

Charley Luz

Ontologia Digital Arquivística



BOOKESS

Ontologia Digital Arquivística

L979o

Luz, Charley dos Santos

Ontologia Digital Arquivística: A representação do contexto de arquivo nos sistemas informatizados e na Web / Charley dos Santos Luz. -- São Paulo,. Bookess Editora, 2018.

Recurso digital : il.

Formato: PDF

Requisitos do sistema: Adobe Acrobat Reader

Modo de acesso: World Wide Web

ISBN 978-85-448-0710-1

1. Ontologia. 2. Interoperabilidade. 3. Arquivística. 4. Descrição Arquivística - Padrões. 6. Metadados. 7. Informação Digital - Preservação. I. Luz, Charley dos Santos. II. Ontologia Digital Arquivística: interoperabilidade e preservação da informação arquivística em sistemas de arquivo e na Web.

CDD 020

“[...] Quem controla o passado, controla o futuro; quem controla o presente, controla o passado... quem controla o passado, controla o futuro. Quem controla o presente agora?! Agora testemunhe, está logo atrás da porta”.
George Orwell



Dedico à minha mãe.

LISTA DE QUADROS

Quadro 1 - Resumo das funções arquivísticas	33
Quadro 2 - Tipos diferentes de documentos digitais e eletrônicos	43
Quadro 3 - Características de um SIGAD	51
Quadro 4 - Conteúdos abordados no e-ARQ Brasil	54
Quadro 5 - Comparativo de um SIGAD e do e-ARQ	54
Quadro 6 - Características de um RDC-Arq	56
Quadro 7 - Componentes do Contexto de Proveniência	64
Quadro 8 - Públicos envolvidos no Modelo OAIS	66
Quadro 9 - Estrutura de recursos de linguagens da Web	75
Quadro 10 - Níveis Descritivos e Instrumentos de Pesquisa	87
Quadro 11 - Áreas Informacionais e Elementos Descritivos ISAD-G	90
Quadro 12 - Os Sistemas Ontológicos da Filosofia	105
Quadro 13 - Os usos de Ontologias na Ciência da Computação	106
Quadro 14 - Conceitos de Ontologia para a Ciência da Informação	107
Quadro 15 - Componentes de uma especificação de uma ontologia	107
Quadro 16 - Componentes de um Domínio de uma ontologia	109
Quadro 17 - Estrutura de especificação de uma ontologia	112
Quadro 18 - Comparativo Especificação Ontológica com Classificação funcional	116
Quadro 19 - Tipos de ontologia por granularidade ou especialização	116
Quadro 21 - Taxonomia de instrumentos KOS	119

LISTA DE FIGURAS

Figura 1 - Diagrama / Esquema do Artefato do Conhecimento Interoperável	71
Figura 2 - Parte da Estrutura de Ontologias de Open Links Data	80
Figura 3 - Encoded Archival Description Tag Library, Version 2002	98
Figura 4 - Ficha de inscrição da ontologia no catálogo online Linked Open Vocabularies (LOV)	113

LISTA DE ABREVIATURAS E SIGLAS

ALTO	Technical Metadata for Optical Character Recognition
AN	Arquivo Nacional
AAB	Associação dos Arquivistas Brasileiros
BAD	Associação portuguesa de Bibliotecários, Arquivistas e Documentalistas
CA	Ciência Arquivística
CI	Ciência da Informação
CIA	Conselho Internacional de Arquivos
CONARQ	Conselho Nacional de Arquivos
CTDE	Câmara Técnica de Documentos Eletrônicos
DCMI	Dublin Core Metadata Applications Initiative
DTD	Document type definition
e-ARQ	Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos
EAC-CPF	Encoded Archival Context - (Corporate bodies, Persons and Families)
EAD	Encoded Archival Description

ECM	Enterprise Content Management
GED	Gestão Eletrônica de Documentos
HTML	Hypertext Markup Language
InterPARES	International Research on Permanent Authentic Records in Electronic Systems
ISSAR-CPF	Norma Internacional de Registo de Autoridade Arquivística para Pessoas Coletivas, Pessoas Singulares e Famílias
ISAD-G	International Standard Archival Description - General
KOS	Knowledge Organization System
LOC	The Library of Congress
LOD	Linked Open Data
MADRAS	Metadata and Archival Description Registry and Analytical System
MADS	Metadata Authority Description Standard
MARC	Machine Readable Cataloging
METS	Metadata Encoding & Transmission Standard
NOBRADE	Norma Brasileira De Descrição Arquivística
OAD	Ontology for Archival Description
OAIS	Open Archival Information System
OWL	Web Ontology Language
PCD	Plano de Classificação Documental
PDI	Pacotes de Disseminação da Informação
PIA	Pacotes de Informação Arquivística
PREMIS	Preservation Metadata

PSI	Pacotes de Submissão de Informação
RC	Record Continuum
RDC	Repositório Digital Confiável
RDC-Arq	Repositórios Digitais Confiáveis Arquivísticos
RDF	Resource Description Framework
SAA	Society of American Archivists
SGML	Standard Generalized Markup Language
SIGAD	Sistema Informatizado de Gestão Arquivística de Documentos
SKOS	Simple Knowledge Organization System
TIC	Tecnologias da Informação e Comunicação
UNIRIO	Universidade do Rio de Janeiro
URI	Uniform Resource Identifier
W3C	World Wide Web Consortium
XML	eXtensible Markup Language

Índice

0. Desafios iniciais

0.1 Prólogo	11
0.2 Uma primeira motivação	14
0.3 Arquivos: caminhos interdisciplinares	15
0.4 Uma aplicação da representação e organização da informação	18
0.5 Arquivística: do documento à informação	20

PARTE I: ARQUIVÍSTICA, CONTEXTOS E DESAFIOS

1.1. Aspectos das redes e o Mundo Digital	24
1.2. Representação para a informação digital	27
1.3. Arquivística Contemporânea: interdisciplinaridades, objeto(s) e funções	29
1.4. A Ciência Arquivística e alguns conceitos fundamentais	36
1.5. Archival Bond: o Vínculo entre um Documento e Outro	42
1.6. Documento Arquivístico Digital, seus parecidos, sua preservação e tipologia	44
1.7. Preservação Digital e as tipologias	47

PARTE II: SISTEMAS DE INFORMAÇÃO ARQUIVÍSTICOS E TECNOLOGIA

2. Sistemas de Informação Arquivísticos	51
2.1. SIGAD: Sistemas Informatizado de Gestão Arquivística de Documentos	55
2.2. Repositórios Arquivísticos Digitais Confiáveis - RDC-Arq	62
2.3. O Modelo Record Continuum	66
2.4. Cadeia de Custódia e Contexto Arquivístico	71
2.5. O Modelo OAIS Utilizado Também para Preservar	76
2.6. A Inerência da Interoperabilidade na Web Semântica	80

2.7.	Da Taxonomia à Ontologia: Ideias em Camadas de Processamento	85
------	--	----

PARTE III: METADADOS E ONTOLOGIAS

3.	A Interoperabilidade Semântica por meio de Metadados	92
3.1.	Criação de Padrões de Interoperabilidade por meio da Descrição	97
3.2.	EAD para Descrição de Conteúdos e de Informação arquivística	112
3.3.	EAC para Contextos de Criação de Arquivos	117
3.4.	MADRAS: a Ligação entre Metadados e Semântica	119
3.5.	PREMIS: Encapsulando as Ontologias	121
3.6.	METS: padrão arquivístico de metadados de objetos descritivos	123
3.7.	Ontologias: Estruturando Domínios de Conhecimento	124
3.8.	Ontologias e Padrões Arquivísticos	132
3.9.	Ontologias Digitais Arquivísticas	138
3.10.	Assim sendo: juntar os pontos	145
	Posfaciar sobre ontologias	160
	REFERÊNCIAS	161

0. Desafios iniciais



0.1 Prólogo

Este livro é uma evolução de uma pesquisa que realizei há um bom tempo e objetivou a conquista de grau de Mestre em Ciência da Informação em 2016 na ECA-USP. Com a aplicação e observação de parte dos elementos aqui dispostos, foi possível trazer nesta edição uma visão mais estruturada, em três grandes blocos de textos originalmente publicados na dissertação de mestrado, e aqui adaptados para um livro.

Neste breve prólogo temos textos que ambientam a questão da representação da informação nas plataformas digitais e o impacto da informação registrada em dígitos binários para os arquivos e as instituições que preservam patrimônio documental.

Na primeira parte de textos, que chamo de PARTE I: ARQUIVÍSTICA, CONTEXTOS E DESAFIOS, os aspectos das redes e o

Mundo Digital e a Representação para a informação digital. Numa abordagem epistêmica, explana-se sobre a Arquivística Contemporânea e suas interdisciplinaridades, seu (s) objeto(s) de estudo principal e secundários e as funções arquivísticas clássicas. Ao revisar a Ciência Arquivística e alguns conceitos fundamentais, elege-se aqueles que são importantes para entender os desafios da representação de contextos arquivísticos por meio de ontologias. Daí surge o Archival Bond, que trata do Vínculo entre os Documentos, o Documento Arquivístico Digital e um pouco de sua preservação e tipologia e a Preservação Digital e as tipologias documentais.

No segundo bloco de textos do livro, PARTE II: SISTEMAS DE INFORMAÇÃO ARQUIVÍSTICOS E TECNOLOGIA, é onde revisitamos conceitos de Sistemas de Informação Arquivísticos, aborda-se o modelo SIGAD: Sistemas Informatizado de Gestão Arquivística de Documentos e os Repositórios Arquivísticos Digitais Confiáveis - RDC-Arq. O Modelo Record Continuum surge como comparativo a outros modelos internacionais. A Cadeia de Custódia e Contexto Arquivístico, tão importante na garantia da autenticidade documental e o Modelo OAIS servindo como padrão também para Preservar compõem esta parte. Em relação aos temas deste bloco, aborda-se, ainda, a inerência da Interoperabilidade na Web Semântica e o caminho da Taxonomia à Ontologia.

No último bloco temático, PARTE III: DA DESCRIÇÃO ÀS ONTOLOGIAS, desnuda-se a representação da informação, partindo do entendimento dos metadados EAD, utilizados para Descrição de Conteúdos e de Informação arquivística e o modelo EAC, utilizado para Contextos de Criação de Arquivos. Aborda-se, ainda o modelo MADRAS que propõe a Ligação entre Metadados e Semântica, e o PREMIS, que Encapsula as Ontologias. Outro modelo estudado é o METS, o padrão arquivístico de metadados de transmissão de objetos

descritivos. Estes modelos são a base para entender a Ontologia Digital Arquivística. As Ontologias Estruturam Domínios de Conhecimento e podem representar contextos de arquivos numa linha de tempo digital. Estes três blocos se embrincam e formam uma linha de conhecimento necessária para os novos tempos. Qual abordagem o profissional da informação e o arquivista devem ter a partir de seu papel de propor sistemas de informação e plataformas perenes?

O tempo para o digital é diferente do tempo humano. Nosso longo trabalho de garantir a autenticidade documental e a confiabilidade da informação é uma projeção para o futuro. Neste link que criamos – a possibilidade de que no futuro alguém possa acessar o que geramos de informação nos dias de hoje – devem estar depositadas toda a interdisciplinaridade necessária para garantir uma linha de custódia duradora. Nesse livro exercitamos este diálogo entre a ciência da informação, a tecnologia da informação e a ciência arquivística de forma a contribuir para novas visões sobre o impacto da tecnologia na humanidade ao longo do tempo, contribuindo para que entendam o que geramos hoje de informação.

Charlley Luz

0.2 Uma primeira motivação

Parte da memória humana está sendo registrada em meio digital no mínimo pelos últimos 40 anos. Portanto, podemos considerar esta Era como a do primitivismo digital. Tudo que fazemos agora, em relação às opções tecnológicas de estruturação, terá um impacto nos documentos que serão acessados no futuro.

Porém, parte dessa história corre o perigo de se perder em sequências de bits, arranjados em bytes sem leitura no futuro. É resultado da obsolescência tecnológica e falta de tratamento das informações digitais. Necessitamos de técnicas de preservação digital que poderão ajudar nossos descendentes a entenderem os dias de hoje e acessarem essas informações.

Este livro foi resultado de uma pesquisa que foi ampliada para o mestrado em Ciência da Informação da Escola de Comunicação e Artes da Universidade de São Paulo, onde buscou-se estudar o uso de ontologias em sistemas de arquivo e na web como forma de representar a informação arquivística, ou seja do contexto do arquivo, visando sua preservação a longo prazo. Esta motivação maior foi um movimento no sentido de levar o tema arquivos contemporâneos para a ECA-USP, aproximando assim a missão de Organização e Representação do Conhecimento, tradicional no departamento, com a realidade do mundo das informações digitais.

O campo científico da arquivística também deve se abrir para as relações intradisciplinares, onde a representação do conhecimento, área primordial da Ciência da Informação (CI) age sobre os arquivos, empurrando seus cientistas a analisarem as

relações dos arquivos com a área da organização do conhecimento. As ontologias, uma frente de estudos dentro da CI, já mostra para a área de arquivo um nova realidade, mais estruturada e que usa de forma informação natodigital.

Assim, as ontologias estão condicionadas à atividade de representação da informação, conhecida como representação descritiva, que é basicamente representar o conteúdo de um recurso de informação por meio de palavras-chave, organizá-lo e relacioná-lo aos demais. Neste caso é criar informação sobre a informação, tornando-o um recurso encontrável. As ontologias trazem novas formas de fazer isso, ao incluir dimensões e camadas na representação de relações complexas de um domínio de conhecimento.

0.3 Arquivos: caminhos interdisciplinares

A informação é um fenômeno humano mediado por linguagem e, hoje, parte deste processo está digitalizado e mediado por sistemas e softwares. Logo, as linguagens devem ser entendidas por agentes computacionais. A informação institucionalizada, orgânica, gerada em sistemas e arquivos só é fidedigna e autêntica se estruturada em ambientes de confiança.

Parte desta informação precisa evoluir e migrar durante o tempo, contudo, esta evolução deve carregar todos os dados acerca dos documentos e de seu contexto para garantir a autenticidade, o que poderia ocorrer por meio de uma linguagem em comum entre homem e máquina e neste ponto as ontologias podem contribuir para a preservação destas informações.

A fidedignidade e autenticidade do documento arquivístico é garantida pela credibilidade de sua cadeia de custódia por meio de sistemas de informação de arquivo confiáveis, tanto na gestão como na guarda permanente. Tal tarefa requer que as informações geradas no documento e de seu contexto de criação sejam preservadas, desde a gênese documental por meio de diferentes técnicas de empacotamento e identificação dos conjuntos documentais.

O possível uso de ontologias como forma de mediação entre agentes computacionais e agentes documentais e seu possível uso para a preservação da informação arquivística e dos documentos em longo prazo foram construídos a partir de pesquisa que dialogou com a tecnologia da informação, a ciência arquivística e a ciência da informação.

Nesse livro vamos conhecer as ontologias digitais arquivísticas, geradas por meio dos padrões de descrição arquivísticas e de manutenção do contexto dos arquivos ao longo do tempo nos diferentes tipos de sistemas de informação de arquivo e na rede mundial. O uso de ontologias em repositórios confiáveis pode acompanhar a manutenção da cadeia de custódia de documentos e trazer dados do contexto de criação, garantindo a preservação, acesso e a confiabilidade documental.

Em relação à Tecnologia da Informação, abordamos a Web 3.0 e o uso de ontologias aplicadas. Em relação ao campo informacional, tratamos tanto da organização e representação e mediação da informação de acervos documentais (Ciência da Informação), como o compartilhamento e preservação da Informação Arquivística em sistemas de arquivo (dentro do campo específico da Arquivologia) e a identificação desta na rede.

Os documentos informáticos, também identificados como informação digital, precisam passar por processos que garantam sua padronização para facilitar as operações de busca e revocação em interfaces de plataformas na rede mundial e em sistemas informatizados.

Com o advento da web 3.0 é necessário criar links semânticos que façam sentido aos agentes computacionais, o que gera uma necessidade de trabalho, tanto teórico como prático, para a área da Ciência da Informação e seus profissionais. É uma nova prática de tratamento da informação que precisa ser exercitada por profissionais da informação: a preparação da interoperabilidade dos objetos informacionais, dos registros e recursos online.

Na área da Arquivologia, também conhecida como uma das Ciências da Informação, possui *corpus* científico definido e tem como objeto o documento arquivístico (e tudo relativo à sua geração, uso e guarda). As tentativas de padronização de dados são recentes, remontando as primeiras articulações à década de 1980, culminando com a publicação da primeira Norma internacional da área, a ISAD-G (*International Standard Archival Description - General*), em 1994.

Tais iniciativas sempre buscaram organizar a informação arquivística e sua representação descritiva para acesso e uso. Logo, a informação arquivística é aquela disponível num arquivo organizado como um sistema, seja informatizado ou não, que representa o conteúdo e a descrição dos arquivos. Este livro mostra a existência das ontologias digitais de cunho arquivístico, geradas por meio de padronização de metadados e utilizadas tanto para a difusão como para a preservação de informação geradas em sistemas de arquivo e disponíveis na internet. Lopes (2009, p. 362) já prospectava esta ideia em seu livro intitulado “Nova Arquivística”, quando afirma que

competem aos profissionais da informação analisar o fenômeno da informática no aspecto da “necessidade de adoção de normas e padrões de trabalho arquivístico intercambiáveis internacionalmente, para que sejam compatíveis e de possível integração no ciber mundo”.

Neste livro mostramos como, no contexto tecnológico digital, como a relação dos padrões de descrição arquivísticos, representados por meio de metadados, pode-se estabelecer a criação de ontologias leves a serem utilizados na interoperabilidade e na preservação da informação arquivística, tanto em sistemas informatizados de arquivos quanto na Web. Assim, buscamos formas de identificar a informação arquivística em sistemas documentais, além de sua relação com a Web 3.0.

Ao estabelecer como um dos resultados do processo arquivístico digital as ontologias, este processo pode ser aprimorado, planejando desde sua concepção e pode melhorar a interoperabilidade de seu produto. Por isso este livro pode também indicar boas práticas para quando houverem iniciativas de interoperabilidade de dados arquivísticos brasileiros.

0.4 Uma aplicação da representação e organização da informação

A informação digital resultado da informatização trouxe para os arquivos a realidade dos sistemas de gestão documental tão utilizados em grandes corporações e na estrutura estatal por meio de plataformas de uso de informações. Afinal, “vive-se em uma nova etapa, em que a presença da informação registrada mensurável em bytes, tem uma significação muito maior do que se possa, a priori, imaginar” (LOPES, 2009, p. 359). Os sistemas informatizados de

gestão e guarda documental geram dados sobre seu acervo, criando metadados que tornam possíveis seu gerenciamento.

Nessa realidade, parte da informação que transita pela rede, originada em sistemas digitais, possui características arquivísticas, pois, conforme afirma Lopes (2009, p. 360) “os equipamentos de informática são usados rotineiramente como meios de produzir, acumular e transmitir informações arquivísticas e não-arquivísticas”. Ainda, de acordo com o autor, parte dessa documentação é informação de arquivo, pois muitas dessas informações “se cristalizam como documentos informáticos, não existindo em outros suportes” (LOPES, 2009, p. 360). Por isso, há a necessidade de se identificar informação disponibilizada na rede com potencial arquivístico.

Os recursos informacionais estruturados são resultantes dos processos de organização da informação, ou seja, as informações são rotuladas, combinadas e disponibilizadas em interfaces dinâmicas por meio de links utilizando, sobretudo, o modelo conhecido como *Open Linked Data*. Este modelo, incentivado por um movimento autônomo internacional, conecta dados relacionados e que não estavam vinculados entre si e reduz as barreiras de conexão de dados ligados por meio de outros métodos, como por exemplo, vocabulários controlados que são de estrutura sintática.

A Web Semântica, ou Web 3.0 é estruturada numa rede de computadores via internet, com informações ligadas por meio de semântica estruturada em ontologias. Utiliza-se uma série de recursos informacionais e tecnológicos de forma que é possível garantir que computadores entendam o conteúdo e ofereçam aos usuários respostas significantes às questões e inferências de buscas realizadas.

Essa representação do conhecimento também ocorre na informação arquivística (tanto na classificação, como no arranjo e descrição arquivística). Os serviços de informação de arquivos, por custodiarem a informação orgânica e representar a estrutura funcional das instituições, necessitam ter metadados para representar fidedignamente a informação contida, mantendo a autenticidade por meio da manutenção de cadeia de custódia.

Para Alves (2010), a catalogação nasceu da necessidade de estabelecer regras para a construção de catálogos, porém, foi sendo desenvolvida e aprimorada, tornando-se uma metodologia para processamento e tratamento descritivo e temático da informação. Ao longo do tempo, utilizou-se das tecnologias disponíveis em cada época como forma de aprimorar o processo de representação, com o intuito de facilitar a recuperação e disseminação dos recursos informacionais.

0.5 Arquivística: do documento à informação

É necessário destacar o que torna a informação arquivística especial: a possibilidade de estruturar a informação em sua gênese, ou seja, criando visões sobre a informação que se possui em acervo, desde seu gerador/acumulador até as unidades documentais. Ou, como afirma Silva (2012), “a função dos arquivos estaria na perspectiva das possíveis reutilizações da informação gerada e estruturada por processos de trabalho, que lhes impõem uma interpretação contextual”. Logo, entende-se que a informação arquivística possui formas de expressar o conjunto da informação orgânica contextual, mas trata-se de uma representação desta que estará custodiada num ambiente seguro.

A informação arquivística é abordada como toda informação gerada pelas instituições que disponibilizam arquivos e dados sobre os fundos documentais que administram, possuem, descrevem e disponibilizam. Ela ainda traz informações de contexto dos fundos (as respectivas estruturas de organização deste acervo, ou seja, a instituição, as seções e séries documentais) que são contidas em instrumentos descritivos como guias, em perfis de metadados ou em *links* semânticos. Afinal, é premente o uso da rede mundial de computadores pela internet para divulgar acervos e em especial o “uso deste meio para a difusão de informações, em especial a descrição arquivística” (LOPES, 2009, p. 362).

Há iniciativas internacionais de normalização de dados, de tecnologias e normas que indicam caminhos de padronização de dados para a área dos arquivos, aplicável em documentos orgânicos e em plataformas digitais. Portanto, para isso que se destinou o presente estudo: indicar como o produto da descrição arquivística (que possui iniciativas internacionais de uniformização) pode ser utilizado como padrão de interoperabilidade de dados de acervos de arquivos e de informação arquivística, sendo aplicado por meio de ontologias, utilizando os perfis de metadados existentes, tal como o *Encoded Archival Description* (EAD) e outros analisados.

A necessidade de interoperabilizar a informação é básica para os tempos de comunicação em rede. A interoperabilidade garante o uso e a encontrabilidade dos metadados estruturados dos objetos informacionais. As formas de estruturar a informação arquivística, isto é, aqueles referentes aos acervos documentais disponíveis, não é obrigatoriamente padronizada. Isso representa que cada serviço informacional, ou unidade de tratamento documental (um serviço de arquivo, por exemplo), pode definir a forma como vai estruturar este tipo de informação, gerando suas

específicas políticas de interoperabilidade e de descrição, além do seu próprio padrão de metadados.

Identificamos neste livro a Ontologia Digital Arquivística, que é um produto de linguagem, da descrição de um contexto, de uma estrutura e de um registro ao longo de um tempo, realizado num ambiente digital por meio de plataformas tecnológicas utilizando aplicações de ontologias leves para tal. Isto pode auxiliar no processo de perenizar o contexto, importante na preservação da informação digital e para validar a cadeia de custódia, o que é primordial na garantia da autenticidade documental, missão essencial do arquivista.

