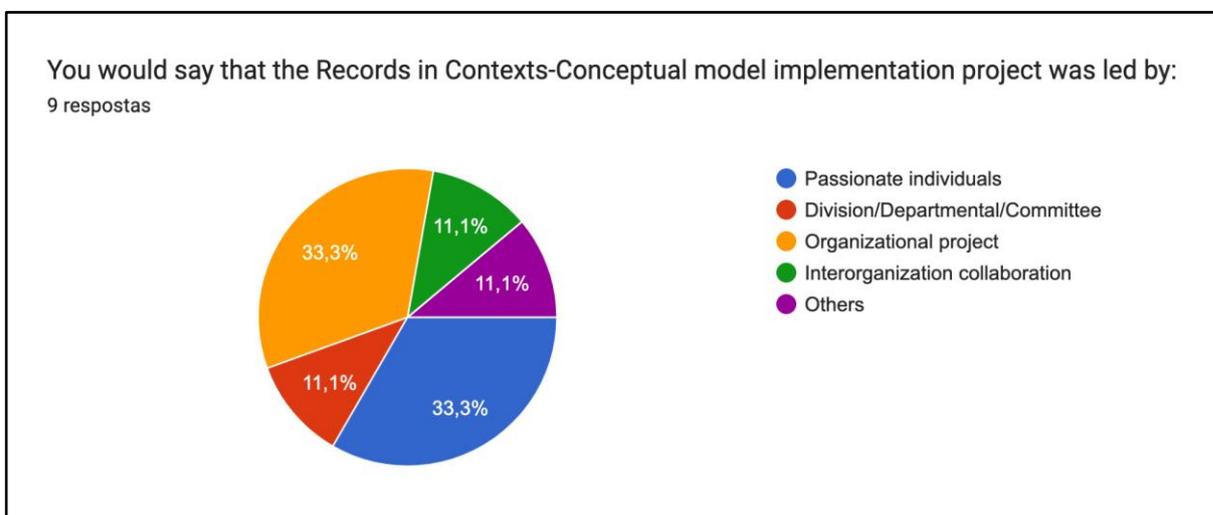
Gráfico 9: Estimativa de uso do modelo *RiC* pela comunidade arquivística.



Fonte: Elaborada pelo autor (2023).

A análise da **questão 14** (fechada) sobre como se deu a iniciativa e quem foram os patrocinadores dos projetos de aplicação do modelo *RiC* sinaliza variados níveis de formalidade, predominando projetos patrocinados pela hierarquia superior da instituição como forma mais citada. Por outro lado, foram também registradas as iniciativas individuais de profissionais e as parcerias entre instituições, bem como as iniciativas originadas na base da hierarquia (departamentos ou comitês). As diversas formas de patrocínio do uso do *RiC* mostram a complexidade e a natureza multifacetada destes projetos de aplicação do modelo.

Gráfico 10: Nível de institucionalidade de projetos de implementação do *RiC*.



Fonte: Elaborada pelo autor (2023).

A **questão 15** (aberta) se propôs a identificar as principais dificuldades na aplicação do modelo *RiC*. O Arquivo da Cidade de Amsterdam e o Arquivo Nacional da Noruega mencionaram aspectos técnicos relacionados ao desenvolvimento de *softwares* no padrão *Linked Open Data* para gestão de dados de arquivos e à adaptação de desenvolvedores de sistemas aos padrões *RDF/OWL* e *RiC-O*, respectivamente.

Quanto à aplicação da ontologia do *RiC* e à modelagem de dados, a Biblioteca Estatal de Berlim e a Fundação do Patrimônio Cultural da Prússia, do projeto SoNAR, relataram a complexidade de certas partes do *RiC-O*. Por sua vez, o Arquivo Nacional da Noruega assinalou incertezas de aspectos do modelo, como a aplicação do *RiC* para dados de arquivos híbridos (analógicos e nato-digitais), e alertou que a flexibilidade do modelo em permitir adições de classes pode levar à inconsistência na padronização, além das lacunas do modelo em relação ao conceito de meta-metadados.

O Arquivo Nacional da Suécia citou problemas relacionados à grande quantidade de dados legados a serem formatados para o uso do *RiC* e à mudança cultural nas práticas arquivísticas tradicionais. O Serviço Interministerial de Arquivos da França notou a pouca qualidade dos dados como um obstáculo, enquanto os Arquivos Cantonais de Vaud da Suíça fizeram alusão à escassez de soluções tecnológicas baseadas no *RiC* para descrição, preservação e difusão dos dados, como o aplicativo *docuteam cosmos*³⁵.

O projeto Memoriav salientou a complexidade de integrar o *RiC* ao modelo *CIDOC-CRM* dos museus. Já o Arquivo Nacional de Liechtenstein argumentou sobre as barreiras linguísticas e a necessidade de tradução do *RiC*, além da falta de conhecimento técnico básico em tecnologias semânticas entre os arquivistas. Por outro lado, a Docuteam AG considerou que não existem barreiras para o uso do modelo *RiC*.