A descrição arquivística como base para ontologias foi um dos pilares dessa investigação, e analisou-se até que ponto devemos considerar os padrões da área - aplicados pela *International Standard Archival Description - General* (ISAD-G) ou a Norma Brasileira de Descrição Arquivística (NOBRADE) – como uma fonte de organização e representação da informação em sistemas e na rede e, neste caso, passando a configurar uma ontologia digital arquivística.

Este livro também trata de reflexão sobre a padronização da informação digital, que visa a interoperabilidade, tanto da informação arquivística, como sua preservação. Por isso o contexto trata tanto da Web Semântica como os sistemas informatizados de arquivo, abordando a questão da Representação, principalmente em como a descrição arquivística pode ser considerada uma ontologia e em que nível.

PARTE I: ARQUIVÍSTICA, CONTEXTOS E DESAFIOS

Nesta primeira parte, parte-se dos novos desafios para um mundo digital e que explicita a motivação que levou a escrever este livro e ao desafio da informação digital. Tal abordagem possibilita o vislumbre de um enfoque contemporâneo da arquivística.

Para tanto, é preciso elucidar alguns conceitos importantes como Arquivo, documentos e organicidade; analisa-se o *Archival bond*, que é o vínculo entre um documento e outro e, por fim, é necessário explanar o conceito de documento arquivístico digital para assumi-lo como foco da Ciência Arquivística e elemento de análise deste trabalho.

1.1. Aspectos das redes e o Mundo Digital

Propiciadas pelo advento da rede mundial de computadores baseada na internet surgem as redes de colaboração com base em informação digital, que são caracterizadas por conjuntos de unidades ou serviços de informação voltados para um interesse comum, como a troca de conhecimento, materiais ou serviços. Afinal, como observou Tomaél (2005, p. 3), as “redes de informação reúnem pessoas e organizações para intercâmbio de informações, ao mesmo tempo em que contribuem para a organização de produtos e a operacionalização de serviços que sem a participação mútua, não seriam possíveis”. Deve-se considerar que o funcionamento básico dessa rede se dá por meio de acordos de cooperação e adoção de normas comuns para a padronização de seu conteúdo.

Com base nesse aspecto, uma rede é a intersecção de contatos e inteligências, e seu resultado é a inteligência coletiva (LÉVY, 2003), que interage em estruturas que dão acesso à informação, sendo possível controlar e organizar informação e tratar a difusão do conhecimento. O modelo tecnológico disseminado para essas redes é o *Open Data*, que é a estrutura original da Web Semântica, com seus serviços de mapeamento e coleta de dados (chamado de *harvesting*) e de distribuição de informação por meio de portais públicos.

Os serviços de informação e seus sistemas, portanto, estão inseridos em redes, o que requer a adoção de padrões que possibilitem importar e exportar seus dados em caráter global. O *Dublin Core* e o MARC, por exemplo, são formatos sugeridos para

esta padronização em relação a dados gerais e o EAD e EAC para dados arquivísticos.

As informações que antes eram encontradas em livros e documentos, atualmente compõem complicadas redes dispostas em plataformas digitais, como a internet. Tratam-se de funcionalidades, recursos, conteúdos colaborativos, gestão de documentos e bibliotecas virtuais, entre outros sistemas, que disponibilizam a informação em tempo real (frizando também a importância dos recursos Web 2.0 de participação e compartilhamento). Mesmo contando com todo esse aparato digital, porém, o processo de organização da informação é extremamente necessário.

Para Café e Sales (2010) a organização da informação é necessária para descrever e estruturar o conteúdo, uma vez que “organizamos um acervo para compreendê-lo melhor e assim podermos recuperar objetos informacionais, ou seja, informações registradas nos mais variados suportes (textos, imagens, registros sonoros, representações cartográficas e páginas Web)”. Ao acrescentar as páginas Web ou as plataformas digitais no rol de suportes, admite-se que a interface e os metadados são instrumentos para a gestão desses recursos. Além do mais, é preciso verificar como a descrição arquivística pode servir de instrumento para padronização de informação na Web 3.0.

A informação digital, como forma estruturante do conhecimento contemporâneo e para ser tratada por agentes computacionais, precisa ser organizada em suas principais características, tais como os temas (taxonomias) ou a semântica (ontologias). Ao definir os processos comunicativos e de navegação de sistemas e linguagens, objetiva-se que, ao utilizar uma plataforma digital de sistema de informação de arquivo, isto leve o usuário a um

novo conhecimento, estabelecendo um *locus* de mediação entre instituições e pessoas. Afinal, informação não usada é apenas dado.

Quanto ao aspecto temático e de significações, iniciativas com dados abertos ou públicos já operam, criando redes semânticas de informações abertas. De forma ampla, a aplicação do esforço em interoperabilizar os dados também necessitam de processos de organização específicos e de definições, como a criação de ontologias e a aplicação de descrições, atividades já partilhadas por museus, bibliotecas, arquivos, instituições e pessoas no tratamento da informação em ambientes digitais.

No ponto de vista local, focando nas instituições que intermediam conteúdo, as instituições coletoras de cultura no conceito de Smith (2012), aspecto em comum das “3 irmãs” da Ciência da Informação: arquivos, bibliotecas e museus - a organização de informações passa a ter um papel decisivo no acesso realizado diretamente pelos usuários, por meio de plataformas digitais e portais, interna ou externamente.

Aos arquivistas significa, segundo Quisbert (2006), devido à sua responsabilidade de guarda e preservação, dar acesso a informações, por vezes usando diferentes gerações de softwares. Todo software é substituível e atualizado, mas a informação deve permanecer original. No caso de arquivos os documentos arquivísticos são preservados da mesma forma: por meio de diferentes gerações de softwares. Os documentos devem permanecer como originais, a fim de cumprir suas exigências como prova.

Nesta perspectiva, é necessário definir algumas das principais funções que envolvem a organização dos objetos

informacionais, quer seja identificar a existência de todos tipos de recursos informacionais, assim que estiverem disponíveis, ou agrupar sistematicamente recursos informacionais em coleções (de bibliotecas, arquivos, museus, e ambientes digitais da internet), além de produzir listas desses recursos informacionais, oferecendo metadados úteis de identificação.

1.2. Representação para a informação digital

Parte do processo de organização da informação digital nada mais é do que evidenciar a essência informacional de cada registro, aplicando técnicas de catalogação (representação) e estabelecendo o processo de descrição e de indexação por meio de metadados e, assim, elegendo este recurso para preservação. Esse tratamento da informação é objeto de trabalho dos profissionais da informação, incorporando, assim, estas duas dimensões no processo (catalogação e descrição).

Neste sentido, considera-se as duas dimensões da essência informacional, de acordo com Café e Sales (2010), como sendo “a dimensão descritiva, voltada aos elementos relativos à forma dos documentos (como na catalogação descritiva) e a dimensão temática, voltada aos conteúdos informacionais (como na catalogação de assuntos, na classificação, na indexação e na análise documental”.

A informação digital alterou substancialmente a relação dos indivíduos com a busca de conhecimento e acesso à informação, tornando-o um ser seletivo. A mediação é realizada diretamente pelos usuários, exigindo esquemas intuitivos de classificação e de

recursos de navegação em interfaces. Além disso, a informação arquivística ainda terá uma classificação conforme a função e atividade que a gerou, ou seja, haverá nesta estrutura informações de seu contexto. Como afirma Ribeiro (2005, p. 10), “com a internet criaram-se possibilidades de uma consulta à distância em frações de segundo sobre volumes incomensuráveis de informação, o que alterou, radicalmente, os comportamentos de pesquisa, os perfis dos utilizadores e as necessidades de informação, agora à escala planetária”.

A própria gênese documental, para o arquivo, foi alterada por conta da realidade digital. Se antes as burocracias e as estruturas corporativas realizavam seu processo baseado em fluxos documentais registrados em papel, a nova realidade coloca para o mundo da gestão documental os workflows, fluxos estruturados em software em plataformas digitais, que se tornam a ferramenta ideal. Como destaca Innarelli (2015, p. 114), *Workflow* é um sistema informatizado que automatiza fluxos de trabalho nas instituições, ferramenta que produz e gerencia documentos digitais. Os *workflows* são estruturados em interfaces que fornecem uma decisão ou tarefa ao usuário, este passo do processo então é registrado até a finalização do fluxo de trabalho e a geração de evidência final (que pode ser considerado, ao final, um documento arquivístico).

As interfaces de relacionamento de informações não solucionam os problemas de organização da informação digital, mas tem um papel importante nisso, ao serem a forma de mediar e dar acesso as informações organizadas. Planejar essa mediação, utilizando a estrutura de descrição de links da Web 3.0, ou as abordagens semânticas e ontológicas na descrição e indexação é, portanto, uma forma de organizar a informação e disponibilizar para o usuário final, facilitando o trabalho dele na busca por

conhecimento e no esclarecimento de um estado anômalo de conhecimento (CHOO, 2003). Nos capítulos seguintes são estudadas a informação de arquivo nestes aspectos, sendo criada para a interoperabilidade, fornecendo informações para a Web semântica e sistemas informatizados de arquivos, e se pode ser considerada uma ontologia a ser utilizada na preservação da informação arquivística digital.

1.3. Arquivística Contemporânea: interdisciplinaridades, objeto(s) e funções

Para este livro são utilizados os termos Arquivologia, Ciência Arquivística e Arquivística como sinônimos. Devido aos debates semânticos acerca da utilização destes termos, e por uma questão de fluidez de escrita, estes conceitos serão considerados o mesmo objeto de análise.

O próprio Dicionário Brasileiro de Terminologia Arquivística propõe tal generalização, quando explica no verbete “Arquivologia: Disciplina que estuda as funções do arquivo e os princípios e técnicas a serem observados na produção, organização, guarda, preservação e utilização dos arquivos. Também chamada arquivística” (ARQUIVO NACIONAL, 2005, p. 37). Aqui são acrescentados os demais termos para ampliar as possibilidades de expressão. Este livro, por sua característica dialógica entre Tecnologia da Informação e Ciência Arquivística e da Informação, já a define com um enfoque contemporâneo, independente do debate acerca do(s) objeto(s) científico(s) e de trabalho da área Arquivística.

A visão das ontologias digitais arquivísticas aqui exposta, extrapola o documento arquivístico, mas foca na informação descritiva resultante do processo de arquivamento de séries

documentais, ou seja, na chamada informação arquivística (que são informações do fundo documental e do seu contexto). Ao vincular este livro à informação arquivística, considera-se, portanto, um objeto de análise, mas sem a preocupação de definir este como objeto único ou atual ou principal da área. Assume-se, desde então, uma visão informacional, ou como menciona Santos (2011, p. 119), onde os:

[...] objetos da Arquivística são, em ordem de frequência de adoção pelos diversos teóricos: os arquivos (instituição e conjunto documental), os documentos arquivísticos, o trabalho (função profissional) arquivístico e a informação orgânica. Inicialmente, se poderia dizer que os objetivos da Arquivística podem ser sintetizados na manutenção contínua do acesso às informações orgânicas. (SANTOS, 2011, p. 119).

Porém, isso não significa descartar ou negar os preceitos e técnicas tradicionais e o objeto declarado. Pelo contrário, estes pontos estão presentes neste livro, mas focados no acesso contínuo à informação gerada em arquivos, que é uma premissa desta abordagem, considerada contemporânea. Isso engloba, além do conteúdo dos documentos arquivísticos, todo o contexto arquivístico, tanto que o mesmo autor também destaca que:

Embora se reservando o direito de não adjetivar o paradigma atual como “pós-custodial”, “contemporâneo” ou qualquer outro termo, propomos o entendimento de que o objeto da disciplina é a informação orgânica registrada, principalmente em sua manifestação estruturada e em seu conjunto, quais sejam os documentos e os fundos arquivísticos e, também, o papel dos arquivos como instituição (arquivos públicos, arquivos institucionais) na

preservação e na concessão de acesso às informações. (SANTOS, 2011, p. 121).

Portanto, esta abordagem aplica a valorização do conteúdo informacional dos documentos, ou seja, aborda-se arquivisticamente a informação. Neste caso, esta informação é aquela originada ou localizada nos acervos compostos por informações orgânicas originais (os documentos arquivísticos com seus respectivos vínculos). Ela deve estar contida em séries de documentos registrados em suporte convencional, ou estão em suporte que permitam a gravação eletrônica e servem como prova. Tais documentos arquivísticos são mensuráveis pela sua ordem binária e produzidos ou recebidos por pessoa física ou jurídica, decorrentes do desenvolvimento de suas atividades (LOPES, 1996). O contexto desta informação é representado por meio de metadados e deve estar vinculado ao documento e ao seu conjunto.

Para a UNIRIO (2015) a Arquivologia é uma área do conhecimento das Ciências Sociais Aplicadas que possui um quadro conceitual e de uma metodologia própria e específica (o que caracteriza uma ciência) e, portanto, estuda e trata os dados contidos nos documentos arquivísticos transformando-os em informação capaz de produzir conhecimento e desenvolvimento social. É uma visão de tratamento e uso do documento arquivístico.

Para a universidade, o objeto de estudo da Arquivologia e área de intervenção é a informação arquivística, isto é,

[...] a informação de natureza orgânica e funcional, pública ou privada, coletiva ou pessoal, produzida, recebida e acumulada por pessoa física ou jurídica em razão de seus objetivos. Com a gestão da informação arquivística assegura-se a constituição e a preservação da memória institucional e pessoal. (UNIRIO, 2015).

Vê-se, neste caso, a área de atuação da Arquivologia como a gestão da produção, do processamento e da disseminação da informação corrente, necessária e básica para a tomada de decisões na administração contemporânea e, complementa ao pronunciar que:

[...] seu objeto de estudo e intervenção é a informação arquivística, isto é, uma informação de natureza orgânica e funcional, pública ou privada, coletiva ou pessoal, produzida, recebida e acumulada por pessoa física ou jurídica em razão de seus objetivos. (UNIRIO, 2015).

Para o curso, com a gestão da informação arquivística assegura-se a constituição e a preservação da memória institucional e pessoal. Neste caso, também traz como objeto científico a disseminação (ou difusão de acervo) como partícipe do objeto de ação da área. Cabe destacar que, para ambientar o foco dado a este trabalho dentro do campo arquivístico, considera-se a descrição arquivística como instrumento de padronização e disseminação de informação, tratando-a como uma das funções arquivísticas. Por conseguinte, ao tratar de descrição e acesso, também se trata de organização da informação e as funções arquivísticas fazem parte da disciplina.

Nesse sentido, a organização da informação se torna uma linguagem comum entre as áreas da Arquivologia e da Ciência da Informação. Para ambas áreas se trata de representação, indexação, organização e difusão, no caso específico da arquivologia tudo que envolve a informação arquivística. Muitos elementos da Arquivologia são referenciados como organização da informação, resultando na informação arquivística.

Dentro de uma Ciência há disciplinas. Nas disciplinas estão presentes as técnicas e as metodologias. Nesse sentido, entender os aspectos funcionais da Arquivística é importante para admitir a importância da discussão e entender o contexto dentro da realidade da organização da informação. Na área da Arquivologia, existem as funções arquivísticas, que acabam determinando as principais atividades do trato arquivístico.

Utilizou-se, para este estudo, uma abordagem determinada por Rousseau e Couture (1998), que afirmam que são 7 as funções arquivísticas: produção, avaliação, aquisição, conservação, classificação, descrição e difusão. Portanto, fez-se necessário explicar um pouco acerca de cada uma, mas como parte do escopo deste trabalho é preciso destacar a Descrição Arquivística. A produção ou criação compreende os procedimentos técnicos relacionados à manutenção de padrões na produção documental, incorporando a definição de normas, conteúdo, modelos, formato e trâmites ou fluxos. A avaliação é realizada a partir de critérios pré-estabelecidos, como a definição dos prazos de guarda e destinação (eliminação ou preservação) do documento arquivístico. A aquisição contempla a entrada de documentos nos arquivos correntes, intermediários e permanente. Refere-se ao arquivamento corrente e aos procedimentos de transferência e recolhimento de arquivo.

A conservação ou preservação, para Rousseau e Couture (1998), incorpora os procedimentos relativos à manutenção da integridade física e (ou) lógica dos documentos ao longo do tempo, e as tecnologias e métodos que permitem o seu processamento e recuperação. A classificação refere-se à criação e a utilização de planos de classificação que reflitam as funções, atividades e ações ou tarefas da instituição acumuladora dos documentos arquivísticos, nas

fases corrente e intermediária, e a elaboração de quadros de arranjo na fase permanente, além de revisão desse instrumento, efetivando-as quando cabíveis. A difusão ou acesso se refere à acessibilidade do documento. Não se restringe ao acesso, mas também à proliferação das informações contidas nos documentos (ao que se considera a oferta de informação interoperável para a Web semântica). Com base em Rousseau e Couture (1998), foi concebido um quadro contendo o resumo das funções arquivísticas:

Quadro 1 - Resumo das funções arquivísticas.

Função	Componentes
Produção / Criação	• Garante a integridade e autenticidade; • Padronização de elementos de descrição; • Fluxos informacionais.
Avaliação	• Critérios preestabelecidos (classificação e temporalidade); • Definição dos prazos de guarda; • Destinação (eliminação ou preservação).
Aquisição	• Entrada de documentos no sistema de arquivo; • Procedimentos de transferência e recolhimento de arquivo.
Conservação / Preservação	• Manutenção da integridade física e (ou) lógica; • Tecnologias e métodos de

	processamento e recuperação.
Classificação	• Criação e à utilização de planos de classificação nas fases corrente e intermediária; • Elaboração de quadros de arranjo na fase permanente.
Difusão	• Acesso; • Divulgação das informações contidas nos documentos.
Descrição	• Multinível; • Definição de padrões de informação; • Indexação e linguagens documentárias.

Fonte: Adaptado de Rosseau e Couture (1998).

Sobre a descrição arquivística é importante destacar que é uma função que perpassa todo o ciclo de vida do documento (uma visão integrada dos processos de descrição), devendo ter seus elementos adequados para cada uma de suas fases. Deve considerar a unidade documental a qual se refere e as necessidades do usuário (aplicando o princípio multinível, que condicionou a criação das normas descritivas).

Em relação à informação digital, a descrição arquivística utiliza um conjunto de metadados para evidenciar seu conteúdo e devemos considerar que “os metadados fazem parte do processo de preservação dos documentos arquivísticos digitais e que os mesmos devem ser preservados juntamente com os documentos arquivísticos digitais” (INNARELLI, 2015, p. 209).

A indexação e acesso, portanto, são componentes da descrição arquivística, efetivado por meio de processos de estabelecimento de pontos de acesso (temas), para facilitar a

recuperação dos documentos ou informação. Compreende a criação e utilização de índices e de vocabulários controlados e, em alguns casos, podem ser desenvolvidos tesouros, seguindo os princípios e normas oriundas da Documentação.

Ao analisar as estruturas de metadados com ontologia, imagina-se que eles podem suprir o papel de linguagem documentária, e sendo ontologia, não será só vocabulário de substantivos ou relação entre sinônimos, pois incorpora outras variações e relações que são garantidas além de um tesouro. Neste trabalho, que visa a interoperabilidade da informação arquivística, focamos na função Descrição, porém entende-se que as Ontologias Digitais Arquivísticas podem participar de outros processos, e principalmente na preservação e difusão.

1.4. A Ciência Arquivística e alguns conceitos fundamentais

Com o desenvolvimento da escrita surge a necessidade de registro de atividades administrativas. Foi por conta do registro de transações comerciais que a escrita foi difundida pelos Sumérios. No momento que criaram símbolos inscritos em suportes, começou a necessidade de acumular documento para prova e pesquisa futura. Então, é plausível afirmar que a Arquivologia surgiu com a escrita, ou até que a escrita surgiu após a formação dos arquivos primitivos da humanidade.

Os conceitos mais tradicionais tratam a Arquivologia como um campo que tem por objetivo o conhecimento dos arquivos e das teorias, métodos e técnicas a serem observados na sua

constituição, organização, desenvolvimento e utilização. Outro ponto a se destacar é quanto ao nome. O mais comum é tratar Arquivologia igual a Arquivística, isto é, como sinônimos - como assumimos neste trabalho. Porém, alguns autores da área dividem os conceitos: Arquivologia para a Ciência e Arquivística como disciplina ou técnica. A tendência atual é identificar como Arquivística, pois conforme afirma Santos (2011, p. 103):

Essa é uma opção pragmática no sentido de aproveitar uma situação que já se encontra estabelecida no país – o uso de ambos os termos –, reorientando a interpretação atual de sinonímia entre eles. Entretanto acreditamos que, no âmbito internacional, com reflexo na realidade brasileira, é provável que a mudança paradigmática ora em curso na área tem influenciado a consolidação da denominação “Arquivística”. (SANTOS, 2011, p. 103).

Quanto ao termo Arquivo, segundo o Dicionário de Terminologia Arquivística, trata-se de um “conjunto de documentos que, independente da natureza ou do suporte, são reunidos por acumulação ao longo das atividades de pessoas físicas ou jurídicas, públicas ou privadas” (ARQUIVO NACIONAL, 2005, p. 27). É um conceito clássico e que destaca alguns princípios que será abordado adiante, como o de cumulatividade (que são reunidos por acumulação) e de organicidade (ao longo das atividades). Do ponto de vista legal, a Lei federal n. 8159 de 1991 Art. 2o, destaca que:

Arquivo são os conjuntos de documentos produzidos e recebidos por órgãos públicos, instituições de caráter público e entidades privadas, em decorrência do exercício de atividades específicas, bem como por pessoa física, qualquer que seja o suporte da

informação ou a natureza dos documentos. (BRASIL, 1991).

Cabe ressaltar, ainda, o papel polissêmico do termo Arquivo. Segundo o Dicionário Brasileiro de Terminologia Arquivística (ARQUIVO NACIONAL, 2005, p. 27) o conceito possui quatro verbetes:

1. Arquivo como conjunto de documentos produzidos e acumulados por uma entidade coletiva, pública ou privada, pessoa ou família, no desempenho de suas atividades, independentemente da natureza do suporte; 2. Arquivo como instituição ou serviço que tem por finalidade a custódia, o processamento técnico, a conservação e o acesso a documentos; 3. Arquivo como instalações onde funcionam os arquivos; E por último, 4. Arquivo como móvel destinado a guarda de documentos.

É importante sempre ressaltar qual arquivo conceito está sendo utilizado e, neste subcapítulo, o Arquivo é um dos objetos de trabalho do campo Arquivístico, ou da disciplina científica arquivística.

Outro conceito básico a ser explorado é o de documento. Primeiramente, o “documento” denota um conceito lato para, a seguir, abordar o documento arquivístico. Por isso, tratam-se de abordagens diferentes. Para um senso comum, documento poderia ser considerado o suporte (desde *bytes* até lâminas de pedra) e alguma informação registrada. Então, o suporte é o material que carrega a informação e o local no qual são registradas as informações. Já a informação é o elemento referencial, noção, ideia ou mensagem contidas em um documento. Logo, pode-se entender o documento como uma “unidade de registro de informações, em

qualquer suporte ou formato, suscetível de consulta, estudo, prova e pesquisa” (ARQUIVO NACIONAL, 2005, p. 18).

É importante destacar a não necessidade de algum suporte específico (qualquer que seja o suporte da informação) ou em relação a sua natureza. Esse conceito também incorpora a visão de que são conjuntos de documentos produzidos e recebidos, mostrando o vínculo com a função desempenhada pela instituição. Assim, o documento se vincula ao contexto que o gerou.

A questão funcional já pressupõe a necessidade de haver processos, responsáveis pelo desempenho das funções. Esse conceito está presente em Thomassen (2006, p. 6), que afirma que os arquivos são compostos por informação vinculada aos processos, ou seja, informação gerada e estruturada por processos de trabalho”. Logo, o que compõe o arquivo “é informação gerada e estruturada por processos de trabalho funcionalmente interrelacionados”, destaca o autor.