Na **questão 16** (aberta), explorou-se a percepção dos participantes sobre a exequibilidade do modelo *RiC*. A empresa suíça Docuteam AG recomendou menos atenção nas tecnologias atuais, como o *RDF*, e mais no próprio modelo *RiC*, destacando a relevância de se pensar nos conceitos e na descrição arquivística em si.

Os Arquivos Cantonais de Vaud propuseram a harmonização do *RiC* com o padrão PREMIS, com o modelo CIDOC dos museus e com a norma *ISO TC46/SC11* – que abrange a produção e gestão de documentos e de arquivos, incluindo multimídias digitais. O Arquivo Nacional da Noruega manifestou confiança na operacionalidade do *RiC* dada a crescente

³⁵ Disponível em: <https://www.docuteam.ch/en/software-fur-archive/>.

familiarização da comunidade arquivística com o modelo e com as tecnologias relacionadas. O Arquivo Nacional da Suécia argumentou que a exequibilidade do *RiC* seria reforçada pela sua incorporação em sistemas de gestão de arquivos comercializados e de código aberto, com a complexidade do modelo oculta para os que trabalham com descrição arquivística. O projeto *Memoriav* vê a ratificação do *RiC* atrelada ao desenvolvimento de ferramentas de fácil utilização.

O Serviço Interministerial de Arquivos da França constatou que o *RiC-O* não pode ser a solução para todos os problemas e pode não ser suficientemente robusto para a gestão de dados, sugerindo que a continuidade do modelo estaria melhor associada às novas versões dos padrões EAC-CPF 2.0 e EAD4. O Arquivo Nacional de Liechtenstein sugeriu oferecer o padrão em mais idiomas, promover palestras e treinamentos para facilitar a sua compreensão, além de desenvolver sistemas com interfaces amigáveis para não especialistas. A Biblioteca Estatal de Berlim e a Fundação do Patrimônio Cultural da Prússia, do projeto *SoNAR*, considerou que a questão não se aplicava a eles, enquanto os Arquivos da Cidade de Amsterdam não responderam.

A análise das respostas dos questionários junto às instituições que aplicam o *RiC* forneceu valiosas informações para a pesquisa, ao destacar tanto as críticas à atual descrição arquivística quanto os benefícios e soluções potenciais do *RiC*. As vantagens do modelo são confirmadas não apenas teoricamente ou com base em experiências de outras comunidades com seus respectivos modelos para Web semântica, mas a partir de casos práticos bem-sucedidos na área dos arquivos.

5 CONSIDERAÇÕES FINAIS

No atual contexto do dataísmo, caracterizado pela centralidade do uso de dados em rede, observa-se uma importante transformação: ao longo dos últimos trinta anos, a *Web* de documentos evoluiu para uma *Web* de dados. A mudança impõe aos arquivos o desafio de se adaptarem, integrando-se a esta realidade tecnológica para manterem a sua relevância como instituições de acesso à informação. Diante das inovações os arquivos buscam integrar-se à *Web* 3.0 ao adotar ferramentas como a Web semântica, hoje, a mais moderna plataforma para a recuperação de dados. Por sua vez, as humanidades digitais e as pesquisas em arquivos têm se orientado para o aproveitamento de dados de descrição e de todas as ferramentas digitais em rede.

Nesta pesquisa, vimos teóricos destacarem a importância de estruturar os dados na *Web* 3.0, incluindo os dados dos arquivos, para evitar que sejam obscurecidos pelo vasto volume de informações no mundo digital. Para tal feito, as ontologias, a Web semântica e os Dados Abertos Conectados (*Linked Open Data*) são fundamentais, pois conferem estrutura e expressividade aos dados, ultrapassando as limitações das buscas convencionais em páginas. O Conselho Internacional de Arquivos apoia a estruturação de dados na lógica da Web semântica, sugerindo aos arquivistas o modelo *RiC* e a sua respectiva ontologia *RiC-O*, apesar de críticas e questionamentos da comunidade arquivística internacional.

As ontologias de domínio, como o *RiC*, são centrais para representar de forma precisa o mundo real dos arquivos, tanto para humanos quanto para computadores. Elas não só identificam entidades e suas características, como estabelecem relações entre si. Assim, o modelo *RiC* e sua ontologia contribuem para uma representação aprimorada dos arquivos na *Web*. O *RiC* também questiona os padrões tradicionais de descrição arquivística – como as normas de descrição –, inserindo-se em uma natural renovação científica e técnica. Outra vantagem do *RiC* é a possibilidade de integração de dados com bibliotecas e museus, por todos usarem padrões como o RDF e OWL em seus modelos.

Esta investigação identificou limitações nas práticas de descrição e nas tecnologias tradicionais utilizadas nos arquivos, como a descrição hierárquica e multinível; o entendimento convencional do princípio da proveniência; a redundância e falta de integração das normas do ICA; a fragilidade dos relacionamentos entre as unidades de descrição; a limitada inclusão de soluções para documentos digitais; a escassa conexão com dados externos; e a falta de abertura para parcerias com instituições do patrimônio documental e cultural. A análise de conteúdo

desta pesquisa permitiu reconhecer, segundo a visão do EGAD/ICA, as limitações da função da descrição, categorizadas em termos de isolamento, problemas de integração, inflexibilidade dos atuais padrões e dificuldades de adaptação às inovações tecnológicas.

A estudo também ressalta os benefícios do *RiC*, como uma ontologia unificada para o domínio arquivístico, a ampliação do princípio da proveniência, a prioridade da descrição multidimensional, a relevância dos relacionamentos entre entidades, ao lado do uso de formatos abertos e expressivos. As soluções propostas pelo *RiC* foram classificadas em termos de precisão na descrição arquivística, valorização dos contextos, integração, interoperabilidade, flexibilidade e fortalecimento do acesso e recuperação de dados.

Além do avanço tecnológico ocorrido desde o final do século XX, em especial dos novos recursos da *Web*, nota-se que muitos questionamentos e críticas às normas de descrição do ICA não se direcionam diretamente a elas, mas ao seu uso indevido ou por interpretações equivocadas por parte das instituições de arquivo. Por conseguinte, o *RiC* – que propõe superar os padrões obsoletos no contexto digital dos arquivos – também enfrentará problemas em termos de aplicação nos arquivos.

Os projetos demonstram inúmeras qualidades do *RiC*, como indicam as respostas aos questionários enviados às instituições que testaram o modelo. Apesar de estar em fase inicial, o *RiC* mostrou-se eficiente, com instituições satisfeitas com as possibilidades de tratamento e uso dos dados arquivísticos. No entanto, como visto nas respostas, ainda há uma forte resistência institucional ao uso do *RiC*. As ferramentas tecnológicas desenvolvidas para o *RiC* são mínimas e não comercializadas a baixo custo. Além disso, é necessário conhecimento especializado em tecnologias *Web* que os arquivistas ainda não dominam, e nem consta na grade da maioria dos cursos de arquivologia. Mesmo a flexibilidade proporcionada pelo *RiC* pode levar a uma certa desestruturação e despadronização dos dados na aplicação do modelo, se não for adequadamente controlada pelo ICA. Soma-se a isto a baixa qualidade de dados e o grande volume de dados legados – todos pontos de atenção citados pelos arquivos que testam o modelo *RiC*.

Ainda, como criticado quando da publicação das versões preliminares, o modelo *RiC* foi concebido numa realidade de países do norte, com pouca participação ou comentários dos arquivistas da América Central, América do Sul, África e Ásia. O mesmo se deu com as instituições que testaram o modelo e que responderam o questionário desta pesquisa, preponderantemente da Europa. Portanto, faltam pesquisas com aplicações do *RiC* e experiências com arquivos e *Web* semântica fora do eixo Europa-América do Norte. Os

resultados poderiam mostrar-se divergentes do ponto de vista dos países com pouco tratamento de arquivos ou com menos acesso à Internet.

De todo modo, o modelo *RiC* não só responde às demandas tecnológicas, mas também reafirma a visão do ICA sobre a descrição arquivística, consolidada em 1989. Na Declaração de Princípios de Descrição Arquivística da organização internacional, consta a intenção da comunidade arquivística em desenvolver uma descrição consistente, relevante e explícita, facilitando o acesso e a troca de informações sobre acervos arquivísticos e promovendo a integração das descrições de diferentes instituições em sistemas de informação unificados.

Chenard (2021) argumenta que a Web semântica e o modelo *RiC* representam uma evolução na descrição de fundos documentais, propondo uma mudança da leitura de inventários para a consulta de bancos de dados conectados. O arquivista reforça a necessidade de continuamente aprimorar a qualidade das descrições, preparando a transição para o *RiC* e a Web semântica:

Atomizar ainda um pouco mais a informação em múltiplos campos, segundo um vocabulário mais controlado, indexar em grande escala, registrar informações sistematicamente em certos campos até o momento um pouco esquecidos. Em síntese, preparar a mudança para o *RiC* e para a Web semântica é continuar realizando o que fazemos desde sempre: aperfeiçoar continuamente a qualidade de nossas descrições (Chenard, 2021, p. 27, tradução nossa).

No vasto universo de dados disponíveis na *Web*, destacam-se os da descrição arquivística. É primordial reconhecer a sua natureza predominantemente permanente, pois referem-se aos arquivos e aos documentos históricos preservados por períodos que variam de anos a séculos. Graças à sua permanência e relevância histórica, esses dados, que não serão eliminados das instituições custodiadoras, merecem o mais alto padrão de tratamento técnico, visto constituírem uma parte essencial do patrimônio arquivístico global.