Já vinculando à atividade do profissional de Arquivo, o autor descreve que “um sistema de gerenciamento arquivístico é desenvolvido para estabelecer, manter e explorar a ligação entre estes processos de trabalho e a informação que geram, a fim de otimizar os potenciais informacionais decorrentes de suas relações” (THOMASSEN, 2006, p. 6). Estes sistemas serão abordados em mais profundamente em tópicos posteriores.

No que se refere ao documento de arquivo, o que a caracteriza e o diferencia do sentido lato é seu vínculo à ação realizada e que ele apoiou. Portanto, de acordo com o Conselho Nacional de Arquivos (CONARQ, 2004, p. 7), trata-se da “Informação registrada, independente da forma ou do suporte, produzida ou

recebida no decorrer das atividades de uma instituição ou pessoa, dotada de organicidade, que possui elementos constitutivos suficientes para servir de prova dessas atividades”.

Para Gonçalves (1998, p. 20) o documento arquivístico é um “documento que um determinado organismo, seja ele pessoa física ou jurídica, produz no exercício de suas funções e atividades”. A organicidade, logo, é o diferencial entre esses dois conceitos (o lato e o arquivístico) e o destaque que se dá quanto ao valor arquivístico, pois, como destaca Sousa (2014, p. 7) se o “documento é o resultado da atividade de uma pessoa física ou jurídica, podemos falar do caráter orgânico desse registro”. O aspecto orgânico é importante, pois é o que caracteriza a informação arquivística, como utilizada neste trabalho.

O aspecto orgânico se destaca nas abordagens que demonstram se uma informação produzida ou recebida no decorrer das atividades de uma instituição ou pessoa, ou da “produção que pode significar tanto a elaboração do documento pelo próprio organismo, como a recepção e guarda”, conforme expõe Gonçalves (1998, p. 20.).

Ainda, o documento arquivístico carrega em si dois importantes pressupostos, a confiabilidade, que é a capacidade de um documento arquivístico de “sustentar os fatos que se refere”. Que é garantido por meio de sua completude, ou seja, “Tem que ser completo, ser criado pela autoridade competente e ter seus procedimentos de criação bem controlados”, de acordo com o Glossário da Câmara Técnica de Documentos Eletrônicos (CTDE) (CONARQ, 2009). Portanto, está relacionada ao momento da sua produção e à veracidade do seu conteúdo.

E também a autenticidade, que é uma característica da qualidade de um documento ser de fato o que é, independente se for rascunho ou minuta, original ou cópia, e que seja livre de adulterações ou qualquer outro tipo de intervenção. Este pressuposto está relacionado com a forma de transmissão, além de suas estratégias de custódia e preservação. Logo, o documento arquivístico serve como testemunho da ação que o gerou, por isso ele precisa ser confiável e autêntico (CONARQ, 2009).

Cabe destacar, sobre o caráter orgânico do Arquivo, que se trata do vínculo da gênese documental ao próprio ciclo de produção e uso da documentação. O ciclo que compreende uma entrada que são materiais, serviços e informações. Essa entrada é processada e agregada a cadeias de valores, onde são executadas atividades para fim e para alimentar o meio. Durante esse processo ocorre uso e produção documental. Ao final, também há na saída os materiais, serviços e informações que alimentam uma nova cadeia de produção e geração. Esse caráter orgânico é originado nas estruturas das instituições por intermédio de funções desempenhadas e atividades realizadas. Pode-se imaginar que a “organicidade é revelada pelo inter-relacionamento e pelo contexto de existência e de criação”, ressalta Sousa (2014, p. 7).

Vale lembrar, finalmente, sobre o conceito de “fundo” que surge no quadro teórico-conceitual da Arquivologia para orientar suas práticas organizativas e que equivale a noção de “coleção”, comumente utilizada na Biblioteconomia. Foi no período de maior aproximação interdisciplinar da Arquivologia com o campo da História, Direito e Administração. Após esse processo, houve o desenvolvimento de princípios teórico-conceituais e metodológicos próprios e o desenvolvimento de terminologia própria para explicitar as diferenças teórico-conceituais das áreas. Um destes conceitos

específicos que engloba arquivos, documentos e organicidade diz respeito ao vínculo arquivístico, ou *Archival bond*, que será abordado a seguir.

1.5. **Archival Bond: o Vínculo entre um Documento e Outro**

O conceito de *Archival bond*, ou vínculo arquivístico, engloba arquivos, documentos e organicidade. Ela é a característica primordial do documento arquivístico, isto é, fazer parte de algo que vincula o documento a uma estrutura ou instituição. Para Rocha (2011, p. 85) “alguns arquivistas se referem ao conceito de *Archival bond* como vínculo arquivístico. Entretanto, existe um outro termo utilizado tradicionalmente pelos arquivistas brasileiros, com o mesmo significado de *Archival bond*, que é o de relação orgânica”. Este termo é derivado do termo organicidade, portanto, ao qual *Archival bond* está relacionado. Podemos considerar a similaridade de sentido, neste ponto, entre *Archival bond*, vínculo arquivístico e relação orgânica do documento.

Para Thomassen (2006, p. 6), o que compõe o arquivo “é informação gerada e estruturada por processos de trabalho funcionalmente interrelacionados”. Novamente o relacionamento entre os documentos em uma série é parte integrante do sentido de Arquivo para o autor. Para a prática do profissional de arquivo, como dito, o autor ainda destaca que “um sistema de gerenciamento arquivístico é desenvolvido para estabelecer, manter e explorar a ligação entre estes processos de trabalho e a informação que geram, a fim de otimizar os potenciais informacionais decorrentes de suas relações”.

A rede de relações que cada registro tem com os registros pertencentes de uma mesma série mostra que o vínculo de arquivo é originário, conforme Duranti (1997), porque passa a existir quando um registro é criado; ou seja, um documento pode ser considerado um documento arquivístico somente se adquire um vínculo de arquivo. Para a autora, os documentos que são resultados de uma transação não são documentos arquivísticos até serem relacionados com outros documentos, pois para que os documentos que não são resultado de uma transação se tornem registros orgânicos é necessário adquirirem o vínculo de arquivo, por meio do relacionamento com outros documentos que participam da mesma atividade.

Para a autora, o vínculo de arquivo pode ser revelado pelo ordenamento, o seu código de classificação ou o seu número de registro. Além de determinar a estrutura dos fundos arquivísticos, o vínculo de arquivo é o componente principal de identificação de cada registro, como diversos documentos são idênticos os registros tornam-se distintos depois de adquirir o vínculo de arquivo (ou seja, serem depositados em sua série ideal).

Duranti (1997) também descreve que o vínculo de arquivo é a expressão do desenvolvimento da atividade em que o documento participa, porque ele contém em si a direção da relação de causa-efeito (ou poderia ser a Função-Atividade). Portanto, o vínculo de arquivo determina o significado do registro. Esta é a razão pela qual a descrição no nível de item não é recomendável, pois poderia destruir o vínculo de arquivo e, conseqüentemente, os registros restantes sobre os registros (a metainformação arquivística).

A autora também salienta que esta é, também, a razão pela qual a descrição arquivística, como meio de elucidar a natureza do vínculo de arquivo em seu contexto documentário, tem sido tradicionalmente considerada a principal forma de perpetuar e autenticar o significado dos documentos e, em tempos de migrações de ambientes informáticos, de que é, provavelmente, o melhor método para garantir a autenticidade em longo prazo. Isto é importante para garantir a integridade da cadeia de custódia documental e dos ambientes e plataformas digitais que a compõem e onde as ontologias teriam um papel em sua manutenção. Pelo elucidado, vê-se, aqui, o uso da informação descritiva do contexto na preservação da informação digital arquivística.

Por sua vez, para Rocha (2011, p. 85), *Archival bond* está relacionado à organicidade, que é característica de arquivo. Já a organicidade é consagrada na área de arquivos e extremamente referido na literatura clássica da área. O diferencial do arquivo em relação a uma coleção de documentos é a organicidade e, além disso, os documentos de arquivo apresentam entre si o *Archival bond*. Logo, o *Archival bond* (ou o vínculo de arquivo) é parte indissociável da informação arquivística e, portanto, pertence à estrutura de um documento arquivístico digital, como podemos ver adiante.

1.6. Documento Arquivístico Digital, seus parecidos, sua preservação e tipologia

Os chamados documentos em “suporte não convencionais” são aqueles num suporte diferente do papel, e englobam os eletrônicos e os digitais. Podem ser arquivísticos ou podem apenas ser de apoio a processos de conhecimento, por exemplo. Já um documento arquivístico digital (DAD) é um documento digital que é tratado e gerenciado como um documento

arquivístico, ou seja, incorporado ao sistema de arquivos. Já um documento eletrônico quando reconhecido pode ser tratado como um documento arquivístico (CONARQ, 2009). Assim cabe diferenciar o DAD do documento eletrônico.

Um documento arquivístico digital é diferente de um documento arquivístico eletrônico, pois é um documento arquivístico codificado em forma analógica ou em dígitos binários, acessível por meio de um equipamento eletrônico. Dessa forma, são exemplos de documentos arquivísticos eletrônicos a fita audiomagnética, a fita vídeomagnética, os documentos processados por computador, etc. Enfim, todos os documentos que precisam de equipamentos eletrônicos para serem acessados.

Já um documento arquivístico digital é um documento codificado em dígitos binários em plataformas digitais ou portais corporativos. Podem se apresentar em formato digital, textos, imagens fixas, imagens em movimento, gravações sonoras, mensagens de correio eletrônico, páginas da Web, bases de dados, entre outros.

A diferença entre documento digital e documento digitalizado, é que a digitalização é a conversão de um documento em qualquer suporte ou formato para o formato digital, utilizando de tecnologias de digitalização, como os *scanners*, criando um representante digital do documento. É utilizada para reprodução e acesso de documentos convencionais a fim de agilizar a sua pesquisa e localização.

A digitalização também pode ser utilizada no processo de capturar documentos arquivísticos convencionais

para um SIGAD. Neste caso, o documento digitalizado e capturado pelo sistema é um documento arquivístico digital e, desta forma, obedecerá aos mesmos procedimentos e operações técnicas da gestão arquivística como tramitação, armazenamento e destinação. (CONARQ, 2016).

A digitalização então tem um objetivo de criar cópias, onde o original deve ser preservado, porém ao criar um SIGAD, estará criando um acervo que pode vir a ser permanente. Então é necessário planejar a preservação digital separando os documento nato digitais dos representantes digitais, pois tem destinação diferentes.

Quadro 2 – Tipos diferentes de documentos digitais e eletrônico.

Documento arquivístico digital	Documento arquivístico codificado em dígitos binários em plataformas digitais ou portais corporativos.
Documento arquivístico eletrônico	Documento arquivístico codificado em forma analógica ou em dígitos binários.
Representante Digital	Com objetivo de criar cópias para promover acesso. Não pode ser eliminado original.

Fonte: Adaptação de Conselho Nacional de Arquivos (2016).

A preservação digital engloba ações visando manter a integridade e a acessibilidade dos documentos digitais ao longo do

tempo. “Devem alcançar todas as características essenciais: físicas (suporte), lógicas (software e formato) e conceituais (conteúdo exibido), conforme destaca o CONARQ (2016).

1.7. Preservação Digital e as tipologias

Uma estratégia de preservação deve ser aplicada desde o planejamento da gestão da informação arquivística, definindo-se prazos de guarda e eliminação e aplicado diretamente no objeto digital via sistemas de gerenciamento, desde sua gênese. “As ações de preservação digital têm que ser incorporadas desde o início do ciclo de vida do documento. Tradicionalmente, a preservação de documentos arquivísticos se concentra na obtenção da estabilidade do suporte da informação”, descreve o CONARQ (2011, p. 35). Ainda, o Conselho delinea que, para realizar a preservação, é necessário realizar diversas operações, entre elas “mudanças de suporte e formatos, bem como atualização do ambiente tecnológico. A fragilidade do suporte digital e a obsolescência tecnológica de hardware, software e formato exigem essas intervenções periódicas”.

A gestão de metadados para preservação, por meio dos metadados administrativos e estruturais, é definida ao entender um documento em suas características intrínsecas e extrínsecas, com a Diplomática, que ajuda a definir parâmetros para garantir a autenticidade, fidedignidade e fidelidade dos documentos. A Diplomática, originalmente, foi criada para avaliar se determinados documentos envolvidos em disputa entre nações eram autênticos e serviriam como prova. Trata-se de uma ciência forense como a própria arquivística e todos seus princípios que garantem a prova documental. “O termo **forma** é a palavra chave da Diplomática,

entendendo por ela não só os caracteres externos do documento, mas também sua disposição material e a ordenação interna do texto, o discurso diplomático”, descrevem Togolês e Guimarães (2009, p. 28-29). Os autores destacam uma importante autora para a área Arquivística, a Luciana Duranti e sua obra intitulada “Diplomática Arquivística Contemporânea”:

Para Duranti (1995) a forma é definida como um complexo de regras de representação usado para transmitir uma mensagem. Essa forma é constituída das características do documento manifestando-se física e intelectualmente. É papel do método diplomático repartir e analisar essa forma, através de seus elementos internos e externos, com o objetivo de analisar a gênese documental, sua evolução, tradição e autenticidade. (TOGNOLI; GUIMARÃES, 2009, p. 29).

Para a diplomática arquivística contemporânea “os estudos realizados por Duranti provam que é perfeitamente possível a aplicação da crítica diplomática aos documentos contemporâneos”, que auxilia sobre a compreensão dos contextos de produção “e com isso os arquivistas passam a ver o método diplomático, ou estudo da Tipologia Documental como uma ferramenta de suma importância para o conhecimento da documentação gerada no século XXI” (TOGNOLI; GUIMARÃES, 2009, p. 31).

Duranti coordenou o Projeto internacional InterPARES, que identificou as características de um documento arquivístico digital. Definiu a importância da forma documental fixa, ou seja, a apresentação da mesma forma que tinha quando o documento foi armazenado. Deve ter o conteúdo estável, com o documento tendo que permanecer completo e inalterado.

A tipologia documental pode ser subsídio para compreender contextos da produção arquivística e, no caso, auxiliar a garantir a autenticidade em ambientes e plataformas digitais. O tipo documental é a "configuração que assume uma espécie documental, de acordo com a atividade que a gerou" (CAMARGO; BELLOTTO, 1996). Por exemplo: uma espécie é um boletim. Quando se qualifica esta espécie, ela se torna um tipo e magnetiza sempre uma qualificação. Um boletim de ocorrência é um tipo. Uma certidão é uma espécie, uma certidão de nascimento é um tipo documental. Como insumo analítico, ajudam a entender os processos e o ambiente que as geraram possibilitando aplicar uma abordagem arquivística e compreender o contexto que gerou aquele documento arquivístico.

PARTE II: SISTEMAS DE INFORMAÇÃO ARQUIVÍSTICOS E TECNOLOGIA

Nesta segunda parte do livro, ao avaliar os sistemas de informação arquivísticos, admite-se sua importância no cenário de gestão da informação digital e o entendimento da necessidade de preservação e interoperabilidade.

Aborda-se os SIGAD (Sistemas Informatizado de Gestão Arquivística de Documentos), os Repositórios Arquivísticos Digitais Confiáveis - RDC-ARQ, e o Modelo *Record Continuum*, como ambientes de custódia documental. Falamos da cadeia de custódia e mais do contexto arquivístico, e da forma de garantir, em longo prazo, a confiabilidade do ambiente de gênese documental. Por fim, foi avaliada a preservação pelo modelo OAIS e a constatação da inerência da interoperabilidade na Web semântica.

2. Sistemas de Informação Arquivísticos

Um Sistema de informação, no sentido amplo, é a totalidade de todas as atividades formais e informais de representação de dados e de processamento dentro de uma organização, incluindo a comunicação vinculada, tanto internamente quanto com o mundo exterior. O autor Quisbert (2006) acredita que é possível qualificar um Sistema de Informação em sentido estrito como um subsistema baseado em computação, destinado a fornecer o registro e apoio a serviços de operação e gestão organizacional.

No lado da ciência a informação, Robredo (2003) afirma que o Sistema de Informação é uma entidade complexa e organizada e que capta, armazena, processa, fornece, usa e distribui informação. O sistema deve incluir os recursos organizacionais relacionados, tais como os recursos humanos, tecnológicos e financeiros. “É de fato um sistema humano, que inclui provavelmente recursos computacionais para automatizar determinados elementos do sistema” (ROBREDO, 2003 p. 110).

Para Ruesta (2012), ao abordar o uso do termo Sistema, afirma que “embora se possa discutir com muitos argumentos, as TIC (tecnologias da informação e comunicação) apropriaram-se do termo e a maior parte das pessoas interpreta sistema como sistema de informação ou aplicação informática” Ruesta (2012, p. 8), conforme destaca a autora.

Um Sistema de Informação de Arquivo, operando em uma plataforma tecnológica, é um tipo especial de sistema de informação. Como cada tipo especializado de sistema de informação,

o desenvolvimento de sistemas de informação de arquivo deve ser precedido de uma forma especializada de pensar, pois quem lidar com Sistemas de Informação de Arquivo está lidando em tornar a informação digital gravada acessível, como descreve Quisbert (2006). O autor afirma, ainda, que a compatibilidade é possível por meio de atitude aberta e padrões abertos. Por fim, a acessibilidade e a preservação em longo prazo devem esticar a longevidade dos registros ao longo do tempo-espço.

Numa tentativa de definição de Sistemas de Informação de Arquivo Quisbert (2006) afirma que poderia ser definido como sistemas de processamento de informações que servem qualquer organização e que tem a responsabilidade, ou a necessidade, de preservar e tornar a informação acessível a uma comunidade designada ao longo do tempo.

Numa visão mais ampliada em relação ao conceito de Sistema de Gestão Documental, a BAD (RUESTA, 2012, p. 9) acredita que “os sistemas de gestão definem-se como o conjunto de elementos interrelacionados ou que interagem numa organização com o fim de estabelecer políticas e objetivos, bem como os processos para os alcançar”. E mais adiante tal conceito é aplicado em relação às TIC, segundo o qual “as aplicações que se comercializam ou se desenvolvem para gerir os documentos eletrônicos também são designadas muitas vezes por sistemas de gestão documental”.

Neste estudo considerou-se os termos sistemas informatizados de arquivo, sistemas de gestão documental, sistema de informação de arquivo e sistemas de gerenciamento de arquivos como sinônimos, pois são aplicações de tecnologia da informação

que apoiam, por intermédio da “informatização”, as operações de gestão arquivística de documentos.

Cabe ainda destacar, que o termo Sistemas de Arquivos tem amplo uso na área arquivística e de forma polissêmica, segundo Feitoza e Silva (2012, p. 41), o termo Sistema de Arquivo “é comumente utilizado como órgãos de ascendência técnica e como estratégias; como métodos que facilitam o arquivamento de documentos e como redes de arquivos”. Aqui não é abordado este conceito, por isso sempre se qualifica o tipo de sistema, seja com os substantivos eletrônico, informatizado ou digital.

Neste livro, e a partir de agora, verificamos os seguintes tipos de sistemas informatizados de arquivo: o SIGAD (modelo brasileiro definido pelo e-ARQ), o RDC-Arq (Repositório Digital Arquivístico Confiável) e a solução proposta pelo modelo RC (*Record Continuum*), definido pela norma AS 4390. O modelo AS4390 foi internacionalizado em formato ISO, criando

La Norma ISO 15489 es lá primeira norma internacional para lá gestión de documentos. Para su elaboración se tomó como punto de partida la norma australiana AS 4390 -1996: *Records Management (Australian Standard for Record Management 4390)*, la cual estableció un conjunto de medidas y desarrolló una metodología para el diseño de sistemas orientados a la gestión de documentos de archivos electrónicos (*electronic records*) con garantías de veracidade. (DEL CASTILLO GUEVARA; MUGICA, 2001, p. 55).

Ele se estabeleceu no modelo australiano de *Record Continuum* e na implementação de sistemas para a captura, classificação, fluxo de trabalho, decisões de valorização e eliminação dos documentos arquivísticos unificando os dois preceitos num

mesmo sistema e possibilitando a melhoria da descrição e preservação em longo prazo. Contudo, o modelo australiano será analisado para contrastar com o modelo brasileiro definido pelo SIGAD.

A condição para a existência do SIGAD é, de fato, a utilização de instrumentos de gestão antes mesmo de um sistema informatizado. “O sucesso do SIGAD dependerá fundamentalmente da implementação prévia de um programa de gestão arquivística de documentos”, é o que diz o e-ARQ quando fala no SIGAD (CONARQ, 2011). Ou seja, o planejamento da gestão e da estrutura da informação é primordial no modelo brasileiro.

Para o modelo de ciclo de vida da informação, esta deve ser utilizada, a fim de evitar que seja considerada apenas dados. Portanto, a informação é conhecimento que pode ser trocado. A intenção de um sistema de informação de arquivo, assim sendo, não deve ser apenas o de armazenamento, mas de garantir a autenticidade, a perenidade da cadeia de custódia, além de estimular o uso de seu acervo por meio da adoção e difusão de metadados aos agentes computacionais por meio da Web semântica, diz Quisbert (2006).

Quisbert (2006) também menciona que o uso de Sistemas de Informação deve ser visto como praticar a disseminação da informação, a fim de promover o cultivo de conhecimento, não só dentro da comunidade onde atua, mas também na sociedade. A utilização dos registros como fontes de informação mantém os documentos vivos. Logo, estes dados devem ser preparados tanto para migração como para difusão.

Uma visão importante a ser compartilhada é do conjunto SIGAD e RDC-Arq, pois somados representam sistemas informatizados suficientes para atender ao modelo das três idades documentais (corrente, intermediário e permanente). Ao utilizar um SIGAD, estamos definindo a estrutura de gestão e de uso dos documentos e da informação arquivística correntes. Com o RDC-Arq se utiliza um ambiente de custódia definitiva que herdará os pacotes de informação preparadas para a interoperabilidade. Estes pacotes, portanto, devem ser descritos utilizando os padrões arquivísticos e são considerados os, assim que transformados em pacotes de arquivamento, os Arquivos Permanentes Digitais.

2.1. SIGAD: Sistema Informatizado de Gestão Arquivística de Documentos

A partir de agora conheceremos o SIGAD e a Norma que o viabiliza (Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos), e que representam a realidade brasileira em relação a normatização e modelo tecnológico de sistema de informação de arquivo correntes, procura-se relatar a diferença entre os dois (norma e modelo), visto comum a confusão inicial entre seus conceitos, e destaca-se o profissional da informação frente à tecnologia como definidor de requisitos, e a relação destas interfaces de gestão com a preservação digital.

Um SIGAD, Sistema Informatizado de Gestão Arquivística de Documentos, permite ter uma compreensão da organização como um todo, ao estruturar os instrumentos de gerenciamento que propiciem a existência da Gestão de Documentos Arquivísticos, principalmente o plano de classificação e a tabela de temporalidade.

Trata-se de um sistema desenvolvido para produzir, receber, armazenar, dar acesso e destinar documentos arquivísticos digitais. Pode compreender um software particular, um determinado número de softwares integrados, adquiridos ou desenvolvidos por encomenda, ou uma combinação desses recursos. Deve realizar todas as operações técnicas da gestão arquivística (da criação até a destinação final - a eliminação ou a guarda permanente).

O funcionamento de um SIGAD depende de implementação de procedimentos e políticas de gestão de documentos como, por exemplo, um plano de classificação e de destinação. Eles carregam as regras do sistema que fazem a gestão dos documentos arquivísticos digitais. Estes são os requisitos que veremos mais adiante, no e-ARQ (Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos). A seguir vemos algumas características de um SIGAD:

Quadro 3 – Características de um SIGAD.