Considerando os investimentos já feitos na preservação física desses documentos, torna-se imperativo alocar recursos adequados no seu tratamento e acesso, especialmente no atual contexto digital. A estruturação semântica dos dados, conforme proposta pelo modelo *RiC*, representa uma apropriada possibilidade. Esta metodologia não apenas valoriza o significado intrínseco dos dados arquivísticos, mas também assegura que o acesso a essas fontes vitais não seja obscurecido ou perdido no vasto labirinto de informações disponíveis na *Web*. Assim, a adoção da estruturação semântica de dados por intermédio do modelo *RiC* é um avanço ao qual os arquivistas tenderão a aderir para promover a recuperação e o uso eficiente dos dados do valioso patrimônio arquivístico mundial.

REFERÊNCIAS

- ALMEIDA, M. B. Uma abordagem integrada sobre ontologias: ciência da informação, ciência da computação e filosofia. **Perspect. Ciênc. Inf.**, Belo Horizonte, v. 19, n. 3, p. 242-258, jul./set. 2014. Disponível em: <http://dx.doi.org/10.1590/1981-5344/1736>. Acesso em: 3 dez. 2023.
- ARCHIVES NATIONALES. Service Interministériel des Archives de France. Bibliothèque Nationale de France. **PIAAF**: pilote d'interopérabilité pour les autorités archivistiques françaises: démonstrateur. [S. l.]: Logilab, [2017]. Disponível em: <https://piaaf.demo.logilab.fr/>. Acesso em: 3 dez. 2023.
- ARQUIVO NACIONAL (Brasil). **Dicionário brasileiro de terminologia arquivística**. Rio de Janeiro: Arquivo Nacional, 2005.
- BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 2021.
- BERNERS-LEE, T. Linked Data. Design Issues. **World Wide Web Consortium (W3C)**. jul. 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 3 dez. 2023.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. **Scientific American**, v. 284, n. 5, p. 28-37, may 2001. Disponível em: https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf. Acesso em: 3 dez. 2021.
- BRANDT, M. B. **Modelagem da informação legislativa**: arquitetura da informação para o processo legislativo brasileiro. 2020. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (Unesp), Marília, SP, 2020. Disponível em: <https://repositorio.unesp.br/handle/11449/191740>. Acesso em: 5 jan. 2024.
- BRASIL. Congresso Nacional. Câmara dos Deputados. Coordenação de Arquivo. **Arquivo Histórico**. 2023. Disponível em: <https://arquivohistorico.camara.leg.br/index.php/>. Acesso em: 3 dez. 2023.
- BRASIL. Congresso Nacional. Câmara dos Deputados. Coordenação de Arquivo. **Inventário analítico do arquivo da Assembleia Geral Constituinte e Legislativa do Império do Brasil, 1823**. 2. ed., rev. e reform. Brasília: Câmara dos Deputados, Edições Câmara, 2015. Disponível em: <https://arquivohistorico.camara.leg.br/index.php/assembleia-geral-constituente-e-legislativa-do-império-do-brasil-1823>. Acesso em: 3 dez. 2023.
- BURDICK, A.; DRUCKER, J.; LUNENFELD, P.; PRESNER, T.; SCHNAPP, J. Um breve guia para as humanidades digitais. **TECCOGS Revista Digital de Tecnologias Cognitivas**, [S.l.], n. 21, p. 69-98, 14 dez. 2020. Disponível em: <https://opus.lib.uts.edu.au/handle/10453/158973>. Acesso em: 20 maio 2023.
- CAMPOS, M. L. A. O papel das definições na pesquisa em ontologia. **Perspect. Ciênc. Inf.**, Belo Horizonte, v. 15, n. 1, p. 220-238, 2010. Disponível em: <https://www.scielo.br/j/pci/a/tJr4GnX9Xp7pj5pf44gK4yD/?lang=pt>. Acesso em: 3 dez. 2023.

CASTELLS, M. A revolução da tecnologia da informação. *In*: CASTELLS, M. **A era da informação: economia, sociedade e cultura: sociedade em rede**. 3 ed. São Paulo: Paz e Terra, 2000. p. 49-86.

CERVO, A. L.; BERVIAN, P. A.; SILVA, R. **Metodologia científica**. 5. ed. São Paulo: Prentice Hall, 2002. Disponível em: <https://doceru.com/doc/ne01xv0>. Acesso em: 2 jan. 2024.

CHENARD, G. Des points du vue multiples: RiC et moi. **Archivistes!** Paris, n. 136, janv./mars 2021, p. 27.

CLAVAUD, F. RiC aux Archives nationales de France: perspectives pour 2021. **Archivistes!** Paris, n. 136, janv./mars 2021, p. 24-25.

COOK, T. **O conceito de fundo arquivístico: teoria, descrição e proveniência na era pós-custodial**. Rio de Janeiro: Arquivo Nacional, 2017.

DATA CENTRIC MANIFESTO. 2023. Disponível em: <http://datacentricmanifesto.org/>. Acesso em: 3 dez. 2023.

FREITAS, C. G. Ciência de dados: Big Data e gestão dos dados. **Revista Científica Multidisciplinar Núcleo do Conhecimento**, ano 05, ed. 10, vol. 08, p. 134-140, out. 2020. Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/gestao-dos-dados>. Acesso em: 3 dez 2023.

GARTNER, INC. **Gartner says 70% of organizations will shift their focus from big to small and wide data by 2025**. Stamford: Gartner, Inc, 19 may 2021. Disponível em: <https://www.gartner.com/en/newsroom/press-releases/2021-05-19-gartner-says-70-percent-of-organizations-will-shift-their-focus-from-big-to-small-and-wide-data-by-2025>. Acesso em: 3 dez. 2023.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GUARINO, N. Understanding, building, and using ontologies. **International Journal of Human and Computer Studies**, [S. l.], v. 46, n. 2-3, p. 293-310, 1997.

GUEGUEN, G.; FONSECA, V. M. M.; PITTI, D. V.; GRIMOÛARD, C. S. Para um modelo conceitual internacional de descrição arquivística. **Acervo**, Rio de Janeiro, v. 26, n. 2, p. 100-116, 2013a. Disponível em: <https://revista.arquivonacional.gov.br/index.php/revistaacervo/article/view/517>. Acesso em: 3 dez. 2023.

GUEGUEN, G.; FONSECA, V. M. M.; PITTI, D. V.; GRIMOÛARD, C. S. Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives' Experts Group on Archival Description. **The American Archivist**, [S. l.], v. 76, n. 2, p. 566-583, Fall/Winter, 2013b. Disponível em: https://www.ica.org/sites/default/files/EGAD_English.pdf. Acesso em: 3 dez. 2023.

HACK, U. What's the real story behind the explosive growth of data? **Red Gate Software Blog**. 8 set. 2021. Disponível em: <https://www.red-gate.com/blog/database-development/whats-the-real-story-behind-the-explosive-growth-of-data>. Acesso em: 3 set. 2023.

HARARI, Y. N. A religião dos dados. In: HARARI, Y. N. **Homodeus**: uma breve história do amanhã. São Paulo: Companhia das Letras, 2016. p. 370-399.

HEATH, T., BIZER, C. **Linked data**: evolving the web into a global data space. Synthesis lectures on the semantic web: theory and technology. Kentfield: Morgan & Claypool, 2011. Disponível em <https://www.iro.umontreal.ca/~nie/IFT6255/Books/LinkedData.pdf>. Acesso em: 3 dez. 2023.

HERNÁNDEZ SAMPIERI, R.; FERNÁNDEZ COLLADO, C.; BAPTISTA LUCIO, M. P. **Metodologia de pesquisa**. 5. ed. Porto Alegre: Penso, 2013.

HOLZE, J. **DBpedia Snapshot 2022-12**. 2023. Disponível em: <https://www.dbpedia.org/blog/dbpedia-snapshot-2022-12-release/>. Acesso em: 3 dez. 2023.

INTERNATIONAL COUNCIL ON ARCHIVES. Committee on best practices and standards. **Progress report for revising and harmonizing ICA descriptive standards**. 2012. Disponível em: <https://www.ica.org/en/cbps-progress-report-revising-and-harmonising-ica-descriptive-standards>. Acesso em: 3 dez. 2023.

INTERNATIONAL COUNCIL ON ARCHIVES. Experts Group on Archival Description. **Introduction to Archival Description (RiC-IAD Consultation Draft v0.2)**. 2021. Disponível em: https://www.ica.org/sites/default/files/ric-cm-02_july2021_0.pdf Acesso em: 3 dez. 2023.

INTERNATIONAL COUNCIL ON ARCHIVES. Experts Group on Archival Description. **Records in Contexts Conceptual Model 1.0**. 2023. Disponível em: <https://www.ica.org/en/records-in-contexts-conceptual-model>. Acesso em: 3 dez. 2023.

INTERNATIONAL COUNCIL ON ARCHIVES. **ISAD(G)**: norma geral internacional de descrição arquivística. 2. ed. Rio de Janeiro: Arquivo Nacional, 2004.

INTERNATIONAL COUNCIL ON ARCHIVES. **Records in Contexts**: a conceptual model for archival description. Consultation draft v.1, Sept. 2016. Disponível em: <https://www.ica.org/sites/default/files/RiC-CM-0.1.pdf>. Acesso em: 3 dez 2023.

INTERPARES Trust. **InterPARES Trust response to EGAD-RiC**. 2016. Disponível em: <https://interparestrust.org/trust/article/inter pares-trust-response-to-egadric>. Acesso em: 9 dez. 2023.