SIGAD	QUE GARANTEM:
CONJUNTO DE FERRAMENTAS E METODOLOGIAS	As propriedades do documento arquivístico;
	O cumprimento do ciclo vital dos documentos;
	O impedimento de que esses documentos arquivísticos sofram alterações ou que sejam eliminados, exceto em situações previamente determinadas;
	A inclusão de prazos de guarda e controles de segurança rigorosos;
	A formação de um repositório seguro de documentos arquivísticos necessários para a realização das atividades e funções das organizações;
	A facilidade de acesso a esses documentos arquivísticos produzidos;
	A organização eficiente e eficaz da documentação arquivística de acordo com as previsões de um plano de classificação.

Fonte: Adaptação de Negreiros (2007).

É comum a confusão entre Gestão Eletrônica de Documentos (GED) ou *Enterprise Content Management* (ECM, em inglês) e um SIGAD. Um GED é um sistema que capaz de gerenciar os documentos eletrônicos, seja em forma de catálogo ou como ambiente de uso social. A estrutura de armazenamento de documentos pode ficar sob o controle dos usuários, permite a criação de versões dos documentos e que sejam eliminados pelos autores. O GED também gerencia os documentos digitalizados, as imagens de documentos, e as telas do GED exibem as imagens dos documentos para os usuários.

Já o SIGAD possui uma série de requisitos mínimos para garantir a fidelidade e autenticidade documental, isto é, o controle é permitido a determinados perfis e possui regras de retenção, aplicados por meio de instrumentos arquivísticos (classificação e arranjo).

O conjunto de requisitos de um SIGAD está determinado pelo e-ARQ, proposto pelo CONARQ (2011), já que um sistema de GED tem por objeto o depósito de documentos e não o gerenciamento do ciclo de vida de documentos arquivísticos. Dessa forma, estes sistemas tornam os profissionais da informação definidores de requisitos funcionais. Afinal, como aborda Luz (2010, p. 26) “a tecnologia da informação sempre será um conjunto de requisitos que são atendidos de formas diversas por uma interface de sistema, sendo assim um mecanismo de interação do humano com o computador”. Com isso, o papel de definidor de requisitos para o profissional da informação está cada vez mais claro, pois ele:

[...] não precisa programar um sistema, mas deve ter claros os requisitos de seu “negócio” (esta é a linguagem utilizada), e também deve saber o que é arquitetura tecnológica, arquitetura de informação e arquitetura de software, por exemplo. Como um cliente assíduo das áreas de TI (Tecnologia da Informação), ele deverá ser capaz de saber solicitar requisitos, de validar documentações como especificações funcionais e até acompanhar o gerenciamento de projetos. (LUZ, 2010, p. 26).

Para o caso específico da Arquivística, muitas vezes os requisitos indicam os caminhos, mas não são fórmulas de aplicação.

O e-ARQ, por exemplo, define os elementos de descrição, no entanto, não define como eles serão aplicados nos diferentes arquivos existentes, não havendo para a área o equivalente à Norma de Catalogação para a biblioteca. Os requisitos do e-ARQ Brasil são determinados por especificações técnicas e funcionais capazes de nortear a aquisição e/ou a especificação e desenvolvimento de sistemas informatizados de gestão arquivística de documentos. Os requisitos servem para que os sistemas informatizados possam garantir a confiabilidade, autenticidade e acesso ao longo do tempo dos documentos arquivísticos.

Vale lembrar que a confiabilidade é a capacidade de um documento arquivístico sustentar os fatos a que se refere. Para isso, deve ser completo, ou seja, ser criado pela autoridade competente e ter seus procedimentos de criação bem controlados. A confiabilidade está relacionada ao momento da sua produção e à veracidade do seu conteúdo. Além disso, a autenticidade caracteriza a qualidade de um documento ser de fato o que diz ser e que seja livre de adulterações ou qualquer outro tipo de corrupção. Logo, isto “está relacionado com a forma de transmissão e estratégias de custódia e preservação”, estabelece o CONARQ (2011, p. 22).

O documento arquivístico serve como testemunho da ação que o gerou e, por isso, ele precisa ser confiável e autêntico. Para tanto, precisa trazer as informações de contexto e metadados quanto à sua transmissão e custodiadores. Uma ontologia digital arquivística pode ser capaz de centralizar as informações de cada objeto informacional identificado como documento arquivístico e do contexto que o gerou.

Quanto à acessibilidade, o documento acessível é aquele que pode ser localizado, recuperado, apresentado e interpretado

sempre que necessário, ou quando houver uma inferência. A interoperabilidade também é inerente à organização, afinal “o sistema deve garantir a transmissão de documentos para outros sistemas sem perda de informação e de funcionalidades, e ser capaz de recuperar qualquer documento em qualquer tempo e apresentá-lo com a mesma forma de sua criação”, descreve o CONARQ (2011, p. 62).

O e-ARQ Brasil é um compêndio de requisitos para sistemas informatizados de gestão arquivística de Documentos e traz na sua introdução um conteúdo sobre gestão de documentos e a sistemas informatizados de gestão de documentos, tanto conceitos quanto metodologias como pode-se verificar:

Quadro 4 – Conteúdos abordados no e-ARQ Brasil.

e-ARQ BRASIL	Conceitos e metodologias: gestão de documentos sistemas informatizados de gestão de documentos
	Aspectos de funcionalidade Especificação dos Requisitos
	Glossário
	Metadados

Os requisitos do e-ARQ determinam como o SIGAD deve tratar o documento arquivístico, este sendo uma unidade complexa, produzido e/ou recebido por uma pessoa física ou jurídica, no decorrer das suas atividades, qualquer que seja o suporte, e dotado de organicidade. Ele sempre participa ou apoia uma ação. Gerencia fluxos entre pessoas envolvidas (autor, destinatário, redator, por exemplo).

O SIGAD recupera todos os componentes digitais (objeto físico e lógico) que formam o documento arquivístico (objeto conceitual), atendendo aos requisitos de gestão dos documentos a partir do plano de classificação para manter a relação orgânica entre os documentos. Ele gerencia o ciclo de vida dos documentos arquivísticos, que significa classificá-los de acordo com as funções e atividades da organização e aplica prazos de guarda específicos para cada classe.

Quadro 5 – Comparativo de um SIGAD e do e-ARQ.

e-ARQ	Especifica os requisitos funcionais de um sistema informatizado de Gestão Arquivística de Documentos (SIGAD);
SIGAD	É o sistema desenvolvido para realizar operações técnicas da gestão arquivística de documentos.

Num SIGAD a classificação registra e aplica um código de classificação e, além disso, a organicidade é expressa nas relações que os documentos guardam entre si, ao refletirem as funções e atividades das pessoas ou da organização que os produziram.

Por meio do registro de metadados dos documentos, há um controle do ciclo de vida, da autenticidade, do acesso e da preservação e, do mesmo modo, a segurança no armazenamento e gestão garantem a autenticidade dos documentos e a transparência das ações (trilha de uso auditável). Dessa forma, o SIGAD trata sistematicamente a seleção, a avaliação dos documentos

arquivísticos e a sua destinação (eliminação ou guarda permanente), conforme legislação em vigor. Não há eliminação automática e é necessário realizar a exportação dos documentos para transferência e recolhimento.

Todavia, para preservação é necessária uma abordagem de complementação de sistemas de gerenciamento de arquivos e de preservação em longo prazo, complementado a gestão documental com os repositórios digitais confiáveis. Assim, o SIGAD incorpora procedimentos para a preservação de longo prazo dos documentos arquivísticos, garantindo pela preservação do vínculo arquivístico e à manutenção da cadeia de custódia.

2.2. Repositórios Arquivísticos Digitais Confiáveis - RDC-Arq

Neste tópico é abordado o repositório arquivístico digital, considerando seu contexto característico e requisitos que qualificam esse tipo de Sistema de Informação de Arquivo.

Com base na resolução do CONARQ n. 43 (2015) chamada “Diretrizes para a Implementação de Repositórios Arquivísticos Digitais Confiáveis - RDC-Arq” que atualizou a resolução n. 39 de 2014, existem algumas considerações acerca do cenário para garantias de confiança de um sistema informático de Arquivo.

Estes elementos englobam a responsabilidade pelo repositório, o processamento técnico, a independência de repositórios e as garantias de interoperabilidade. Assim, verifica-se que a interoperabilidade da informação arquivística já é garantida a partir da visão de seu repositório que será capaz de preservar a sua integridade, como podemos ver a seguir:

Quadro 6 – Características de um RDC-Arq.

Responsabilidade pelo repositório	• Compartilhada por profissionais de arquivo e de tecnologia da informação; • Cumprir os requisitos tecnológicos e os procedimentos do tratamento arquivístico.
Tratamento arquivístico	• Capaz de organizar e recuperar os documentos, de forma a manter a relação orgânica entre eles, • Organização a partir de um plano de classificação de documentos; • Descrição multinível, de acordo com a norma internacional para descrição arquivística.
Independência dos repositórios	• Um repositório digital deve ter independência; • Seu funcionamento e o acesso aos documentos não podem depender das aplicações que funcionam em conjunto com ele.
Interoperabilidade	• Conformidade com as normas e padrões estabelecidos; • Níveis de interoperabilidade com outros repositórios digitais e sistemas informatizados que tratam de documentos arquivísticos.

Fonte: Adaptação de Conselho Nacional de Arquivos (2015).

A mesma norma estabelece alguns princípios de preservação digital para garantir o acesso de longo prazo aos

documentos arquivísticos autênticos, o que implica na adoção de alguns princípios para que um RDC-Arq deva ser capaz de atender aos procedimentos arquivísticos em suas diferentes fases, e aos requisitos de um repositório digital confiável.

Deve-se atender a estes requisitos pressupondo que a autenticidade dos documentos arquivísticos digitais esteja sob ameaça no momento da transmissão em rede (entre pessoas e sistemas) e em relação ao tempo (atualização e/ou substituição de hardware ou software usados para armazenar, processar e comunicar os documentos). Por isso, é necessário considerar que a preservação digital é um processo contínuo, que começa na concepção do documento (CONARQ, 2015).

Um ponto importante para ressaltar é que a autenticidade dos documentos arquivísticos digitais tem por base os procedimentos de gestão e preservação, além da confiança tanto no repositório como no órgão responsável pela guarda desses documentos. Por isso, a elaboração de manuais e os procedimentos de preservação realizados pelo repositório digital apoiam a presunção de autenticidade desses documentos, assim como o registro em metadados das intervenções de preservação em cada documento apoia a presunção de autenticidade desses documentos (CONARQ, 2015).

Essa operação pode, por exemplo, ser acompanhada com uma ontologia leve de origem descritiva, que evolui no espaço-tempo (acompanha o processo de avaliação e eliminação) e registra esses tipos de alterações até definir um quadro final de arranjo que represente as funções e atividades de um domínio. As ontologias leves são mais focadas nos aspectos estruturais, incluem conceitos, relações e instâncias. Já as ontologias pesadas possuem os aspectos

de uma ontologia leve acrescentando-se axiomas e restrições. (CORCHO; FERNÁNDEZ-LÓPES; GÓMEZ-PÉREZ, 2003).

Os repositórios confiáveis focam nos documentos arquivísticos autênticos, e não genericamente em objetos digitais, pois é necessário arbitrar o que se considera como documento original, uma vez que a preservação digital implica a necessidade de conversão de formatos e atualização de suportes. Outro fator destacado pelo CONARQ é quanto a distinção da autenticidade e da autenticação de documentos, considerando que a primeira é a qualidade de o documento ser verdadeiro, e a segunda é uma declaração dessa qualidade, feita, em um dado momento, por uma pessoa autorizada para tal.

A autenticidade dos documentos digitais deve ser avaliada e presumida no momento de sua submissão ao repositório. Por fim, o repositório digital é responsável pela manutenção permanente da autenticidade dos documentos a ele submetidos.

Como vemos, os princípios almejados pelo CONARQ representam as condições mínimas para considerar uma plataforma digital confiável e de que forma devem ser tratados os registros. Alguns deles poderiam englobar ontologias, mas sempre se trata de gestão de metadados e registros, do estado de integridade e de processamento da informação declarada como arquivística.

Para Santos e Flores (2015, p. 205) a confiabilidade é também concedida pela relação de segurança e credibilidade, pois “o repositório digital deve ser o ambiente autêntico para a preservação em longo prazo, dispondo, por exemplo, de ferramentas para a implementação das estratégias de preservação e inserção de padrões de metadados”. Os autores também afirmam que, para este

repositório, todas as ações realizadas em relação aos documentos digitais (migrações, por exemplo) devem ser registradas detalhadamente, criando um histórico de cada objeto digital armazenado, acrescentando confiabilidade aos conteúdos.

Ainda, destacam a interoperabilidade como fermenta de gestão e preservação, associada as políticas e ao plano de preservação para servirem de artifícios para a criação de um repositório arquivístico digital confiável em longo prazo (SANTOS; FLORES, 2015, p. 204-205).

Face o exposto, pode-se considerar que uma ontologia do tipo leve pode ter uma função no quesito preservação da informação digital e em documentos arquivísticos, principalmente ao analisarmos sua participação num RDC-Arq, no qual podemos verificar, anteriormente, seu papel de representar contextos arquivísticos (conceitos e instâncias) ao longo do tempo, nas migrações entre ambientes de custódia documental.

2.3. O Modelo Record Continuum

O conceito de *Record*, na língua inglesa é traduzido no Brasil (pelo Conselho Nacional de Arquivos) como Documento Arquivístico. Em outros países os chamados *files* representam o sentido lato de documento, enquanto *Record* (podendo ser traduzido como registro orgânico), representa o sentido que trabalhamos na arquivística. Conforme aponta Upward (1996, p. 4) sobre o *Australian Standard AS 4390, Records Management* (norma australiana de gestão documental, que implementa a visão em sistema de informação de arquivo do *Record Continuum*) “recorded information

in any form, including data in computer systems, created or received and maintained by an organisation or person in the transaction of business or the conduct of affairs and kept as evidence of such activity". Quanto ao *Continuum*, o autor descreve as características como a sua continuidade, as suas partes indiscerníveis, e a maneira como seus elementos passam de um período para o outro.

O *Record Continuum* (RC) é um modelo internacional, foi formulado na década de 1990 pelo teórico de arquivo australiano Frank Upward (1996, p. 5) com ajuda de colegas, com base em quatro princípios: O primeiro princípio diz respeito ao conceito de que o "documento arquivístico" sublima as suas utilizações para fins de memória transacional, probatório e seu *continuum* unifica as abordagens de arquivamento e manutenção de documentos arquivísticos, independe se os registros são mantidos por uma fração de segundo ou por um longo prazo de tempo. O segundo princípio do *Continuum* diz que há um foco em registros como entidades lógicas, em vez de entidades físicas, independentemente de se eles estão em suporte convencional (em papel), ou em formato eletrônico (natodigitais ou digitalizados). O terceiro princípio diz respeito à institucionalização do papel do profissional da informação e arquivista, que requer uma ênfase na necessidade de integrar documentos arquivísticos em sistemas e processos, tanto para fins comprobatórios como sociais. E, por fim, o quarto princípio seria de que a Ciência Arquivística é a base para organizar o conhecimento sobre a gestão e a manutenção de informação arquivística. Tal conhecimento é passível de revisão, mas pode ser estruturado e explorado em termos da operação e de princípios de ação tanto para documentos do passado, do presente como os do futuro.

Para o autor, a perda da fisicalidade que ocorre quando os documentos são capturados eletronicamente está forçando os

arquivistas a reavaliarem entendimentos básicos sobre a natureza dos documentos de atividade organizacional (e social) e de suas qualidades como prova. Assim, a custódia, para o arquivista pós-custodial é apenas um dos muitos problemas que devem ser enfrentados nas "múltiplas realidades" dos arquivos digitais (UPWARD, 1996, p. 5).

Ainda, conforme Upward (1996, p. 4), princípios estruturais em arquivos foram discutidos por Luciana Duranti usando os conceitos de 'vínculo de arquivo' e 'limite de arquivo'. Esses termos referem-se, respectivamente, às relações estabelecidas entre os documentos arquivísticos durante a manutenção desses documentos e o papel das instituições arquivísticas na autenticação de documentos arquivísticos que tenham passado por sua custódia. A ligação dos registros (*Archival bond*) e a fé de arquivo tornaram-se um processo de ordenação no tempo e no espaço. O limite de arquivo estabelecido e a manutenção da cadeia de custódia torna-se parte do modelo físico para a implementação.

Nos arquivos digitais, para o autor, a localização dos documentos (dentro de uma classificação funcional) e serviços oferecidos será de pouco interesse para aqueles que utilizam sistemas de informação de arquivos. Desse modo, todos os processos de garantia de autenticidade de documentos serão executados de acordo com novas estratégias que levem em consideração essas novas realidades. O autor ainda destaca que abordagens pós-custodiais para arquivos e documentos arquivísticos não podem ser compreendidas se continuarem a serem tratados como dualismo, pois conforme ressalta, o pós-custodial não é o oposto de custódia, o pós-custodial também engloba a custódia (UPWARD, 1996, p. 3-4).

Segundo “*A Glosar of Archival and Records Terminology*” da *Society of American Archivists* (SAA, 2016) uma perspectiva de *Record Continuum* pode ser contrastada com um modelo de ciclo de vida. O modelo de ciclo de vida defende que sejam claramente definidas as etapas de gestão documental para ser criada uma nítida distinção entre a manutenção de registros correntes e históricos. O conceito de *Records Continuum*, por outro lado, fornece para os gestores e arquivistas australianos uma maneira de reflexionar sobre a integração dos processos de gestão documental e arquivamento.

O modelo de ciclo de vida, para a SAA, trata de documentos que passam por fases até que eventualmente sejam eliminados, exceto para alguns selecionados, que são considerados como documentos de arquivos. Já uma abordagem baseada em *Continuum* sugere dimensões do espaço-tempo integrados. Os registros são 'fixados' no tempo e no espaço a partir do momento de sua criação, e então regimes de gestão documental o encaminham adiante e permitem a sua utilização para fins múltiplos, difundindo-os e fornecendo informação para pessoas que vivem em diferentes tempos e espaços.

O ciclo *Continuum*, para a mesma SAA, é um padrão unificado de vida de um registro, composto de quatro fases interrelacionadas: (1) Criação ou o recebimento; (2) Classificação; (3) Regimes da gestão documental, incluindo a manutenção no local de criação, uma área de armazenamento de ativos ou centro de registros, ou um arquivo; e (4) Uso (primário ou secundário). Nessa visão o objetivo final, seu uso, determina seu processo de disponibilização.

Para o *Australian Standard AS 4390, Records Management*, o *Records Continuum* é um regime uniforme e

coerente dos processos de gestão, a partir do momento da criação de documentos (e antes da criação, no projeto de sistemas de arquivo aplicando-se o desenho de sistemas), chegando à preservação e utilização dos documentos arquivísticos em arquivos, como descreve An (2003, p. 25). Para a autora, os gerentes de gestão documental e arquivistas são reunidos sob um Quadro Integrado de Documentos Digitais com o mesmo objetivo: garantir a confiabilidade, autenticidade e integridade dos registros.

A estrutura fornece a compreensão comum, padrões consistentes e unificados, os critérios de melhores práticas e abordagens interdisciplinares e colaborações nos processos de registros e arquivamento tanto no papel como no mundo digital. O *Record Continuum*, para An, fornece documentos arquivísticos sustentáveis para ligar o passado ao presente e do presente para o futuro. Ele pode coerentemente existir em um contexto mais amplo dinâmico, mutável, que pode ser influenciado por variáveis jurídicas, políticas, administrativas, sociais, comerciais, tecnológicos, culturais e históricas ao longo do tempo e espaço. O Quadro Integrado de documentos digitais poderia facilitar a proveniência, apoiar a prestação de contas, constituir a memória, facilitar a construção de identidade e fornecer fontes oficiais de informação de valor confiável.

A norma *Australian Standard AS 4390, Records Management* (STANDARDS AUSTRALIA INTERNATIONAL, 1996) possui características tanto para o padrão eletrônico como para registros tradicionais, é uma metodologia para a elaboração e implementação de estratégias de sistemas de documentos arquivísticos, para assegurar a criação e a captura de identificação de atributos que os documentos precisam para funcionarem efetivamente como evidência. Ainda, de acordo com a norma, a avaliação para o *Record*

Continuum é o processo de apreciação de atividades de negócios e que determinam quais registros precisam ser capturadas e quanto tempo eles precisam ser mantidos, para atender às necessidades de negócios, as exigências de prestação de contas da organização e as expectativas da comunidade

Observa-se que na norma australiana a classificação serve a múltiplos propósitos, envolve a elaboração e aplicação de um sistema baseado nas atividades de negócios que geram os documentos arquivísticos. A norma possibilita a gestão documental como processos de gerenciamento do conjunto de documentos desde o momento da criação destes, além de sua preservação e uso como arquivo.

Estes elementos, embora tratados diferentemente no sistema de ciclo de vida (na qual o SIGAD é baseado) ou no modelo de sistema *Record Continuum* possuem pontos que são comuns: ambos fazem gestão de temporalidades, de classificações, usos e acessos e garantem, com suas estratégias de descrição e preservação a manutenção da cadeia de custódia, importante para garantir a confiabilidade em sistemas de informação de arquivos, assunto abordado a seguir.

2.4. Cadeia de Custódia e Contexto Arquivístico

A manutenção da autenticidade em toda a cadeia de custódia visa o acesso em longo prazo e a consequente preservação, numa relação de confiança, independente se for um SIGAD (focado em documento corrente e intermediário), um RDC-Arq (focado em arquivo permanente) ou sistema integrado como na visão de Record

Continuum. A manutenção da cadeia de Custódia deve ser feita por meio de Ambientes Autênticos (que garantem a confiabilidade, como verificado), sejam eles SIGAD (e-ARQ Brasil) nas fases corrente e intermediária, e os RDC-Arq (Repositórios Digitais Confiáveis Arquivísticos) na fase permanente (CONARQ, 2015).

A norma e-ARQ Brasil, contempla plataformas digitais para a Gestão Documental, isto é, no uso corrente e intermediário quando apoiam a administração e as atividades das instituições e famílias. Após o término da fase da Gestão de Documentos, com a alteração da cadeia de custódia, passa-se para a fase de Administração de Arquivos Permanentes e, para isso usa-se os RDC-Arq (CONARQ, 2015).

Para Flores (2014), este processo deve contemplar os processos descritivos de acervos permanentes realizados nas etapas de Arranjo, Descrição, Digitalização, Difusão e Acesso de Documentos de caráter permanente, e não mais permitindo ações ou operações da Gestão de Documentos. O autor também destaca que o documento arquivístico nasce orgânico, mas para permanecer (e manter o *Archival bond*), precisa do Plano de Classificação Documental – PCD. Isso pode ser dito como se o documento não permanecesse vinculado a outros (o PCD servindo como seriador), porém tem-se o risco de perder seu vínculo, o que pode indicar a interrupção da cadeia de custódia.

Quando destinados para guarda permanente, há uma alteração na cadeia de custódia, passando a responsabilidade pela preservação dos documentos dos produtores para a instância de guarda. O processo é migrar de um SIGAD para um RDC-Arq. Para o CONARQ (2015), os documentos digitais em fase permanente são dependentes de um bom sistema informatizado que apoie o

tratamento técnico adequado, incluindo arranjo, descrição e acesso, de forma a assegurar a manutenção da autenticidade e da relação orgânica desses documentos.

Os arquivos devem dispor de repositórios digitais confiáveis para a gestão, a preservação e o acesso de documentos digitais. A preservação dos documentos arquivísticos digitais, nas fases corrente, intermediária e permanente, deve estar associada a um repositório digital confiável. As normas e-ARQ e a Resolução 39 para RDC-Arq apresentam os requisitos para a garantia da integridade da cadeia de custódia, que deve ser ancorada num contexto arquivístico com segurança e que garanta a confiança nos registros (CONARQ, 2015).