ISOTANI, S.; BITTENCOURT, I. I. **Dados abertos conectados**. São Paulo: Novatec, 2015.

JACOBS, I.; WALSH, N. **Architecture of the World Wide Web**. W3C, 2004. Disponível em: <http://www.w3.org/TR/webarch/>. Acesso em: 3 dez. 2023.

- KHVOYNITSKAYA, S. **The future of big data: 5 predictions from experts for 2020-2025.** Itransition. 30 jan. 2020. Disponível em: <https://www.itransition.com/blog/the-future-of-big-data>. Acesso em: 3 dez. 2023.
- LAIR, J. S. Le web sémantique n'est qu'une forme de description pour échanger de la donnée. **Archivistes!** Paris, n. 136, janv./mars 2021, p. 28-29.
- LEROI, M. V. La stratégie sémantique du Ministère de la Culture. **Archivistes!** Paris, n. 136, janv./mars 2021, p. 22.
- LISBOA FILHO, J.; IOCHPE, C. Um estudo sobre modelos conceituais de dados para projeto de banco de dados geográficos. **Revista Informática Pública**, Belo Horizonte, ano 1, n. 2, dez. 1999, p. 67-90. Disponível em: http://pbh.gov.br/informaticapublica/ANO1_N2_PDF/ip0102lisboafilho.pdf. Acesso em: 3 dez. 2023.
- LLANES-PADRÓN, D.; PASTOR-SÁNCHEZ, J.-A. Records in contexts: the road of archives to semantic interoperability. **Program: electronic library and information systems**, v. 51, n. 4, 2017, 387-405. Disponível em: <https://doi.org/10.1108/PROG-03-2017-0021>. Acesso em: 3 dez. 2023.
- MAIA, R. M. C. S.; ALVARENGA, L. Interconexões entre a Teoria da Classificação Facetada (TCF) de Ranganathan e o Modelo Entidade-Relacionamento (MER) de Peter Chen. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB)*, 14., 2013. Brasília. **Anais digitais [...]**. Brasília: Universidade de Brasília, Faculdade de Ciência da Informação, 2013. Disponível em: <https://brapci.inf.br/index.php/res/v/184424>. Acesso em: 3 dez. 2023.
- MARCONDES, C. H. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. **Perspect. ciênc. inf.**, Belo Horizonte, v. 21, n. 2, p. 61-83, jun. 2016. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362016000200061&lng=pt&nrm=iso. Acesso em: 3 dez. 2023.
- MARCONI, M. A.; LAKATOS, E. M. **Técnicas de pesquisa**. 6. ed. São Paulo: Atlas, 2007. Disponível em: https://www.academia.edu/40756711/Marconi_e_Lakatos_2007_cap_1_Técnicas_de_pesquisa. Acesso em: 3 jan. 2024.
- MARIZ, A. C. A. **A informação na internet: arquivos públicos brasileiros**. Niterói: FGV, 2012.
- MIRANDA, J. M. Records in Contexts (RiC): análise da sua aplicação em arquivos, à luz das tecnologias Linked Open Data (LOD). **Acervo**, Rio de Janeiro, v. 34, n. 3, p. 1-26, 2021. Disponível em: <https://revista.an.gov.br/index.php/revistaacervo/article/view/1745>. Acesso em: 3 dez. 2023.
- NAZÁRIO, C. E.; DIAS, C. C. Avaliação do Europeana Data Model para o enriquecimento semântico de objetos. **Múltiplos Olhares em Ciência da Informação**, Belo Horizonte, n. Especial, 2021. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/view/37189>. Acesso em: 3 dez. 2023.

NORONHA, G. F. Representação lógica do conhecimento para ontologias: um guia para formalização de definições em linguagem natural para lógica descritiva. **Fronteiras: Revista Catarinense de História**, Belo Horizonte, v. 1, n. 1, p. 19-50, set. 2021. Disponível em: <https://periodicos.ufmg.br/index.php/fronteiras-rc/issue/view/1703>. Acesso em: 3 dez. 2023.

ORTEGA Y GASSET, J. **Meditación de la técnica**: ensimismamiento y alteración. Madrid: Biblioteca Nueva, 2015.

PAULA, L. B.; DEUS, W. S.; BARBOSA, E. F. Análise de repositórios de REAs em relação ao uso dos padrões de Linked Open Data. *In*: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 31, 2020, Porto Alegre. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 312-321

PICKLER, M. E. V. Web semântica: ontologias como ferramentas de representação do conhecimento. **Perspect. Ciênc. Inf.**, Belo Horizonte, v. 12, n. 1, abr. 2007. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362007000100006&lng=en&nrm=iso. Acesso em: 10 maio 2022.

PITTI, D.; STOCKTING, B.; CLAVAUD, F. Records in Contexts (RiC): a standard for archival description. *In*: INTERNATIONAL COUNCIL ON ARCHIVES, 2016, Seoul. **Annals [...]**. Seoul: ICA Congress, 2016. Disponível em: <https://www.ica.org/en/records-in-contexts-ric-a-standard-for-archival-description-presentation-congress-2016>. Acesso em: 3 dez. 2023.

POUPEAU, G. Dix ans de web sémantique et maintenant? **Archivistes!** Paris, n. 136, janv./mars 2021, p. 19-33.

RESTIF, M. L'orientation web sémantique de FranceArchives. **Archivistes!** Paris, n. 136, janv./mars 2021, p. 19-33.

SANTAREM SEGUNDO, J. E.; CONEGLIAN, C. S. Web semântica e ontologias: um estudo sobre construção de axiomas e uso de inferências. **Informação & Informação**, Londrina, v. 21, p. 217-244, 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/26417>. Acesso em: 3 dez. 2023.

SHADBOLT, N.; BERNERS-LEE, T.; HALL, W. The Semantic Web Revisited. **IEEE Intelligent Systems**, [S.l.], v. 21, n. 3, p. 96-101, jan./fev. 2006. Disponível em: <https://ieeexplore.ieee.org/document/1637364>. Acesso em: 3 dez. 2023.

SHARMA, M. Difference between Web 1.0, Web 2.0, and Web 3.0. **Geeks for Geeks**, [27 jan. 2022]. Disponível em: <https://www.geeksforgeeks.org/web-1-0-web-2-0-and-web-3-0-with-their-difference/>. Acesso em: 3 dez. 2023.

SHAUGHNESSY, J. J.; ZECHMEISTER, E. B.; ZECHMEISTER, J. S. **Metodologia de pesquisa em Psicologia**. Porto Alegre: AMG Editora, 2012. Disponível em: <https://edisciplinas.usp.br/course/view.php?id=76157&lang=de>. Acesso em: 2 jan. 2024.

SOUZA, R. R.; ALVARENGA, L. A web semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 132-141, jan./abr. 2004. Disponível em: <https://www.scielo.br/j/ci/a/sp3XpmZhXw384H5Fw9H89YL/?format=pdf>. Acesso em: 18 out. 2023.

SPLUNK INC. **New industry research shows the volume and value of data increasing exponentially in the data age**. 2020a. Disponível em: https://www.splunk.com/en_us/newsroom/press-releases/2020/new-industry-research-shows-the-volume-and-value-of-data-increasing-exponentially-in-the-data-age.html. Acesso em: 3 dez. 2023.

SPLUNK INC. **The data age is here: are you ready?** 2020b. Disponível em: https://www.splunk.com/en_us/form/data-age/thanks.html. Acesso em: 3 dez. 2023.

THE DATA-CENTRIC MANIFESTO. [2023]. Disponível em: <http://datacentricmanifesto.org/>. Acesso em: 3 dez. 2023.

WOODIE, A. Big growth forecasted for Big Data. **Datanami**. San Diego: Tabor Communications Inc, 11 jan. 2022. Disponível em: <https://www.datanami.com/2022/01/11/big-growth-forecasted-for-big-data/>. Acesso em: 3 dez. 2023.

WORLD WIDE WEB CONSORTIUM. **RDF 1.1 concepts and abstract syntax**. 2014. Disponível em: <https://www.w3.org/TR/rdf11-concepts/>. Acesso em: 3 dez. 2023.

WORLD WIDE WEB CONSORTIUM. **RDF schema 1.2**. 2023. Disponível em: <https://www.w3.org/TR/rdf12-schema/>. Acesso em: 3 dez. 2023.

WORLD WIDE WEB CONSORTIUM. **SKOS Simple Knowledge Organization System**. 2024. Disponível em: <https://www.w3.org/2004/02/skos/>. Acesso em: 8 jan. 2024.

WORLD WIDE WEB CONSORTIUM. **Web Ontology Language (OWL)**. 2013. Disponível em: <https://www.w3.org/OWL/>. Acesso em: 3 dez. 2023.

APÊNDICE A:
MODELO DE QUESTIONÁRIO

Survey on Records in Contexts-Conceptual Model Application Projects

Dear Participant,

Firstly, I want to express my gratitude for taking the time to complete this survey. I am Thiago Borges, the creator of this survey. I work as an archivist at the Câmara dos Deputados (Brazil) and am currently pursuing a master's degree at Faculdade de Ciência da Informação, Universidade de Brasília. This survey is part of my research on the Records in Contexts-Conceptual (RiC-CM) model and its application projects. Please be assured that your personal data will remain confidential and will not be published. This survey consists of 16 questions. Your participation won't take much time. The primary aim of this survey is to enhance the understanding of the RiC model within the archival community, both in Portuguese-speaking countries and internationally. We are interested in exploring its structure, benefits, and solutions. If you have any questions or require clarification, please do not hesitate to contact me via email at tarborges@hotmail.com. Once again, I sincerely appreciate your participation in this survey. Thank you very much.