O contexto arquivístico, que espelha as relações orgânicas, são todos os fatores ambientais que combinados determinam como documentos são criados, estruturados, gerenciados e interpretados. Os fatores ambientais que determinam diretamente os conteúdos, formas e estrutura dos registros podem ser diferenciados em contexto de proveniência, contexto administrativo e contexto de uso (THOMASSEN, 2006, p. 10). Estes fatores são determinados pelo contexto sócio-político, cultural e econômico

Thomassen (2006, p. 10) elucida que o contexto de proveniência é o contexto de produção dos documentos, e está referido a como o ente produtor dos documentos é organizado (contexto organizacional), como suas funções estão estruturadas (o contexto funcional) e como seus processos de trabalho são delineados (contexto de procedimentos administrativos).

Quadro 7 - Componentes do Contexto de Proveniência

Contexto Organizacional	Está na estrutura organizacional (a estrutura interna da organização, bem como suas relações externas), nos agentes com suas interações e mandatos;
Contexto Funcional	Consiste da missão, funções e objetivos do produtor dos documentos, as tarefas por meio das quais desempenha sua missão e as atividades que desenvolve para executar tais tarefas. O contexto funcional é o que, em maior escala, determina o conteúdo do arquivo;
Contexto de Procedimento Administrativo	(O que se refere à estrutura do processo). Pode ser decomposto em processos de trabalho, atividades, ações e transações;
Processos de Trabalho	Determinam a estrutura lógica do arquivo, atividades geram os dossiês e ações e transações geramos documentos individuais.

Fonte: Adaptação de Thomassen (2006, p. 10).

No artigo publicado em português pela extinta Associação dos Arquivistas Brasileiros (AAB), Thomassen (2006) ainda fala do

contexto de administração que é o contexto de gestão de documentos e sua conservação, cujo é constituído de todos procedimentos, métodos, conhecimento, meios e documentos com os quais o produtor assegura a disponibilidade e integridade dos arquivos. O autor relaciona o contexto de administração ao chamado contexto documentário, o qual identifica como estrutura do arquivo.

Ademais, o contexto de uso, para Thomassen (2006), consiste nos usuários, suas competências, as perguntas que fazem, e as maneiras pelas quais tentam responde-las. Este ponto destacado pelo autor aproxima a temática aos sistemas que podem suportar estas inferências deste contexto de uso. Por fim, o autor ainda lista o contexto sociopolítico, cultural e econômico, é tudo que influencia os fatores ambientais, determinando o conteúdo, forma e estrutura dos documentos arquivísticos. O autor conclui afirmando que arquivos não podem ser interpretados corretamente sem informação relacionada aos seus contextos. Informação contextual deve, portanto, ser incluída no sistema de informação do qual os arquivos formam parte.

Os contextos condicionam a formação de domínios onde os documentos são gerados, estruturados, administrados e interpretados. Cada contexto possui características de funções e atividades que são únicas. A manutenção da cadeia de custódia para estabelecer a autenticidade documental e o contexto de criação descrito e registrado podem estar acompanhados por ontologias, cuja uma aplicação é a representação de domínios, podendo tal aplicação fazer parte de uma política de preservação da informação em longo prazo.

2.5. O Modelo OAIS Utilizado Também para Preservar

Um repositório digital confiável é fundamental para assegurar a preservação, o acesso e a autenticidade de longo prazo dos materiais digitais considerados arquivísticos. Arquivos confiáveis contêm evidência confiável a respeito de decisões tomadas, direitos adquiridos e compromissos assumidos, estabelece Thomassen (2006, p. 7).

O Modelo OAIS descreve as funções de um repositório digital e os metadados necessários para a preservação e o acesso dos materiais digitais gerenciados por este repositório, que constituem um modelo funcional e um modelo de informação. Trata-se de uma estrutura conceitual para um sistema de arquivo dedicado a preservar e manter o acesso à informação digital. O modelo é composto por quatro grandes componentes fundamentais: um produtor (pessoa ou a máquina) que produz a informação a ser preservada, um gestor que estabelece uma política ou quadro para a preservação, um arquivo (uma implementação do modelo OAIS) e, por fim, um usuário (pessoa ou máquina), que interage com o arquivo com fim de obter informações (QUISBERT, 2006).

Quadro 8 – Públicos envolvidos no Modelo OAIS.

Produtor	Pessoa ou máquina que gera a informação a ser preservada;
Gestor	Estabelece a política para preservação;
Arquivo	Sistema capaz de preservar em longo prazo a informação (Modelo OAIS);
Usuário	Acessa por meio de busca ou navegação o conteúdo que precisa.

O sistema técnico inclui coleções de arquivos, bases de dados e software. A preservação em longo prazo é conseguida por intermédio da criação de estratégias com base no estado da arte da tecnologia. Isso significa que a atividade de preservação de arquivo é uma atividade contínua que pode ser afetado por mudanças de tecnologia. Estas mudanças têm um impacto apenas sobre os níveis basilares da estrutura (QUISBERT, 2006).

Conforme delinea Quisbert (2006), um pacote de informações, que é constituído de conjuntos de documentos arquivísticos, toma forma no seu momento de criação até que os registros sejam enviados para instituições de custódia para a sua preservação. A lógica de sistemas baseada no OAIS é montada com base em pacotes, como será descrito adiante. O contexto de arquivo deve cuidar, especialmente, da dimensão organizacional e pluralização de documentos, o que só é possível por meio de abertura e uso de padrões abertos. A ideia de continuidade na linha espaço-tempo só é possível com a preservação digital em longo prazo. A abertura dos sistemas e informações (por meio de padrões abertos) possibilita que os dados possam ser acessados por novas gerações de tecnologia. Desse modo, se garante a interoperabilidade na prática.

Arquivos confiáveis contêm evidência confiável a respeito de decisões tomadas, direitos adquiridos e compromissos assumidos. Arquivos devem não só assegurar que o trabalho seja feito de modo eficiente e efetivo, mas também possibilita que terceiros chequem se, e como, foi executado. Arquivos também servem para garantir responsabilidade institucional e evidência (THOMASSEN, 2006 p. 7).

O OAIS prevê seis processos (denominados de alto nível):
1) *Ingest*: entrada e processamento da informação visando

preservação; 2) *Archival Storage*: onde se mantém e recupera as informações preservadas; 3) *Data Management*: coordena informações de descrição de um objeto e sistema de informação de arquivo que é usado para apoiar operações de arquivo; 4) *Access*: ajuda o consumidor a identificar e recuperar informação; 5) *Administration*: planeja as atividades de arquivo, monitora o ambiente e avalia o desenvolvimento de novas normas e políticas; 6) *Preservation Planning*: fornece as recomendações para a conversão, migração, acompanhamento da evolução da tecnologia. Estes processos compõem a estrutura de um sistema de informação de arquivos confiável criado com base no Modelo OAIS. O Modelo de Referência OAIS teve uma influência determinante no desenvolvimento de software de preservação digital. O modelo de referência OAIS é uma estrutura conceitual para um sistema de arquivo dedicados a preservar e manter o acesso à informação digital ao longo do tempo. Este modelo é orientado a processos (QUISBERT, 2006).

O modelo é composto por pacotes de informação, que mudam o seu estado dependendo de onde eles estão no arquivo. *Submission Information Packages*, ou Pacotes de Submissão de Informação (PSI), é o pacote de objeto digital enviado para o arquivo pelo produtor informação, é como chega o conjunto documental para o arquivo. Já os *Archival Information Packages*, ou Pacotes de Arquivamento de Informação (PAI), contém todas as informações necessárias a fim de preservar o objeto digital. Todos os metadados relevantes serão adicionados e especificações técnicas para tornar o objeto. Por sua vez, o *Dissemination Information Packages*, ou Pacotes de Disseminação da Informação (PDI), é o pacote que o usuário receberá, o objeto digital e alguns de seus metadados (QUISBERT 2006).

Na fase de arquivamento os pacotes de submissão (PSI) estão preparados para a preservação. Durante este processo, os materiais digitais enviados para a preservação são conhecidos como objetos de dados de conteúdo e eles são combinados com metadados necessários para administrar a sua preservação no modelo de descrição para a preservação.

Para o OAIS, o PDI possui quatro seções, as Informações de Referência com um identificador único e persistente, as Informações de Proveniência com a história do objeto arquivado, as Informações de Contexto com a relação deste com outros objetos, por exemplo, a estrutura hierárquica de um arquivo. E por último, as Informações de Fixidez, com uma demonstração de autenticidade, como um valor *hash* por meio de assinaturas digitais, por exemplo, ou códigos criptografados.

Quisbert (2006) também expõe que o Modelo OAIS exige que o arquivo mantenha a Informação de Representação necessária para processar o pacote para sua comunidade designada. Isso pode incluir informações sobre o ambiente de hardware e software necessários para visualizar o objeto de dados de conteúdo, além de ontologias capazes de transmitir informações do domínio de conhecimento do qual faz parte. Para o autor, é necessário negociar e aceitar a informação adequada dos produtores, pois é necessário manter controle o suficiente sobre as informações e gerí-las a um nível adequado, necessário para conservação em longo prazo. Além disso, deve-se decidir quais as organizações são partes da Comunidade Designada a fim de compreender as informações recebidas. É necessário certificar-se de que a informação a ser preservada é independentemente compreensível e, dessa forma, a informação está preparada para a interoperabilidade e poderá ser acessada sem a ajuda de especialistas.

2.6. A Inerência da Interoperabilidade na Web Semântica

Tratamos neste subcapítulo os pressupostos tecnológicos para a interoperabilidade, para analisar sua relação com a organização da informação.

O processo de Interoperabilidade, segundo o “*Understanding Metadata*” da *National Information Standards Organization* (2004, p. 2), ocorre ao descrever um recurso com metadados. Promove-se, assim, a interoperabilidade que permite que seja compreendida tanto por seres humanos como por máquinas. A Interoperabilidade, para a NISO (2004) é a capacidade de vários sistemas com diferentes plataformas de hardware e software, estruturas de dados e interfaces em trocar dados com o mínimo de perda de conteúdo e funcionalidade.

Duas abordagens já tradicionais para interoperabilidade são a pesquisa em sistemas diferentes e a coleta dos metadados. Pelo modelo divulgado pela Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) os provedores de dados devem traduzir seus metadados nativos para um núcleo comum de elementos e expor isso para a colheita. Um serviço de busca, em seguida, reúne os metadados em um consistente índice centralizado para permitir a procura em diversos repositórios, independentemente dos formatos de metadados utilizados pelos repositórios participantes.

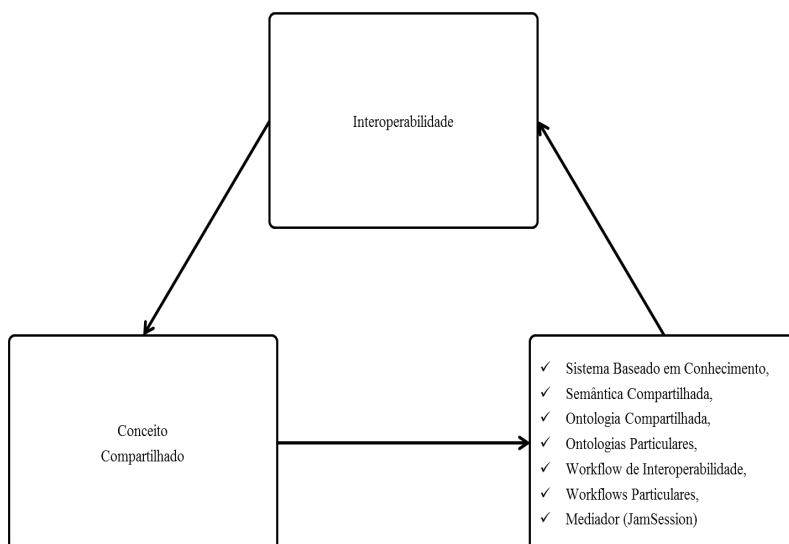
A área de tecnologia possui o conceito de Artefatos do Conhecimento Interoperáveis. Eles são baseados no conceito geral de Artefatos do Conhecimento, ou seja, são:

[...] objetos que contêm e transmitem uma representação utilizável do conhecimento. Os Artefatos do Conhecimento Interoperáveis são padrões arquitetônicos recorrentes que são observados no projeto de mecanismos de interoperabilidade para conectar instituições heterogêneas e são usados como uma descrição de alto nível da arquitetura para um projeto de sistema. (ARAÚJO, 2012, p. 29).

Este conceito é importante para se pensar a interoperabilidade como um processo de preparação da informação digital. Afinal, estes artefatos possuem linguagens onde são estruturados seus metadados e, entre estas camadas, estão ontologias. Isto é, os Artefatos do Conhecimento Interoperáveis são fundamentados sobre uma ontologia de conceitos relevantes destas instituições, cujo nível de abstração pode variar, dependendo do nível de integração necessário para as instituições - quanto mais sofisticada a interação, mais detalhes devem ser representados explicitamente na ontologia (ARAÚJO, 2012, p. 29).

A figura 1 traz o Diagrama/Esquema do Artefato do Conhecimento Interoperável, que mostra a relação da interoperabilidade, do conceito compartilhado (que não é obtido com formas simples de representação) e os chamados instrumentos de representação, que incluem as taxonomias e as ontologias.

Figura 1 – Diagrama / Esquema do Artefato do Conhecimento Interoperável.



Fonte: Araújo (2012, p. 31).

Para compreender a interoperabilidade da informação arquivística, é necessário destacar que a preparação da informação passa pela padronização com a ISAD-G e/ou NOBRADE. Afinal, conforme mencionam Alonso, Urbán e Leguina (2008, p. 14), as normas “permiten el intercambio de descripciones archivísticas gracias a que la información se encuentra estructurada en campos predeterminados”.

A interoperabilidade pressupõe a pré-disposição da informação para o intercâmbio e portabilidade, ou seja, um serviço de informação necessita de ambientes estruturados para tais operações. É possível, portanto, disponibilizar a informação arquivística principalmente se estruturada em pacotes, fornecendo dados para *harvestings* ou em mecanismos de busca, seja em

motores de mercado em um próprio sistema, ou em plataformas digitais.

Outro aspecto que reforça a interoperabilidade é a forma como são estruturados estes dados padronizados, no modo que são lidos e interpretados. Para esta propriedade se dá o nome de “Interoperabilidade Semântica”, pois como destacam Alonso, Urbán e Leguina (2008, p. 15) isso “que denota la capacidad de una información para ser interpretada, compartida e intercambiada por diferentes sistemas de tratamiento basados en el paradigma de la Web Semántica”.

A interoperabilidade é um pressuposto da Web Semântica, considerada uma evolução estrutural da Web sintática, sendo um campo importante de pesquisa e acumulando algumas décadas de concepção, estudo e aprofundamento, pois:

A partir do final da década de 1990, começaram a formalizar-se pesquisas relacionadas ao desenvolvimento de uma nova geração da Web, com o objetivo de possibilitar a incorporação de ligações semânticas aos recursos informacionais, de modo que os próprios computadores possam “compreendê-las” de forma automatizada. Machine understandable information, com esta sucinta expressão Berners-Lee, Hendler e Lassila (1998) impulsionou os primeiros estudos em direção ao projeto da Web Semântica. (RAMALHO; VIDOTTI; FUJITA, 2007, p. 1).

Conforme percebido nos aspectos mencionados, a semântica, o sentido oferecido à informação, é o que diferencia a

Web sintática da Web 3.0, é onde estão as ontologias pesadas. “Como a Internet, a Web Semântica será tão descentralizada quanto possível. [...] A descentralização implica compromissos: a Web deve atingir uma consistência total em suas interconexões [...]”, enfatiza Robredo (2010, p. 25). O desafio de estruturar para compartilhar é o que direciona os projetos de Web Semântica, ancorados no modelo de *Open Linked Data* (ou dados abertos ligados).

Deve-se planejar a disponibilização da informação por meio de plataformas digitais, pois a produção exponencial de informação na Web vem “conduzindo a um estágio em que os modelos clássicos de representação e recuperação de informações precisam ser (re)pensados sob diferentes perspectivas, pois considerando a representação como elemento fundamental, para a garantia de qualidade na recuperação”, advertem Ramalho, Vidotti e Fujita (2007, p. 1).

Em relação à Representação do Conhecimento na forma simbólica, para Vickery (1986, p. 145) é uma questão de preocupação para a área da documentação desde seu início. A estrutura de registros e arquivos em bases de dados, a estrutura de dados dos programas de computador, a estrutura sintática e semântica da linguagem natural, a representação do conhecimento em inteligência artificial e os modelos de memória humana. Para a autora, para todos estes campos há a necessidade de decidir como o conhecimento pode ser representado, de forma que as representações sejam manipuláveis.

Vickery (1986, p. 146) descreve que um símbolo corresponde a uma referência no mundo exterior, ou seja, uma entidade real ou em um mundo imaginado dentro do conhecimento público – aquele que existe dentro de estruturas de conhecimento

peçoal, em uma ou mais mentes. O conceito pode ser representado por um símbolo, como uma palavra, ou sequência de palavras. Ele também pode que pode ser expressado ou gravado (por exemplo, escrito). O símbolo representa, assim, indiretamente, o referente. O conhecimento pode ser representado por símbolos combinados de várias maneiras, onde o mais conhecido é a linguagem natural.

Cabe destacar que, no âmbito da Web 3.0 e considerando a organização da informação, parte do ferramental de representação do conhecimento, como as taxonomias ou os modelos descritivos, alimentam os computadores (os agentes inteligentes), tornando-os capazes de entender as estruturas e ontologias e, conforme aponta Robredo (2010, p. 19) “a finalidade é acrescentar uma camada de significado no topo da Web atual que faça dela menos um catálogo e mais um guia – e ainda forneça os fundamentos para que os sistemas possam raciocinar de forma semelhante aos humanos”.

2.7. Da Taxonomia à Ontologia: Ideias em Camadas de Processamento

Neste subcapítulo faz-se um esforço de relacionar os conceitos de Tecnologia da Informação, com os conceitos desenvolvidos na Ciência da Informação e na Ciência Arquivística.

A Web Semântica é estruturada pelos sistemas de organização do conhecimento existentes (taxonomias presentes no universo sintático), e são complementadas por ontologias (nos links e objetos informacionais com características semânticas dinâmicas) e,

por isso, cabe destacar o papel da padronização de ferramentas descritivas de informação, tais como as técnicas de descrição arquivísticas realizadas e disponibilizadas em ambientes digitais. A interoperabilidade conta com esses recursos para funcionar com êxito.

Com base neste cenário, as linguagens documentárias (como as taxonomias) são elementos importantes que operam combinados, sinergicamente, com outros aspectos da Web 3.0, e fazendo parte de uma de suas camadas. Isso já estava previsto na formulação do conceito da Web Semântica que foi concebido como um conjunto de tecnologias relacionadas. Em 2000, o W3C divulgou a primeira proposta de arquitetura para Web Semântica, no qual destaca que:

Com base em uma série de camadas sobrepostas, onde cada camada ou tecnologia deveria obrigatoriamente ser complementar e compatível com as camadas ao mesmo tempo em que não deveria depender das camadas superiores, possibilitando assim uma estrutura idealmente escalável, que indicasse os passos e as tecnologias necessários para a concretização do projeto Web Semântica. (RAMALHO; VIDOTTI; FUJITA, 2007, p. 2).

Faz-se imprescindível que se criem instrumentos, tais como as ontologias, que forneçam os sentidos lógico e semântico das informações aos computadores, sendo o uso de linguagens documentárias uma das camadas que interagem e os valores aplicados por meio de links semânticos, "estes instrumentos são os links semânticos" declaram Ramalho, Vidotti e Fujita (2007, p. 2). É a estrutura multinível disponível para a Web Semântica que possibilita a criação de sentido aos links e, conseqüentemente, proporciona

uma melhor experiência na recuperação da informação e acesso ao usuário. Conforme dados do W3C, compõem esta estrutura:

Quadro 10 - Estrutura de recursos de linguagens da Web Semântica.

Formatos	Definições
XML (eXtensible Markup Language)	Conjunto de regras para estruturar documentos. Não impõe restrições semânticas, é um código de linguagem de máquina.
XML Schema	É uma linguagem de regras de validação que restringe a estrutura de documentos em XML.
RDF (Resource Description Framework)	É um modelo de dados para representar objetos e relacionamentos na internet. Possui semântica simplificada.
RDF Schema	Vocabulário para descrever propriedades e classes expressas em RDF; semântica para generalização.
OWL (Web Ontology Language)	É uma linguagem para definir e instanciar ontologias na <i>Web</i> . Maior vocabulário para descrever classes e propriedades: relacionamentos, cardinalidade, igualdade, propriedades, características de propriedades e enumeração dos elementos de uma classe.
SKOS (Simple Knowledge Organization System)	É um modelo de dados comum que partilha a ligação dos sistemas de organização do conhecimento por meio da Web, tais como tesouros, taxonomias, esquemas de classificação e sistemas de cabeçalhos de assuntos, que compartilham uma estrutura semelhante, e são usados em aplicações similares.

Fonte: W3C Brasil (2015).

Na comparação entre a organização da informação da Web sintática com a Web 3.0, destaca-se a importância da ontologia, pois, para que a camada ontológica *Web Ontology Language* (OWL) possa repassar conhecimento, é necessário contextualizar a informação dentro de um cenário concreto. É nesta camada onde se define uma ontologia. A ontologia apresenta conceitualizações de determinado domínio, pois:

Onde o termo "ontologia" refere-se a uma especificação explícita de uma conceitualização. [...] O desenvolvimento de ontologias permite um modelo de um domínio de conhecimento a ser construído em que uma taxonomia acordada, ou sistema de classificação de entidades como um repositório de sentidos, é coberto com descrições explícitas das relações que essas entidades compartilham. (LUMSDEN; HALL; CRUICKSHANK, 2011, p. 247).

Já para Smith (2008), quando se compara as diversas ontologias e de determinados campos, destaca que, para a Ciência da Informação a ontologia pode ser uma aplicação, um dicionário de termos formulados em uma sintaxe canônica, com definições comumente aceitas, projetadas para produzir um quadro lexical ou taxonômico para o conhecimento-representação, e que pode ser compartilhado por diferentes comunidades de sistemas de informação.

O princípio classificatório e de capacidade de linguagem faz parte da ontologia, portanto, “distingue-se das ciências especiais, não só na sua generalidade radical, mas também em seu objetivo ou o foco: não visa predicação, mas sim taxonomia”, diz Smith (tradução nossa, 2008, p. 158). É possível, pelo exposto, inferir que uso similar

se faz de uma linguagem documentária, e que pode se fazer também para uma ontologia pesada, estruturando no caso da semântica, os conceitos, axiomas e restrições de um determinado domínio de conhecimento organizado.

Ao considerar, portanto, a carga semântica apresentada pelas ontologias, combinadas com as características sintáticas dos metadados, deve-se considerar os elementos descritivos propostos pela NOBRADE como uma forma de identificar os links e os seus respectivos registros, como exposto no capítulo anterior deste estudo. Como visto, o resultado da descrição estaria fornecendo informação arquivística padronizada para o contexto 3.0, desde que seja:

[...] expresso em RDF, que o codifica em conjuntos de sujeito, verbo e predicado, sendo cada um desses elementos identificados com URIs (Universal Resource Identifier). Os URIs vão permitir a utilização desses sujeitos e predicados como ligações em qualquer página Web e que qualquer pessoa defina um novo verbo ou conceito na Web. (BERNERS-LEE; HENDLER; LASSILA, 2001, p. 2).

Em analogia, pode-se afirmar que os Identificadores Uniformes de Recursos, ou URI - *Uniform Resource Identifier* (em inglês), da Web 3.0 possuem elementos descritivos, onde suas fichas de metadados que descrevem e dão sentido a um elemento do objeto informacional (aqui usado como sinônimo de registro). Este conjunto, a que se referem Berners-Lee, Hendler e Lassila (2001, p. 2), pode ser considerado como uma ficha descritiva, com as relações representativas criadas pelos elementos descritivos arquivísticos, aplicados aos links.

Para Alves (2010) os metadados são atributos que representam uma entidade, sendo um objeto do mundo real representados em um sistema de informação. Em outras palavras, são elementos descritivos ou atributos referenciais e codificados que representam características próprias ou atribuídas às entidades. Em sua pesquisa, a autora afirma que os metadados podem, ainda, serem considerados dados que descrevem outros dados em um sistema de informação, com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação.

A representação da informação é uma característica intrínseca dos metadados. Já quanto aos padrões de metadados, Alves (2010, p. 47) exemplifica que são estruturas de descrição constituídas por um conjunto predeterminado de metadados. São, portanto, atributos codificados ou identificadores de uma entidade que são metodologicamente construídos e padronizados. O objetivo do padrão de metadados é descrever uma entidade, gerando uma representação unívoca e padronizada que possa ser utilizada para sua recuperação. Esta é uma possível abordagem arquivística para a Web Semântica, considerando a criação de ontologias leves, originadas por processos descritivos normatizados, que evidenciam a informação arquivística e as padronizam em forma de metadados possibilitando sua interoperabilidade, por meio de plataformas e diferentes sistemas digitais, ao longo do tempo, gerando as Ontologias Digitais Arquivísticas e que aprofundaremos na parte III deste livro.

PARTE III: METADADOS E ONTOLOGIAS

Avalia-se a interoperabilidade por meio de metadados definidos em processo de estruturação de informação, para identificar e estruturar a informação arquivística para a rede mundial de computadores.

É possível criar padrões de interoperabilidade por meio da descrição arquivística, os modelos EAD e EAC podem ser os formatos padronizados para a interoperabilidade e possuem potencial para evidenciar a informação arquivística. A descrição como uma das principais funções arquivísticas, esta indica de que forma deve-se tratar o resultado da descrição como potencial ontologia.

Verifica-se as ontologias como estruturas de domínios de conhecimento. A partir disso relaciona-se as ontologias e padrões arquivísticos para apurar a existência de Ontologias Digitais Arquivísticas.

3. A Interoperabilidade Semântica por meio de Metadados

Neste capítulo que abre o terceiro flanco do estudo, relaciona os padrões de metadados e a existência de dados abertos ligados estruturados em XML e o padrão *Dublin Core* como potenciais carregadores de informação arquivística disponível na rede.

Os arquivos e outros serviços de informação sempre produziram metadados por meio de seus catálogos, dossiês, índices, e outros instrumentos descritivos. É indissociável do trabalho dos serviços de informação a geração de dados sobre as informações que possuem em seus acervos. Agora que estão digitalizados, possuem potencial de se tornarem interoperáveis.

De fato, para Alves (2010) o termo metadados está geralmente associado com a definição de padrões de metadados. Existe, por assim dizer, uma relação tácita entre os metadados e os padrões. Para que eles possam existir, os metadados (*metadata*) devem estar codificados em estruturas padronizadas de descrição, denominadas padrões de metadados (*metadata statement*). O conjunto de metadados ou elementos de metadados (*element sets*) irá compor o esquema de metadados (*metadata schema*) do padrão de metadados (ALVES, 2010, p. 47).

Um metadado ou atributo irá representar uma única característica da entidade, enquanto que o conjunto de metadados ou o conjunto de atributos presentes em um padrão irá representar a entidade no todo. É o conjunto de metadados que irá identificar e individualizar a entidade de forma única e inequívoca entre as demais, pressupondo a autenticidade daquele objeto informacional, pois devem ser capazes de identificá-la de forma absoluta garantindo a sua unicidade. Logo, os metadados e seus padrões estão

relacionados com o tratamento descritivo da informação, pois apoiam a descrição e a representação padronizada do recurso informacional, independentemente do tipo de suporte, com o objetivo de facilitar a busca e a recuperação, (ALVES, 2010 p. 48).

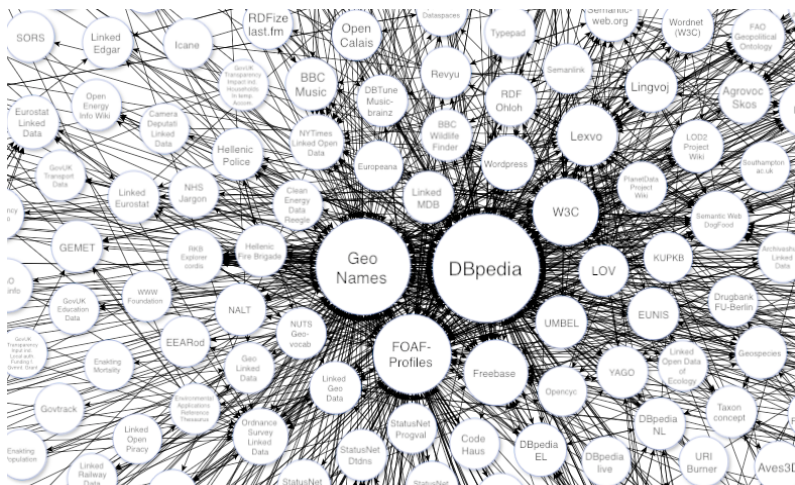
Para difusão, há movimentos internacionais como os de dados abertos ligados, ou *Linked Open Data* (LOD) que é um movimento internacional que prega a abertura dos dados para que sejam acessíveis na rede de computadores *World Wide Web* e que carregam sentido por intermédio de seu esquema de metadados.

As iniciativas que unificam esses metadados são recentes para a área de arquivos, como o projeto *Repository for Linked Open Archival Data* (ReLoad), que:

É um projeto realizado pelo Estado Central, pelo Instituto do Patrimônio Cultural de Emilia Romagna e por Regesta.exe com o objetivo de testar os métodos de tecnologias da Web Semântica e padrões para dados abertos ligados (LOD) para facilitar o compartilhamento de informações de arquivo a partir de uma variedade de fontes. (RELOAD, 2014, tradução nossa).

O movimento *Link Open Data* possui diversas iniciativas e práticas em diversas áreas de conhecimento humano, como pode ser observado na Figura 2. Essa padronização é tanto dos padrões de metadados, como de ontologias, pois é cerne da ontologia servir como padrão interoperável, assim como os tesouros e as taxonomias online estão cada vez mais interoperáveis.

Figura 2 - Parte da Estrutura de Ontologias de Link Open Data.



Fonte: Hasso Plattner Institute (2014). *The Linking Open Data cloud diagram*.

A operação base da interoperabilidade é a identificação e o uso da informação adicionada sobre o link, ou a metainformação (metadados) do objeto informacional, que cria formas de descrever o objeto. Como afirma Patrício (2012):

O desafio da Web Semântica é propiciar uma linguagem que expresse dados e regras lógicas para raciocinar sobre esses dados, de forma interoperável na Web. “Adicionar lógica à Web significa utilizar regras para fazer inferências, escolher ações e responder a questões. As tecnologias da Web Semântica são o XML e o RDF. O XML permite que qualquer pessoa adicione etiquetas para anotar as suas páginas Web ou secções de texto nessas páginas. (PATRICIO, 2012).

Logo, os metadados são representações de informação digital associados a um recurso online, possibilitando seu tratamento técnico. Com base em Pires (2012), é possível afirmar que os metadados podem "ser descritos separadamente em um sistema de armazenamento de informações (bancos de dados), como podem estar embebidos nos arquivos digitais" e, portanto, a padronização e a normatização descritiva garantem dados semelhantes para troca e interoperabilização.

À essa visão da união metadados mais o padrão e/ou norma, que resulta em conteúdo descritivo comum a ser interoperado, soma-se a ideia de que o modelo de tecnologias de dados interligados serve, de acordo com Marcondes (2012, p. 174) para "interligá-lo com outros que lhe agreguem valor semântico, cultural, cognitivo, econômico ou científico".

Assim, alguns recursos de estruturação da Web Semântica podem ser o modelo XML e o esquema de metadados *Dublin Core* (para padronizar os dados da Web). Há um bom tempo, um outro importante passo foi o desenvolvimento da XML (*eXtensible Markup Language*), que enfatiza não mais a aparência, mas o conteúdo das páginas Web. Uma segunda, e igualmente fundamental contribuição, vem da iniciativa *Dublin Core* metadata. Estes dois recursos tecnológicos são básicos para a interoperabilização de dados (TRZESNIAK; KOLLER, 2005, p. 188).

Soma-se a isso o modelo de expansão, ditado pelo movimento *Open Archives*, ou Arquivos Abertos, que possibilita o intercâmbio, a coleta e a distribuição dos conteúdos presentes em metadados estruturados.

A terceira componente é devida ao movimento dos arquivos abertos (Open Archives Initiative, OAI:

www.openarchives.org): consiste no protocolo para coleta de metadados (Protocol for Metadata Harvesting, PMH), através do qual toda a informação na Web, que esteja adequadamente descrita por metadados, pode ser coletada e apropriadamente recuperada de repositórios construídos por provedores de dados. (TRZESNIAK; KOLLER, 2005, p. 188).

Patrício (2012, p. 3) estabelece que "o verdadeiro potencial da Web Semântica será realizado quando forem criados agentes de software capazes de coligir conteúdo Web de diferentes fontes, processar essa informação e trocar os resultados com outros programas". Assim, os computadores coletarão os dados e conseguirão, por meio da semântica, oferecer conteúdo de diversas bases de forma assertiva, de acordo com as necessidades informacionais de usuários. Espera-se ver este modelo aplicado à informação arquivística.

Nesse sentido, objetiva-se a interoperabilidade semântica, que segundo Sayão e Marcondes (2008), é a que:

[...] está relacionada com o significado ou semântica das informações originadas de diferentes recursos e é solucionada pela adoção de ferramentas comuns ou/e mapeáveis de representação da informação, como esquemas de metadados, classificações, tesouros e mais recentemente, ontologias.

Identifica-se como uma possível forma de projetar a interoperabilidade semântica e, portanto, estimular a Web semântica composta por Arquivos e a identificação de informação orgânica disponível na rede. A forma de realizar isto seria por meio da comparação dos elementos descritivos criados por padrões de interoperabilidade padronizados pela descrição com padrões

utilizados e difundidos para a web, alcançando a identificação de informação com potencial arquivístico na Web.

3.1. Criação de Padrões de Interoperabilidade por meio da Descrição

Neste subcapítulo é analisada a descrição arquivística como padrão para a interoperabilidade e o seu potencial para evidenciar a informação arquivística, além de representar uma das principais funções, indica de que forma deve-se tratar o resultado da descrição como potencial ontologia.

Como visto, a descrição arquivística é um dos pilares da Arquivologia, principalmente por conter uma das suas principais funções. Disponibiliza aos arquivistas e aos usuários dos serviços de arquivo as informações sobre os documentos, sem a necessidade do manuseio (ou corruptibilidade) destes, colaborando com a questão da preservação. Portanto, é utilizada com destaque nos acervos permanentes, como o instrumento de indicação da estrutura destes acervos.

Hagen (1998) afirma que “em todos os casos, o trabalho do arquivista é representar ideologicamente as informações contidas nos documentos”, ou seja, a descrição obriga a repassar aos interessados uma ideia do que trata um conjunto de documentos. A forma de acesso às representações ideológicas criadas é que varia, e se dá por intermédio de instrumentos de pesquisa, ou instrumentos de acesso à informação, que padronizam a organização de informações (RIBEIRO, 1998).

A descrição arquivística gera estes instrumentos de pesquisa, que são os produtos de informação e que servem para

localizar objetos informacionais e descrever seu conteúdo, e como o autor exalta:

O modo como esses instrumentos de acesso foram sendo elaborados, os critérios seguidos na sua concepção e a informação descritiva (hoje designada por metadata) que continham assentavam num costume que se foi impondo, numa prática que se foi apurando e em dois objetivos essenciais: controlar fisicamente a localização dos documentos e informar sobre as suas características e, por vezes, sobre o seu conteúdo (a informação). (RIBEIRO, 2005, p. 3).

Assim sendo, neste conceito, a descrição opera desde a gênese documental, isto é, quando é realizada a função de criação e produção, que traz em seu componente a padronização de elementos de descrição, sua classificação e os fluxos informacionais. Sendo um processo analítico “es el analisis realizado por el archivero sobre los fondos y los documentos de archivo agrupados natural o artificialmente, a fin de sintetizar y condensar la información en ellos contenida para ofrecerla a los interesados”, expõe Heredia Herrera (1991, p. 299). A autora ainda destaca a falta de definições nos manuais mais divulgados (na França e no Canadá), pois o termo, apesar de representar uma tarefa realizada no mundo todo, ainda não é um conceito adotado de forma universal.

Segundo Hagen (1998) “Lodolini, em um trabalho de 1984, dedica um capítulo à atividade de descrição, sem usar este termo”. Dessa forma, podemos entender que, apesar de ser uma atividade básica da arquivística no tratamento de arquivos, o termo “Descrição Arquivística” - cunhado por Schellemborg -, mostra um conceito em definição para a Arquivologia internacional.

O próprio Schellemborg (2004) afirma que a descrição deve ser considerada uma tarefa profissional. Ao realizar esse trabalho, o arquivista se inteira da procedência, do conteúdo, do arranjo e do valor dos documentos. Esses dados são registrados em instrumentos de busca que tem duplo propósito: 1) tornar os papéis conhecidos às pessoas que possam vir a se interessar, pelo menos; e 2) facilitar a pesquisa ao arquivista.

Contudo, não existe uma padronização mundial quanto aos termos dos instrumentos e quanto aos métodos empregados na realização de descrição, tanto que Cruz (1996) deixa claro que há acordos a respeito dos objetivos e dos princípios da descrição, mas não é possível dizer o mesmo quanto aos tipos de instrumentos e ao seu processo de elaboração.

A palavra descrição, em termos amplos, como aborda Heredia Herrera (1991, p. 300), “é a enumeração das qualidades e elementos fundamentais de uma pessoa ou objeto”, - visão também compartilhada por Schellemborg (2004). A autora ainda discorre sobre o processo de descrição afirmando que “é a análise realizada pelo arquivista sobre os fundos e os documentos de arquivos agrupados natural ou artificialmente, a fim de condensar a informação neles contidas, para oferecê-las aos interessados. Equivale dar aos documentos de arquivo e suas agrupações seus ‘números de identidade’”.

Pode-se notar que Heredia Herrera (1991) não distingue documento e fundo, atitude comum entre os autores da época. Schellemborg (2004), por exemplo, corrobora com esse pensamento quando afirma que “as técnicas descritivas ideadas pela arquivística norte-americana reportam-se principalmente às grandes e pequenas unidades de arquivo, isto é, aos grupos ou coleções e aos itens

singulares”. Isto demonstra os pensamentos clássicos da Arquivística, quando se aproxima muito da tarefa de um arquivista com um bibliotecário, que descreve cada unidade de conhecimento (um livro, uma revista ou um documento, por exemplo). Bellotto (1991), aproximando a Arquivística da atividade historiográfica, destaca que “a descrição feita no miúdo, isto é, a que incide diretamente sobre o documento unitário, não levando em conta seu meio orgânico, dificilmente revelará ao historiador a real significação do material analisado”.

Podemos vislumbrar nos autores clássicos da área duas vertentes: uma que afirma que a descrição serve tanto para indicar grandes acervos ou o documento; e outra, mais preocupada em repassar significação aos pesquisadores de uma determinada área (história, por exemplo). Ou seja, há os que se aproximam da biblioteconomia e consideram a atividade descritiva comparando-a a uma atividade de descrição de unidades, enquanto àqueles que aproximam a área do campo da história refutam tal ideia. Independente disso, a normatização da área que ocorreu após este debate, mostrou que a descrição tem característica multinível, isto é, podendo ser descrito do fundo (a área mais ampla de descrição) chegando às unidades documentais (o objeto mais restrito).

A descrição como função arquivística é aquela que cria o conteúdo que é acessado pelo pesquisador, informando sobre sua utilidade. Ela prepara a informação para uso, criando a forma como será comunicado aquele acervo e seu potencial informativo. Heredia Herrera (1991, p. 299) defende que a descrição, em definitivo, é o meio utilizado pelo arquivista para obter informações contidas nos documentos e disponibiliza-los aos usuários. Segundo a autora, “os instrumentos de descrição são as representações dos documentos ou de suas agrupações, desde que os transformem mediante uma forma

distinta do original”. Então, vê-se que a descrição representa as informações mais importantes, sem ser uma transcrição *ipsis litteris* do documento. Trata-se da representação da informação arquivística.

A mesma autora também afirma ainda que “a análise aplicada aos documentos deve ser feita com objetividade para respeitar sua natureza que oferece esta anotação substancial”. Para atingir seus objetivos, a descrição deve ser:

Exacta: en cuanto que os documentos no son algo impreciso, sino testimonios únicos y concretos.
Suficiente: par la unidad que se está informando (archivo, fondo, serie o documento), sin ofrecer más de lo necesario, por exceso o por defecto, Oportuna: en cuanto que ha de reflejar una programación que marque una jerarquía de la información. (HEREDIA HERRERA, 1991, p. 301).

A descrição arquivística, aplicada por intermédio de modelos como a ISAD-G e a NOBRADE, é um processo em que o arquivista cria representações de um determinado acervo arquivístico, explicitando o contexto e conteúdo deste acervo num instrumento de pesquisa ou de referência. Conforme destacam Andrade e Silva (2009, p. 6), os “instrumentos arquivísticos de referência são os produtos do processo de descrição arquivística, que se ocupam de criar representações para o acervo ou parcelas deste”, ou seja, a estruturação da informação presente nos arquivos representada digitalmente, conhecida como informação arquivística, é foco deste processo da função arquivística.

Se há uma certa concordância de que a descrição é uma atividade necessária nos arquivos permanentes, o mesmo não vemos em relação aos instrumentos de pesquisa. Couture e Rosseau (1998) tentam dar ordem a essa grande gama de instrumentos, seguindo

suas visões, hierarquizando estes resultados da descrição arquivística com relação às unidades de trabalho nos sistemas de arquivos, indo do mais amplo (patrimônio arquivístico comum entre os estados (Canadá e Reino Unido, no caso), até o mais específico como o dado, componente unitário de um documento. Tal esforço visa facilitar ao utilizador a buscar de forma correta as informações, além de ordenar toda a disponibilidade de instrumentos disponíveis. O quadro a seguir apresenta a análise de Couture e Rosseau (1998) acerca dos instrumentos de pesquisa disponíveis e mostram que o desafio é a padronização da informação arquivística, considerando sua estrutura multinível.

Quadro 10 - Níveis Descritivos e Instrumentos de Pesquisa.

Nível Descritivo	Instrumentos de pesquisa
Patrimônio Arquivístico Comum	• Guia dos arquivos das organizações internacionais • Anuário internacional dos arquivos • Bibliografia internacional dos anuários e guias de serviços de arquivo
Arquivos de Estado	• Anuário dos serviços de arquivo • Guia por serviço de arquivo
Arquivos de um Organismo Documental	• Catálogo dos fundos • Estado geral dos fundos • Guia do serviço de arquivo • Bibliografia dos instrumentos de descrição documental • Índice geral
Grupo de Arquivos	• Guia por grupo de arquivos ou subgrupos de arquivos
Fundos	• Guia Temático • Guia por fundo • Tabela de equivalência

Séries Documentais	● Guia por série ● Guia por subsérie ● Guia por subsubsérie
Unidades de Instalação	● Repertório sumário ● Repertório numérico simples ● Repertório numérico detalhado ● Repertório cronológico
Peça	● Inventário sumário ● Inventário analítico
Documento	● Índice onomástico/geográfico ● Índice Assunto/matéria ● Índice cronológico
Dado	● Índice ● Lista

Fonte: Adaptado de Couture e Rosseau (1998).

O rol de instrumentos demonstra a diversidade do trabalho do arquivista na descrição e o desafio de criar uma informação exata, suficiente e oportuna no nível que está sendo descrito. A padronização da descrição arquivística é um fato recente para a área, contemporânea ao conceito de Web 3.0 e focada no processo de criação de instrumentos descritivos, que serve para dar acesso e permitir a pesquisa nos acervos. Conforme aponta do Conselho Internacional de Arquivos (CIA), esse conjunto de regras gerais para a descrição arquivística faz parte de um processo que visa:

a) assegurar a criação de descrições consistentes, apropriadas e autoexplicativas; b) facilitar a recuperação e a troca de informação sobre documentos arquivísticos; c) possibilitar o compartilhamento de dados de autoridade; e d) tornar possível a integração de descrições de diferentes arquivos num sistema unificado de informação. (CIA, 2000, p. 11).

Existem quatro regras que devem ser aplicadas para estabelecer a relação hierárquica entre as descrições, segundo o Conselho Internacional de Arquivos (2000), são a (1) descrição do geral ao particular que apresenta uma relação hierárquica entre as partes e o todo; (2) informação relevante para o nível de descrição onde as informações devem ser apropriadas para o nível que está sendo descrito; e (3) relação entre descrições que identifica o nível de descrição; e (4) não repetição de informação, ou seja, não repetir as informações em níveis diferentes de descrição.

A norma de descrição que deve ser utilizada no Brasil é a NOBRADE, que é resultante da adaptação da norma mundial ISAD-G. Ao entender a estrutura de uma, entende-se a estrutura da outra, pois ambas possuem abordagem multinível de estruturação, ou seja, conforme descreve Marcondes (2012, p. 183) “a estrutura hierárquica de fundos, séries e dossiês, avalizada pela norma ISAD-G, está baseada no princípio da proveniência, caro à gestão arquivística; acervos arquivísticos são mantidos obedecendo a esse princípio”. Nesse caso, a abordagem multinível possibilita descrever desde a instituição mantenedora aos detalhes do acervo, do produtor ou acumulador do acervo ao detalhe do dossiê onde são organizados os principais temas do arquivo.

Deve-se considerar todo o processo longo e revisado por pares de formação das primeiras regras de descrição de arquivos, inéditas até o fim da década de 90 do século passado:

Normas de descrição arquivística são baseadas em princípios teóricos aceitos. Por exemplo, o princípio de que a descrição arquivística procede do geral para o particular é uma consequência prática do princípio de respeito aos fundos. Este princípio deve ser claramente enunciado caso se deseje construir uma estrutura de

aplicação geral e um sistema de descrição arquivística, manual ou automático, não dependente de instrumentos de pesquisa. (CIA, 2000, p. 2).

A falta de padrão de descrição e organização pode isolar o conteúdo e a informação digital gerados nos serviços de informação, como museus e arquivos, pois os mesmos, muitas vezes, possuem uma abordagem singular, e cada serviço destes possui um contexto singular e, portanto, para a realidade de dados abertos interligados (*Linked Open Data*), devemos considerar que:

[...] Os conteúdos digitais gerenciados por sistemas de gestão de conteúdos de arquivos, bibliotecas e museus seguem padrões e formatos bastante específicos, que os tornam pouco interoperáveis e pouco suscetíveis de serem integrados a outros conteúdos diferentes, tais como música, vídeos, fotos etc. Ao acessar esses catálogos na Web, o usuário fica como que prisioneiro desses contextos sistêmicos e institucionais específicos, praticamente sem possibilidades de navegar de fora para dentro ou de dentro para fora destes. (MARCONDES, 2012, p. 182).

Para alavancar a interoperabilidade e a Web Semântica é necessária a padronização de tecnologias, de linguagens e de metadados descritivos, de forma que todos os usuários da Web obedeçam a “determinadas regras comuns e compartilhadas sobre como armazenar dados e descrever a informação armazenada e que esta possa ser “consumida” por outros usuários humanos ou não, de maneira automática e não ambígua”, destacam Souza e Alvarenga (2004, p. 134). Para estes autores, o projeto da Web Semântica, em sua essência, é a criação e implantação de padrões tecnológicos que não somente facilite as trocas de informações entre agentes

peçoais, mas principalmente estabeleça uma língua franca para o compartilhamento mais significativo de dados entre dispositivos e sistemas de informação de uma maneira geral.