Best regards,

Thiago Borges

E-mail:

Informed Consent Form

The data resulting from your participation in the research will be under the custody of the researcher responsible for the study. Your participation is voluntary and without any remuneration or benefit. You are free to refuse to participate, withdraw your consent, or discontinue your participation at any time. Refusal to participate will not result in any penalty or loss of benefits.

If you have any questions regarding the research, you can contact me at +5561998182021 or via email at tarborges@hotmail.com .

The researcher (Thiago Borges) guarantees that the study's results will be returned to participants through the availability of the master's dissertation and may be subsequently published in the scientific community.

*Based on Termo de Consentimento Livre e Esclarecido/Universidade de Brasília - UnB

I agree.

General Information

Your Name:

Your Organization:

Name of the Records in Contexts-Conceptual model application project:

Names of the organization(s) that have participated on the project:

Type of organization(s):

Library

Archive

Museum

Other

Question 1 - Which is the current implementation stage of the Records in Contexts application project?

Planning

In Progress

Completed

Discontinued (if applicable)

Other (please specify):

Question 2 - Duration of implementation:

Less than 6 months

6 months to 1 year

- 1 to 2 years
- Over 2 years

Question 3 - Why did your organization decide to test the Records in Contexts-Conceptual model? Please describe.

Question 4 - In your opinion, what are the biggest description principles and practice problems currently applied in your organization (online finding aids resembling linear printed documents, self-contained and hierarchical description etc)? Please describe.

Question 5 - In your opinion, the archival description in your organization is currently based on an archival approach that considers the Principle of Provenance:

- Traditional view of the Principle of Provenance (which privileges accumulator context)
- Expanded view of the Principle of Provenance (which includes other persons, groups and context, like “uses contexts”)
- A combination of both

Question 6 - In your opinion, what are the most significant technological problems related to description systems in your organization that justify the Records in Contexts-Conceptual model implementation project? Please describe.

Question 7 - Which era of the Web is the archival description system, currently in use within the organization and available to users and researchers, associated with?

- Web 1.0 - HTML/Internet portal
- Web 2.0 - XML/RSS
- Web 3.0 - RDF/RDFS/OWL
- Unsure/other

Question 8 - On selecting the data for the Records in Contexts-Conceptual model implementation project, what were the general criteria used for selecting it? Please describe.

Question 9 - Select the options below that motivated the start of the Records in Contexts-Conceptual model implementation project:

- () Integrating organization data that are not connected (data silos)
- () Multidimensional description and representation of data on graph structures
- () Semantic technologies innovations and possibilities (Semantic Web and LinkedOpen Data)
- () Interoperability and integrated access to cultural heritage (libraries, archives,museums and curated cultural sites and monuments)
- () RiC-CM model uses on records management activities
- () Sharing archives' data for multiple uses on academic research
- () Others (please describe below on question 10)

Question 10 - Others (question 9):

Question 11 - From 1 (nothing) to 5 (totally), what solutions and benefits suggested onRecords in Contexts-Conceptual model have materialized in the implementation project?

Integrating organization data that are not connected (data silos)

(1) (2) (3) (4) (5)

Multidimensional description and representation of data on graph structures

(1) (2) (3) (4) (5)

Semantic technologies innovations and possibilities (Semantic Web and LinkedOpen Data)

(1) (2) (3) (4) (5)

Interoperability and integrated access to cultural heritage (libraries, archives,museums and curated cultural sites and monuments)

(1) (2) (3) (4) (5)

RiC-CM model uses on records management activities

(1) (2) (3) (4) (5)

Sharing archives' data for multiple uses on academic research

(1) (2) (3) (4) (5)

Question 12 - In your opinion, does the Records in Contexts-Conceptual model, combined with semantic technologies, ensure the preservation of the records' creation,accumulation and uses contexts more precisely and efficiently than the current description standards (ISAD(G), ISAAR-CPF, ISDF, ISDIAH, DACS etc)?

- Strongly Disagree
- Disagree
- Neither Agree or Disagree
- Agree
- Strongly Agree

Question 13 - In the medium and long term, do you believe that the Records in Contexts-Conceptual model will be used on archives?

- Mandatory
- Optional
- Infeasible
- Impossible to determine at this time

Question 14 - You would say that the Records in Contexts-Conceptual model implementation project was led by:

- Passionate individuals
- Division/Departmental/Committee
- Organizational project
- Interorganizational collaboration
- Others

Question 15 - In your opinion what were the biggest barriers and challenges to implementing the Records in Contexts-Conceptual model? Please describe.

Question 16 - What is your opinion about the sustainability of the Records in Contexts-Conceptual model implementation? What do you think would help you in making this work more sustainable?