A utilização de uma regra de padronização de descrição, como a ISAD-G, disponibiliza de forma padronizada os dados descritivos dos acervos e dos geradores de acervos. A descrição multinível, utilizado como padrão para a descrição de acervos arquivísticos pode enriquecer o acesso à informação, pois fornece dados tanto da instituição que possui o acervo como dos documentos. As versões Web da descrição arquivística, representada pelos padrões EAD e EAC, consolidam esta visão.

A utilização de padrão arquivístico de descrição como ISAD-G (estruturado digitalmente em EAD e EAC) aumenta a visibilidade da informação arquivística pois pode ser empregado como um padrão de definição de elementos a serem processados pela Web Semântica. Portanto condicionam a formação de ontologias arquivísticas.

A ISAD-G possui elementos, organizados por áreas que agrupam os elementos descritivos por proximidade de uso. Para este trabalho não é necessário detalhar cada um e, para isso, recomenda-se a leitura da norma. Porém, é preciso listá-las a fim de saber a indicação de sua utilização. Mais adiante aborda-se a EAD e EAC.

Quadro 11 - Áreas Informacionais e Elementos Descritivos ISAD-G.

Áreas informacionais	Elementos descritivos
1 - Área de identificação:	Códigos(s) de referência; título; data(s) Nível de descrição; Dimensão e suporte.
2 - Área de contextualização:	Nome(s) do(s) produtor(es); História administrativa/biográfica; História custodial e arquivística; Fonte imediata de aquisição ou transferência.
3 - Área de conteúdo e estrutura:	Âmbito e conteúdo; Avaliação, seleção e eliminação; Ingresso(s) adicional(ais); Sistema de organização.
4 - Área de condições de acesso e de uso:	Condições de acesso; Condições de reprodução; Idioma/Escrita; Características físicas e requisitos técnicos; Instrumentos de descrição.
5 - Área de fontes relacionadas:	Existência e localização de originais; Existência e localização de cópias; Unidades de descrição relacionadas; Notas de publicação.
6 - Área de notas:	Notas.
7 - Área de controle da descrição:	Nota do(s) arquivista(s); Regras ou convenções; Data(s) da(s) descrição(ões).

Fonte: Conselho Internacional de Arquivos (2000). ISAD-G.

A NOBRADE possui, além dos elementos da ISAD-G, mais uma Área (a área 8 – Pontos de Acesso) e mais dois Elementos de

Descrição (os elementos 6.2 – Notas sobre conservação e o 8.1 – Pontos de acesso e indexação de assuntos).

No caso brasileiro, ao aplicar a NOBRADE, além de aplicar um modelo internacional já adotado por outras iniciativas de *Open Archives* (pois a NOBRADE contém a ISAD-G), amplia-se o potencial de interoperabilidade de dados, pois o oitavo elemento da NOBRADE abre a possibilidade do uso de vocabulários e ontologias leves, enriquecendo a indexação e melhorando a relevância na recuperação da informação.

Considera-se que os metadados de uma forma em geral não apenas descrevem uma entidade como também a contextualizam em um sistema de informação, garantindo seu gerenciamento. Para Alves (2010, p. 51) eles podem auxiliar na contextualização do recurso no sistema, também mantém as relações entre recursos informacionais e possibilitam a distinção entre versões múltiplas de um mesmo recurso. Além disso, para a autora, os metadados permitem representar questões legais, de privacidade, os interesses de propriedade, restrições de uso, direitos autorais entre outras. Assim, vê-se que possibilitam conservar informações que caracterizam os recursos em atualizações de sistemas, entre outras vantagens. Alves (2010, p. 51) ressalta que “logo, considera-se que os metadados não apenas descrevem uma entidade como também a contextualizam em um sistema de informação, garantindo seu gerenciamento”. Tal assertiva valida o posicionamento deste estudo.

O uso de padrões aplicados para a descrição de arquivos ressalta uma mudança de atuação dos profissionais da área, e da necessidade de adotar mudanças conceituais que aprofundam o controle da estruturação das informações (ontologias), a economia de recursos (utilizar padrões), o intercâmbio de dados (XML e RDF), a

melhoria da qualidade metodológica (ISAD-G, EAD e EAC), além da definição de estratégias e de políticas institucionais e de relacionamentos mais amplos, visando o compartilhamento das informações com instituições nacionais e internacionais. Para a Web Semântica, há como manter a custódia (posse da informação), partilhando os metadados e fornecendo o acesso aos objetos informacionais digitais abertos.

Para a *National Information Standards Organization* (NISO, 2004, p. 2), há uma preocupação crescente de que os recursos digitais não irão sobreviver em formato utilizável no futuro. A informação digital é frágil, afinal ela pode ser corrompida ou alterada, intencionalmente ou não. Sua mídia de armazenamento pode tornar-se inutilizável, assim como os hardwares e tecnologias de softwares mudam. Estratégias de preservação da informação digital são os formatos de migração e a emulação de hardware atual e comportamento do software em futuras plataformas de hardware e software.

Os metadados são fundamentais para assegurar que os recursos vão sobreviver e continuar a ser acessíveis no futuro. Segundo o órgão, o arquivamento (definitivo) e a preservação exigem elementos especiais, como rastrear a linhagem de um objeto digital (de onde veio e como ele mudou ao longo do tempo), ao detalhe de suas características físicas. É necessário também documentar o seu comportamento, a fim de imitá-la (por meio da emulação) em tecnologias futuras, o que é condizente com a diretiva de manter a fidedignidade e a autenticidade documental por parte dos arquivistas (NISO, 2004, p. 2).

Assim, ao aplicar a descrição arquivística, sabemos que seu produto pode gerar informação arquivística, que de fato é

resultado da estruturação de um conjunto documental de mesma origem orgânica, ou um fundo, esta é a

[...] informação orgânica registrada, principalmente em sua manifestação estruturada e em seu conjunto, quais sejam os documentos e os fundos arquivísticos e, também, o papel dos arquivos como instituição (arquivos públicos, arquivos institucionais) na preservação e na concessão de acesso às informações (SANTOS, 2011, p. 118).

A informação orgânica de séries, subséries, seções e subseções, além da própria informação de fundo e de autoridade representam o que pode ser determinado como Informação Arquivística e pode estar representado em instrumentos arquivísticos como guias e inventários. Por se tratar de um conjunto descritivo multinível, como previsto nas normas descritivas, fala-se, portanto, do que é fora do documento, o contexto, representado como produto de sua descrição arquivística.

Esquemas de metadados são conjuntos de elementos de metadados projetados para uma finalidade específica, como a descrição de um tipo particular de um recurso de informação. Segundo expõe a NISO (2004, p. 2), a definição ou o significado dos próprios elementos é conhecido como a semântica do esquema. Os valores fornecidos aos elementos de metadados são o conteúdo. Esquemas de metadados geralmente especificam nomes de elementos e sua semântica. Opcionalmente, podem especificar regras de conteúdo, como por exemplo o conteúdo deve ser formulado (por exemplo, como identificar o título principal), regras de representação de conteúdo (por exemplo, regras de capitalização), e os valores de conteúdos permitidos (por exemplo, os termos devem ser usados a partir um vocabulário controlado específico).

Logo, é essa informação que pode ser identificada nos links da Web 3.0, por intermédio da aplicação da descrição arquivística e buscar por seus metadados. A descrição arquivística gera, para cada registro descrito, um conjunto de informações ou metadados. Esses elementos descrevem o que pode ser considerada a informação arquivística do acervo, da seção, da série, do dossiê ou da unidade documental.

A informação arquivística é abordada como aquela sendo resultado da descrição multinível dos arquivos e das instituições que os organiza institucionalmente, e utiliza-se dela como objeto de trabalho. Tal qualificação da informação é uma característica dos arquivos contemporâneos e, principalmente, dos digitais, do ponto de vista prático, no qual é possível verificar que:

A metodologia arquivística atual frequentemente toma o contexto, e particularmente o contexto de proveniência, como ponto de partida para análise. Nesta abordagem, primeiro são analisadas a missão, funções e tarefas do produtor de documentos, os agentes e seus mandatos são mapeados e o sistema de arquivos é criado, ou reconstruído. Esta abordagem analítico-funcional é mais ampla do que a abordagem descritiva clássica e também é mais adequada para a análise de um arquivo ainda em crescimento, de um arquivo que ainda será criado e para arquivos muito extensos ou digitais, os quais não podem ser analisados documento por documento. (THOMASSEM, 2006, p. 15).

Ao aplicar a "metodologia arquivística atual" - abordagem proposta por Thomassem -, que parte da avaliação do contexto orgânico e serial para definir seus padrões, verifica-se que o resultado da aplicação dessa visão é identificado como informação

arquivística nesta pesquisa, ou com informação arquivística interoperável. Essa abordagem faz parte das análises e proposições, compondo a visão de interoperabilidade e Web semântica para a informação arquivística.

Por fim, fez-se uma busca no Google via Web utilizando as palavras “Ontologia” e “Descrição Arquivística”, no qual foi possível selecionar, entre as respostas de busca, o projeto ReLoad (2014), que evidencia que a proposta elaborada nesta pesquisa já é uma realidade. O projeto disponibiliza em seu endereço eletrônico uma Ontologia ISAD-G. O objetivo do projeto é testar os métodos de Web e padrão de tecnologias semânticas para dados abertos ligados para facilitar o compartilhamento de informações de arquivo a partir de uma variedade de fontes e veremos mais adiante quando abordaremos ontologias. A seguir conheceremos os padrões da área para dados digitais, o EAD, EAC e o MADRAS.

3.2. EAD para Descrição de Conteúdos e de Informação arquivística

Discute-se neste tópico se a utilização de padrão arquivístico de descrição como ISAD-G (estruturado em EAD e EAC) aumenta a visibilidade da informação arquivística e o uso de elementos EAD para a construção de ontologias.

O termo EAD é acrônimo de *Encoded Archival Description*, segundo a *The Library of Congress* (LOC, 2016a), EAD significa Descrição Arquivística Codificada em tradução livre, e é um padrão não-proprietário de metadados para a codificação de instrumentos de pesquisa para uso em plataformas digitais em rede online. Por sua vez, Yakel e Kim (2005) descrevem que “Encoded

Archival Description (EAD) is an SGML/XML (Standard Generalized Markup/Language Extensible Markup Language), Document Type Definition (DTD) for Archival Finding Aids”. Um instrumento de pesquisa é um recurso que contém informações detalhadas e padronizadas sobre um conjunto específico de documentos ou registros de um arquivo, resultante da descrição arquivística.

Trata-se, de fato, de informações padronizadas criadas por custodiadores de arquivos para fornecer mais informações sobre seu acervo e coleções específicas. A EAD possui, portanto, informação arquivística que é compartilhada. Não se trata de um sistema utilizado para gestão, mas seu foco de fato é a interoperabilidade da informação de arquivos. Quanto às informações dos próprios custodiadores, existe o modelo Encoded Archival Guide (EAG), que normaliza as descrições das próprias instituições de arquivo e foi desenvolvido pelos Archivos Estatales de Espanha.

Enquanto os instrumentos de descrição podem variar um pouco em estilo, o seu objetivo comum é fornecer descrição detalhada da organização do conteúdo e intelectual de coleções de materiais de arquivo. O uso do EAD permite a padronização das informações de coleções e acervos em instrumentos de pesquisa em rede, dentro e entre repositórios (LOC, 2016a).

Em relação à tecnologia, o *Standard Generalized Markup Language* (SGML) foi escolhido em detrimento de outras soluções possíveis devido a certas características que possui pois é um conjunto de regras para a definição e expressa a estrutura lógica de documentos, o que permite produtos de software para controlar a busca, recuperação e exibição estruturada de documentos. O SGML permite que as estruturas lógicas de documentos eletrônicos sejam

representadas de forma explícita e com rigor, de forma inequívoca e independente, que é reconhecido internacionalmente, independentemente de aplicações e sistemas (LOC, 2003).

As regras são aplicadas na forma de linguagem de marcação (pelo uso de *tags* ou rotulagem) que pode ser incorporado em um documento eletrônico para identificar e estabelecer relações estruturais entre as partes, expressas por metadados. A marcação consistente de documentos de estrutura semelhante é fundamental para o processamento eletrônico e o SGML incentiva a coerência com a introdução do conceito de Definição do Tipo de Documento (ou DTD de *Document Type Definition*).

A DTD prescreve o conjunto ordenado de *tags* de marcação SGML disponíveis para codificar as partes de documentos em uma classe similar, facilitando a interoperabilidade. Instrumentos de pesquisa de arquivo, que partilham peças e estrutura semelhantes, formam uma classe de documentos para que um DTD possa ser desenvolvido (YAKEL; KIM, 2005).

A adoção de DTD, para Yakel e Kim (2005), inclui dois processos distintos, mas relacionados: codificação e publicação. A codificação é o processo de marcação dos elementos dos instrumentos de pesquisa usando o DTD EAD. É uma tarefa que exige reformular ou extrair informações de um instrumento de pesquisa para coincidir com elementos da DTD EAD. Já a Publicação envolve a montagem de instrumentos de pesquisa marcados na *World Wide Web*. Para as autoras, o processo de publicação exige um maior nível de competências tecnológicas do que a codificação, tais como a capacidade de definir os arquivos em um servidor, abrir arquivos para os motores de busca, desenvolver folhas de estilo para instrumentos de pesquisa, e escrever scripts para transformar o XML

nativo em HTML (*Hypertext Markup Language*) para visualização em navegadores.

A utilização de padrão arquivístico de descrição como ISAD-G (estruturado em EAD e EAC) aumenta a visibilidade da informação arquivística pois pode ser empregado como um padrão de definição de elementos a serem processados pela Web Semântica. Portanto, condicionam a formação de ontologias arquivísticas. A EAD possui 146 elementos ao total, que descrevem um acervo arquivístico ou coleção e também servem para codificar inventários multinível (como a maioria das descrições arquivísticas que são multiníveis). Muitos dos elementos da EAD possuem equivalentes em outros formatos normalizados de descrição (como a ISAD-G ou a NOBRADE) e podem tomar formatos estruturais (como o MARC - *Machine Readable Cataloging* ou o *Dublin Core*), o que aumenta sua flexibilidade de uso e interoperabilidade dos dados.

Com isso, pode-se inferir o uso de elementos EAD para possibilitar a construção de ontologias, pois trazem a estrutura multinível função/atividade/documento, representando uma fatia de um domínio de conhecimento, como veremos no capítulo que trata de ontologias.

Os elementos EAD servem para descrever contexto, como por exemplo: <function> - Função, <accessrestrict> - Conditions Governing Access, <appraisal> - Informação de avaliação, <archdesc> - Descrição Arquivística, <archdescgrp> - Archival Description Group, <archref> - Archival Reference. Elementos do EAD podem descrever e representar o próprio Documento, por exemplo: <author> - Author, <abstract> - Abstract, <creation> - Creation, <custodhist> - Custodial History, <indexentry> - Index Entry, <item> - Item, <label> - Label. Alguns metadados do EAD são estruturais, por exemplo: <dao> -

Digital Archival Object, <daodesc> - Digital Archival Object Description, <daogrp> - Digital Archival Object Group, <daoloc> - Digital Archival Object Location, <langmaterial> - Language of the Material.

A Library of the Congress também elaborou um estudo de *crosswalk*¹, onde compara elementos do EAD com elementos da ISAD-G:

Figura 3 - Encoded Archival Description Tag Library, Version 2002.

A.1. ISAD(G) to EAD	
ISAD(G)	EAD
3.1.1 Reference code(s)	<eadid> with COUNTRYCODE and MAINAGENCYCODE attributes<unitid> with COUNTRYCODE and REPOSITORYCODE attributes
3.1.2 Title	<unittitle>
3.1.3 Dates	<unitdate>
3.1.4 Level of description	<archdesc> and <c> LEVEL attribute
3.1.5 Extent and medium of the unit	<physdesc> and subelements <extent>, <dimensions>, <genreform>, <physfacet>
3.2.1 Name of creator	<origination>
3.2.2 Administrative/Biographical history	<bioghist>
3.2.3 Archival history	<custodhist>
3.2.4 Immediate source of acquisition	<acqinfo>
3.3.1 Scope and content	<scopecontent>
3.3.2 Appraisal, destruction and scheduling	<appraisal>
3.3.3 Accruals	<accruals>
3.3.4 System of arrangement	<arrangement>
3.4.1 Conditions governing access	<accessrestrict>
3.4.2 Conditions governing reproduction	<userrestrict>
3.4.3 Language/scripts of material	<langmaterial>
3.4.4 Physical characteristics and technical requirements	<phystech>
3.4.5 Finding aids	<otherfindaid>
3.5.1 Existence and location of originals	<originalsloc>
3.5.2 Existence and location of copies	<altformavail>
3.5.3 Related units of description	<relatedmaterial> <separatedmaterial>
3.5.4 Publication note	<bibliography>
3.6.1 Note	<odd> <note>
3.7.1 Archivist's note	<processinfo>
3.7.2 Rules or conventions	<descrules>
3.7.3 Date(s) of descriptions	<processinfo> <p> <date>

Fonte: Library of Congress (2002). Appendix A: EAD Crosswalks.

¹ Crosswalk é um estudo que mostra elementos equivalentes (ou "campos") em mais de um esquema de banco de dados. Ele mapeia os elementos em um esquema criando um DE-PARA para os elementos equivalentes em outro esquema.

3.3. EAC para Contextos de Criação de Arquivos

Neste subcapítulo é abordado o padrão EAC, sobretudo para representar o contexto de criação de arquivos tanto para a Web Semântica como em sistemas informatizados. O EAC, por tratar do contexto, também compõe uma ontologia, para seus elementos representarem este contexto. Do ponto de vista da preservação, ele representa o *Archival bond* das informações arquivísticas que descreve e valida a ideia de ontologias para preservação.

O acrônimo EAC-CPF significa *Encoded Archival Context - (Corporate bodies, Persons and Families)* ou, em tradução livre - Codificação do Contexto Arquivístico – para entidades Coletivas, Pessoas e Famílias (EAC-CPF) - é um padrão XML utilizado para marcar informações sobre os criadores de documentos de arquivo e acervos gerados por uma entidade coletiva, pessoa ou família. O objetivo é fornecer informação contextual sobre as circunstâncias de criação dos documentos e coleções e sobre seu uso. EAC-CPF pode ser usado em conjunto com EAD e outras normas como a ISAD. O modelo é mantido *Society of American Archivists* (SAA, 2016).

A EAC-CPF aborda principalmente a descrição dos indivíduos, das famílias e órgãos sociais que criam, preservam, usam e são responsáveis e / ou associados a documentos arquivísticos de diversas maneiras. A sua principal finalidade é padronizar a codificação de descrições sobre agentes que permitam o compartilhamento, descoberta e entrega de informações em plataformas digitais. O EAC-CPF é uma estrutura de comunicação de informação contextual de arquivo para os indivíduos, entidades coletivas e famílias. Ele suporta a troca de registros de autoridade ISAAR-CPF - Norma Internacional de Registo de Autoridade Arquivística para Pessoas Coletivas, Pessoas Singulares e Famílias

(SAA, 2016). Assim, os elementos do EAC-CPF refletem a ISAAR-CPF e a ISAD-G, dois padrões gerenciados pelo Conselho Internacional de Arquivos.

A interoperabilidade semântica é uma condição para os arquivistas da Web Semântica. Já existe o uso de links como meio de conexão entre descrições arquivísticas na Web para acrescentar outras informações e aumentar a informação disponível para os usuários que acessam material de arquivo na Web. Segundo Evans e Rouché (2004) o XML fornece o mecanismo para codificar estrutura em um formato universal e também permite o desenvolvimento do registro como uma aplicação Web.

Para Mazzini e Ricci (2011), as tecnologias que melhor servem à descrição de um fundo de arquivo para Web Semântica são RDF e ontologias. Elas são entusiastas da ideia de transformar o esquema EAC-CPF em uma ontologia e da experiência de "abrir" os registros de autoridade EAC-CPF como dados abertos ligados. A opção de usar padrões permite processar, integrar e lidar com dados de acordo com as regras padronizadas que são suportados por grandes comunidades; a oportunidade de integrar com outros recursos da Web descritas com outros vocabulários-padrão. Porém agora podemos complementar esta visão com os padrões EAG e EAD já descritos.

Uma solução concreta é o uso de RDF e ontologias, não só como meio para representar entidades e as relações entre os diferentes elementos da descrição arquivística, mas também como uma ferramenta adequada para qualificar essas relações semanticamente. Para Mazzini e Ricci (2011) são necessárias algumas ações simples a serem feitas, a fim de descrever contexto de arquivo de uma forma "semântica": é necessário identificar os recursos

descritivos por meio de URI (link semântico), fornecer descrições em um formato padrão para que os recursos e as suas relações podem ser reconhecidos imediatamente (em EAD, por exemplo) e incluir nas descrições o maior número possível de links relevantes para outros recursos de informação.

Uma abordagem proativa do profissional da informação para o acesso, ao descrever baseado em padrões da comunidade de arquivística, revela à rede a existência dos documentos arquivísticos. Tanto os padrões de Conteúdo como ISAD-G, ISAAR-CPF ou as estruturas de comunicação: EAD, EAC-CPF, além de outros sistemas de organização do conhecimento como dicionários e taxonomias. Recursos assim são capazes de informar sobre o contexto e, portanto, destacar conexões implícitas e links "acionáveis" para outros recursos (ANGJELI, 2012, p. 5).

A EAC-CPF é capaz de trazer as informações contextuais essenciais para a compreensão dos documentos, ou seja, representa o *Archival bond* ou o vínculo de arquivo. Enquanto ISAAR-CPF concentra-se principalmente na noção de proveniência o EAC-CPF aplica-se a todas as "pessoas" documentados nos registros. (ANGJELI, 2012, p. 7). Do ponto de vista da preservação, por conseguinte, o planejamento estrutural de metadados e de descrição podem representar a informação arquivística e o *Archival bond* dessas informações, o que estimula a ideia de ontologias como ferramenta de preservação dos documentos de arquivo e de seu contexto de criação.

3.4. MADRAS: a Ligação entre Metadados e Semântica

Neste ponto é destacado o processo internacional de padronização de metadados visando a estruturação semântica da informação por meio do *Metadata and Archival Description Registry and Analytical System* (MADRAS).

O *International Research on Permanent Authentic Records in Electronic Systems* (InterPARES) - projeto com a participação de Luciana Duranti, criou um repositório de esquemas de metadados destinado a ajudar na identificação de conjuntos de metadados, ou das combinações de elementos de diferentes conjuntos, que servem para atender a várias necessidades de manutenção e preservação em longo prazo dos documentos. O MADRAS surge neste cenário, trazendo recomendações sobre como cada esquema pode ser expandido ou revisado para atender às necessidades de confiabilidade, autenticidade e preservação dos documentos digitais produzidos dentro do domínio, comunidade ou setor do usuário (INTERPARES, 2007).

O MADRAS, segundo Innarelli (2015, p. 209), é o padrão Metadados, Descrição Arquivística de Documentos e Sistema de Análise que surge para dar suporte ao registro de metadados, além de apoiar a análise dos registros de metadados, servindo também para padronizar a estrutura de metadados e produzir dados analíticos para o grupo que desenvolve o padrão ISO 23081. Ainda, conforme ressalta o autor “o MADRAS foi desenvolvido utilizando a tecnologia XML-DTD para estruturação dos metadados. Esta tecnologia foi escolhida por ser de simples compreensão e implementação nos sistemas informatizados, além de ser uma tecnologia de padrão aberto”. Como não aprofundou em sua tese o estudo sobre o MADRAS, o autor afirma que assume, em relação aos metadados de que “fazem parte do processo de preservação dos documentos arquivísticos digitais e que os mesmos devem ser

preservados juntamente com os documentos arquivísticos digitais” (INNARELLI, 2015, p. 209).

Durante o InterPARES 2, quando se definiu a criação do MADRAS, os estudos estavam indo neste sentido. Durante essa época, Evans e Rouche (2004, p. 316) afirmavam que os pesquisadores aspiravam analisar a papel que esses esquemas de metadados desempenhavam “na criação de registros, controle, manutenção, avaliação, preservação e uso”. Ressaltam também que “a ideia era verificar se eles podem suportar os requisitos de metadados para assegurar a confiabilidade, precisão e autenticidade dos registros ao longo do tempo”. Logo, os esquemas de metadados fornecem definições semânticas e estruturais de metadados, incluindo os nomes dos elementos de metadados, sua estrutura e seu significado. Normas para a descrição arquivística e especificações para sistemas de controle de arquivo são exemplos de esquemas de metadados no domínio de arquivo. Seu propósito foi o de identificar e definir os metadados necessários para o controle de arquivo de registros (EVANS; ROUCHE, 2004, p. 316).

3.5. PREMIS: Encapsulando as Ontologias

Neste subcapítulo é abordado o padrão *Metadata Encoding and Transmission Schema* (METS) como elemento de ligação entre a descrição arquivística e a ontologia, além de apreciar o modelo de preservação *Preservation Metadata* (PREMIS).

A *Library of Congress* desenvolve padrões estadunidense de descrição e estruturação de metadados. Entre eles, existem o *Technical Metadata for Optical Character Recognition* – o *Analyzed Layout and Text Object* (ALTO) - que são metadados técnicos de reconhecimento óptico de caracteres, o AudioMD e VideoMD – que

são esquemas XML que detalham metadados técnicos para gravação áudio e vídeo baseados em objetos digitais (LOC, 2016b; 2016d). O METS trata da estrutura de metadados para codificar descrições, dados administrativos e estruturais. O *NISO Metadata for Images in XML Schema* (MIX) é um esquema XML para codificação de elementos de dados técnicos necessários para gerenciar coleções de imagens digitais. Por sua vez, o PREMIS é um dicionário de dados e de apoio a esquemas XML para metadados de preservação em longo prazo de materiais digitais (LOC, 2016c; 2015b).

Além destes padrões, a LOC (2013) também mantém os padrões descritivos EAD e o MADS (*Metadata Authority Description Standard*) que equivale ao EAC que é mantido pelo Conselho Internacional de Arquivos. Para este trabalho, que avalia os padrões descritivos como possíveis ontologias, interessa dois destes padrões PREMIS e METS (Preservação e Descrição para Transmissão), pois são convergentes aos estudos da pesquisa, isto é, para as ontologias originadas da descrição arquivística e que possam também ser utilizadas na preservação.

O dicionário de dados PREMIS é o padrão internacional de metadados para apoiar a preservação de objetos digitais e garantir a sua usabilidade em longo prazo. Desenvolvido por uma equipe internacional de especialistas, o PREMIS é implementado em projetos de preservação digital em todo o mundo e incorporada a uma série de ferramentas e sistemas de preservação digital *open-source* e comerciais. O Comitê Editorial PREMIS coordena revisões e implementação do padrão, que consiste no Dicionário de Dados, um esquema XML, e documentação de apoio (LOC, 2016c).

O PREMIS foi registrado como um esquema de metadados reconhecido para ser usado como metadados

administrativos, pois operam na preservação. Os esquemas em PREMIS também foram concebidos para terem implementação neutra, e, podem ser aplicados em conjunto com o METS, por conta de sua flexibilidade. A LOC dispõe em seu site um estudo que aprofunda este uso conjunto para representar informação administrativa e para preservação em longo prazo.

Em 2013 o Comitê Editorial do PREMIS, estrutura da LOC, publicou o *PREMIS OWL Ontology*, com base no Dicionário de Dados para Metadados de Preservação. Trata-se de um padrão de preservação digital baseado no modelo de referência OAIS. Anteriormente, a Ontologia PREMIS OWL foi disponibilizada em outubro de 2011, aderindo às definições de unidade semântica do PREMIS (LOC, 2015).

3.6. METS: padrão arquivístico de metadados de objetos descritivos

Já o esquema METS é um padrão para a codificação de metadados sobre objetos descritivos, administrativos e estruturais dentro de um conjunto informacional, que pode ser expresso utilizando a linguagem de esquema XML. O padrão é mantido pela LOC, porém sua origem remonta a inúmeras bibliotecas universitárias e estaduais nos Estados Unidos, que uniram esforços entre 2004 e 2007 para determinar um padrão para identificar, arquivar e preservar os recursos digitais. Ao final do projeto, haviam produzido um sistema de arquivamento digital de código aberto, onde os pacotes de objetos digitais usavam o METS para que os objetos pudessem ser arquivados e acessados em diferentes sistemas (INTERPARES, 2007).

O padrão METS permite o uso de esquemas de metadados desenvolvidos externamente, ou seja, ele pode assumir elementos de diferentes perfis de metadados. O METS em si é agnóstico, isto é, não depende de uma única plataforma ou estrutura. Desse modo, são diferentes esquemas de metadados descritivos ou administrativas que os seus implementadores optam por utilizar (LOC, 2016e).

Com base nos aspectos analisados, verifica-se que de fato o padrão METS, que pode ser utilizado para expressar esquemas como EAD ou o PREMIS, é uma expressão de ontologias (utilizando OWL), mostrando que esquemas descritivos, acompanhado de informações de contexto e outros metadados, criam ontologias.

3.7. Ontologias: Estruturando Domínios de Conhecimento

Neste subcapítulo é abordada as ontologias e seu papel na estruturação dos domínios de conhecimento e seu uso na tecnologia digital. Mas, antes de arrolar mais sobre ontologias para a área da Ciência da Informação e prospectar seu uso para a área de Arquivos, é importante saber sobre seus conceitos enquanto disciplina, que surgiu na Filosofia.

Em relação ao campo filosófico da Ontologia, Almeida (2014, p. 244) descreve que “existe consenso de que o estudo da Ontologia diz respeito aos tipos de coisas que existem. Nesse contexto, “tipo” quer dizer “categoria”, um termo que foi usado ainda por Aristóteles para discutir que declarações sobre uma entidade”.

A categorização, algo que é comum à Documentação e Ciência da Informação, já nos seus primórdios é objetivado por ontologistas, pois:

Uma teoria das categorias é o mais importante tópico do estudo da Ontologia. Tais teorias especificam sistemas de categorias estruturados em níveis hierárquicos, em geral, na forma de uma árvore invertida na qual a categoria de mais alto nível é nomeada “entidade. (ALMEIDA, 2014, p. 244).

Com base em Almeida (2014), foi feito um quadro, a seguir, que traz os sistemas ontológicos indicados pela Filosofia, que poderemos ver a seguir:

Quadro 12 - Os Sistemas Ontológicos da Filosofia.

F I L O S O F I A	Sistema Realista baseado em Aristóteles	Kategoria é um termo para Predicação Entidade é a Essência real Distinção de essências -> Gênero/Espécie e dicotômica Tradição Aristotélica: ordenar categorias de acordo com características essenciais
	Sistema Cognitivo baseado em Kant	Julgamento e proposições definem categorias Quatro aspectos para classificar os julgamentos: quantidade, qualidade, relação e modalidade Categoria do Entendimento humano
	Sistema Descritivo baseado em Husserl	Ontologia é estudo dos fenômenos -> fenomenologia Redução fenomenológica: posição transcendental para descrever consciência Reino dos fatos: entidades concretas (indivíduos, estados e eventos empíricos) Reino das essências: entidades ideais, não-reais,

		essências formais e essências materiais, naturais Reino dos significados: conteúdo de experiências intencionais, Significados não se localizam no tempo e espaço, eles são ideais, mas não são essências
--	--	--

Fonte: Adaptação de Almeida (2014).

O conceito de ontologia também é tratado na área de Ciência da Computação, sendo comum à área tecnológica desde o seu desenvolvimento no pós-guerra, e é possível saber que:

Em Representação do Conhecimento (RC), um sub-campo da Inteligência Artificial (IA), o termo ontologia tem sido usado desde os anos de 1960 para se referir a uma estrutura de conceitos representados por um vocabulário lógico. Nos anos de 1990, o termo manteve sua presença no contexto do conjunto de tecnologias que se convencionou rotular de “Web Semântica”. (ALMEIDA, 2014, p. 248-249).

Com informações de Almeida (2014), foi possível sistematizar o quadro 7, que traz os usos de ontologias indicados pela Ciência da Computação.

Quadro 13 - Os usos de Ontologias na Ciência da Computação.

C O M P U T A Ç Ã O	Princípios Ontológicos: Entender e modelar a realidade	● Papel original na filosofia, lógica ● Fornecer uma descrição do que existe ● Caracterizar entidades
	Representação de um domínio em uma linguagem de representação computacional	● Conjunto de declarações expressas em linguagem de representação ● Processamento por mecanismo de inferência (busca)

Fonte: Adaptação a partir de Almeida (2014).

Apesar deste breve histórico, ainda não há muitas definições acerca do uso. Em relação à rede, as ontologias também são artefatos fundamentais para a Web Semântica, como destacam Berners-Lee, Hendler e Lassila (2001), pois utiliza-se de ontologias e metadados com o intuito de estruturar e dar significado ao conteúdo das páginas Web.

Sabe-se que uma ontologia é a “soma de uma série de conceitos relevantes que representam o conhecimento compartilhado em um domínio”, como descreve Moreira Gonzáles (2011a, p. 77). Para o autor, é descrita como uma especificação explícita, onde a conceituação é testada por meio das múltiplas relações entre os conceitos e validados pelos axiomas (assertivas sobre o domínio) destes conceitos, como veremos em Componentes do Domínio de uma Ontologia. Trata-se, portanto, de uma

conceituação formal de um domínio específico ou parcela da realidade. O domínio pode ser uma área temática ou uma área de conhecimento. O autor, ainda, profunde o conceito de Ontologias, do qual elaborou-se o quadro a seguir:

Quadro 14 – Conceitos de Ontologia de Domínio para a Ciência da Informação.

Ontologia	Codifica o conhecimento de um domínio
	Descrição explícita e formal de conceitos de um domínio de discurso
	Base de conhecimento

Fonte: Adaptação de Moreiro Gonzáles (2011a, p. 77).

Existem outros tipos de ontologias além das de domínio, porém em relação à esta, a especificação dela trata-se de uma ferramenta e é composta de conceitos estruturados, em classes e suas relações entre propriedades e indivíduos, conforme destacado a seguir:

Quadro 15 – Componentes de uma especificação de uma ontologia.

Classes	Aspectos gerais em muitos domínios de interesse
Propriedades	Atributos que as coisas podem ter e restrições

Indivíduos	Instâncias ou exemplares
-------------------	--------------------------

Fonte: Adaptação de Moreiro Gonzáles (2011a, p. 77).

Por isso, muitas vezes há comparação com as linguagens documentárias, tal como fizeram Pedraza-Jiménez, Codina e Rovira (2007, p. 574) ao afirmarem que “Las ontologías pueden considerarse lenguajes documentales con distintos niveles de estructura, pero a diferencia del tesauro tradicional están elaboradas con una sintaxis comprensible para los ordenadores”. As ontologias serviriam, portanto, para permitir acesso em longo prazo a documentos, inclusive os arquivísticos, e tanto seria possível que para a área isso já seria praxe, pois:

Em Ciência da Informação, sistemas de categorias também têm sido criados, não com o objetivo de representar as coisas do mundo, mas sim o conteúdo de documentos visando recuperação da informação. Um exemplo de sistema de categorias bem conhecido é o trabalho de Ranganathan (1967). (ALMEIDA, 2014, p. 252).

Neste sentido, a ontologia atende aos objetivos das duas dimensões da essência informacional, condizentes com a dimensão descritiva, voltada aos elementos relativos à forma dos documentos e a dimensão temática, voltada aos conteúdos informacionais. Uma ontologia, portanto, é uma estrutura que permite organizar a informação e auxiliar na sua recuperação pois define as regras que regulam a combinação entre os termos e suas relações. Utiliza um

vocabulário de termos que descrevem uma determinada realidade. Esses termos são categorizados para representar um determinado domínio. Dessa forma, uma ontologia pode ser concebida como uma linguagem, pois estabelece um conjunto de termos que, posteriormente, poderão ser usados para formular consultas em sistemas e na rede (ALMEIDA; BAX, 2003, p. 9).

Quanto à sua estrutura, os componentes do domínio de uma ontologia são os conceitos, suas relações, as funções determinadas, as instâncias e os axiomas como podemos ver no quadro dos componentes de um domínio de uma ontologia:

Quadro 16 – Componentes de um Domínio de uma ontologia.

Categorias	Ideias básicas que se tenta formalizar (objetos, métodos, estratégias, processos de raciocínio)
Relações	Interação e enlace entre os conceitos do domínio (taxonomias de domínio)
Funções	Identifica elemento por função, por meio de uma relação concreta.
Instâncias	Objeto determinado de um conceito
Axiomas	Assertiva sobre relações que elementos devem cumprir

Entretanto, se o sistema de categorias pode ser representado por um tesouro, o que a ontologia tem em relação a

estrutura e disponibilização da informação? As ontologias são mais amplas ao comportarem outros tipos de relações entre os conceitos, conforme pode-se verificar:

Como vimos, as ontologias contemplam uma ampla gama de relações de classes e subclasses (como em uma taxonomia) ou as de sinonímia e meronímia (como em um tesouro) já que em princípio essas relações não estão fechadas, mas que em parte dependem das relações reais que se dão entre as classes os indivíduos do domínio modelados pela ontologia. Portanto, uma ontologia permite maior riqueza na definição dos seus conceitos e relacionamentos que um tesouro(PEDRAZZA-JIMÉNEZ; CODINA; ROVIRA, 2007, p. 574 tradução nossa)

O que há em comum entre estes sistemas de organização do conhecimento e as ontologias é de fato o uso de linguagem. Conforme Gilchrist (2003) o que existe claramente comum entre tesouros, taxonomias e ontologias é que todos eles lidam com a linguagem natural. Se existe um número de objetos, segundo o autor, é possível organizá-los em grupos e aplicar rótulos a estes grupos – e os bibliotecários têm tradicionalmente feito isso na classificação de livros para arranjo nas prateleiras.

O processo de agrupamento (categorização) e de classificação (rotulagem) torna comum o uso de ontologias e sistemas de organização do conhecimento. Considera-se, portanto, a ontologia gerada da aplicação do padrão descritivo como uma conceitualização formal de um domínio ou de uma parcela de realidade, com a qual podem operar diferentes aplicações software: Os conceitos ou termos utilizados para a descrição servem como vocabulário comum (sintático e semântico) que favorece a

comunicação e a interoperabilidade de recursos. Dão sentido pleno à informação ao situá-la dentro de um contexto (MOREIRO GONZÁLES, 2011b, p. 151).

Cabe destacar que, de acordo com Gilchrist (2003), em relação a informação arquivística, numa analogia, quando cita a “informação do contexto” tal informação é evidenciada por meio da descrição arquivística. Ainda, segundo o autor, a classificação é tornada possível pela análise das relações entre conceitos e suas associações com objetos e redes semânticas. Contudo os relacionamentos podem depender do contexto e, por isso, torna-se necessário adicionar mais informações sobre o conceito na forma de definições e notas sobre o uso.

Com base nesses aspectos, a classificação é tornada possível por intermédio da análise das relações entre os conceitos e suas associações com objetos; e os relacionamentos dependem do contexto e, assim, torna-se necessário adicionar mais informações sobre o conceito na forma de definições e notas sobre o uso. Verificou-se, no capítulo que discorre sobre descrição arquivística, que estes são elementos informacionais descritivos presentes na ISAD-G e como veremos também na EAD e EAC. Lembrando que o nível 0 de descrição, a instituição custodiadora também pode ser descrita por meio do EAG (Encoded Archival Guide).

3.8. Ontologias e Padrões Arquivísticos

Neste subcapítulo, é verificado que as ontologias arquivísticas podem ser consideradas ontologias de tarefas, e aplica-se a abordagem arquivística para comparar a estrutura de especificação de ontologias com a estrutura de classificação funcional da arquivística.

Quanto aos padrões arquivísticos, ao se referir ao conhecimento explícito digital (ou seja, entre eles o documento arquivístico digital), sabe-se das ferramentas que permitem a organização, a gestão e o acesso da documentação arquivística e seu conteúdo explícito, sendo eles: o Plano de Classificação, as Tabelas de Temporalidade, o Tesauro (extensível às ontologias, observação deste autor), isto considerando um armazenamento seguro da documentação (ou seja, contexto confiável, n.a.), com acesso envolvendo direitos, regras de ferramentas de busca, recuperação e uso da informação, além dos instrumentos de pesquisa (INNARELLI, 2012).

Para Lopez (2002) os padrões resultam da aplicação de normalização que contribui não apenas para o intercâmbio entre diferentes instituições, como também facilita o acesso e a consulta em geral. Ou seja, a aplicação de padrões normalizados de estruturação da informação prepara a informação arquivística para a interoperabilidade.

Os instrumentos de pesquisa representam os meios de identificação, localização ou consulta a documentos ou a informações neles contidas: os mais comuns são chamados de catálogo, guia, índices e inventário. Estes são resultado da Descrição Arquivística e retratam a informação arquivística como já vimos. Quanto ao Tesauro, afirma que este permitirá controlar o vocabulário de termos derivados da linguagem natural, normalizados e preferenciais, agrupando as palavras por afinidade semântica, com indicação de equivalência, hierárquica, partitivas, de negação e funcionais estabelecidas entre eles (ARQUIVO NACIONAL apud INNARELLI, 2012).

Compreende-se que as ontologias servem aos agentes computacionais como as linguagens documentárias e os vocabulários controlados. Tratam-se de instrumentos para representar e descrever conteúdos de objetos informacionais tanto para homens quanto para inteligências artificiais. Conforme citam Sayão e Marcondes (2008, p. 141), as "ontologias estabelecem uma compreensão compartilhada de um domínio de interesse para apoiar a comunicação entre seres humanos e agentes computacionais". É a mesma função declarada para outras taxonomias e até para tesauros e operação base do processamento informacional.

Há neste contexto um processo de representação e de estruturação de meta-informação, pois, de acordo com Sayão e Marcondes (2008, p. 141), "as ontologias são simbolizadas caracteristicamente por uma linguagem de representação processada por computador". Estas linguagens de representação estão estruturadas e padronizadas por diversas iniciativas internacionais e públicas, criando variados modelos e esquemas de metadados. Dessa forma, sendo "considerada uma tecnologia-chave para o desenvolvimento da Web Semântica", exatamente por corresponder a uma das camadas básicas da estrutura disponível da Web Semântica. Estas camadas, segundo a W3C, a *Web Ontology Language* (OWL), principalmente, e as camadas RDF, utilizando-se da semântica para generalização e da semântica para simplificação (é na RDF que também são estruturadas as diversas espécies de taxonomias e representações de domínios da Web).

Essa constatação serve para determinar o procedimento de processamento informacional, que ocorre por meio de metadados temáticos e descritivos e estabelecer a interoperabilidade, pois, a partir daí é onde se dá sentido à informação, visto que:

[...] as ontologias têm contribuído para facilitar a comunicação e o processamento de informação semântica tanto entre humanos quanto entre sistemas computacionais. Sendo comumente utilizadas para promover a interoperabilidade entre sistemas ao representarem os dados compartilhados por diversas aplicações. (BORGES; MORALES, 2013, p. 23).

Em relação ao uso de ontologias e sua aplicação em plataforma digital e sistemas informatizados de arquivos, pode-se identificar que os padrões descritivos geram metadados. Estes metadados são diferentemente estruturados, conforme camada da ontologia. Estas camadas são a forma como se especificam as descrições (e, portanto, se estruturam as ontologias)

Quadro 17 – Estrutura de especificação de uma ontologia.

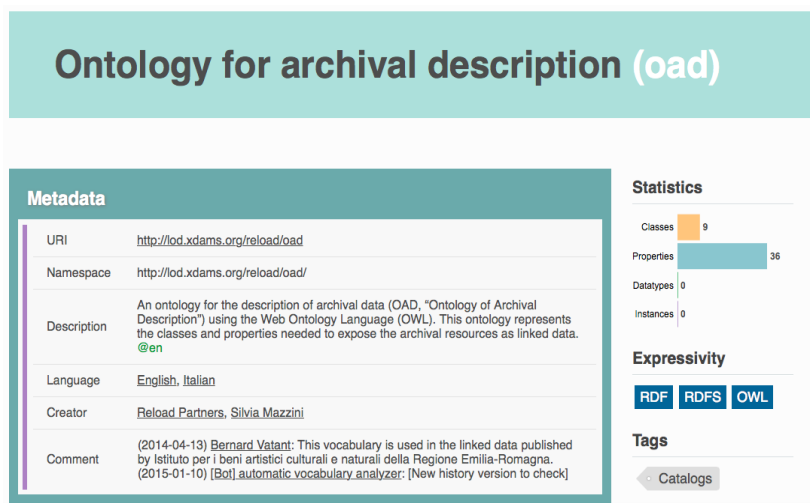
Classes	Aspectos gerais em domínios de interesse
Relações	As trocas entre os recursos
Propriedades	Atributos que os recursos podem ter
Indivíduos	Instância ou exemplares

Fonte: Adaptação de Moreiro Gonzales (2011a, p. 77).

Ao aplicar esta abordagem arquivística, pode-se considerar a ideia de representação da informação de arquivos poder ser estruturada como um domínio de conhecimento (ou um domínio funcional), mantendo seu contexto (representado pelo *Archival bond*) que garante a organicidade, sendo o retrato da própria estrutura de um arquivo, espelhado pelo contexto e pela informação arquivística. Isso tudo é codificável e exibido por meio de metadados (EAD e/ou EAC, por exemplo) e mantido ao longo do tempo por meio de ontologias que registram suas modificações e evoluções.

Ao tratar de acervos descritos pela ISAD-G a estrutura desses metadados (EAD e EAC) e a relação de seus elementos nominativos já pode ser considerada uma Ontologia de Descrição Arquivística (*Ontology for Archival Description* - OAD), apresentada na figura a seguir, que é uma iniciativa internacional em padronizar a informação arquivística aplicando os elementos definidos pela ISAD-G, e que visa interoperabilizar as informações geradas por unidades arquivísticas, arquivos e instituições com arquivos.

Figura 4 - Ficha de inscrição da ontologia no catálogo online Linked Open Vocabularies (LOV).



Fonte: Linked Open Vocabularies (2015).

O projeto ReLoad, desenvolvido pelo *Archivio Centrale Dello Stato e Istituto per I Beni Culturali dell'Emilia Romagna Regesta*, promove a experiência com normas técnicas de arquivo e

Web Semântica, além de testar métodos relativos a dados abertos interligados, no qual dispõe que:

A definição de uma ontologia de descrição arquivística é necessária a fim de testar o potencial da Web de dados para descrições arquivísticas. A descrição arquivística visa a representação de uma unidade de descrição por coletar, analisar, organizar e registrar as informações necessárias para identificar, gerir, localizar e explicar o contexto e o material documental, além dos sistemas de armazenamento que os produziram (ISAD-G). (RELOAD, 2014, tradução nossa).

O projeto ReLoad também divulga a ontologia de descrição arquivística OAD, que tem como objetivo a representação formal de descrições de níveis informacionais - entendido como objetos da própria descrição de arquivo:

Em particular, a OAD renuncia em levar em conta todos os elementos individuais das descrições, priorizam os elementos de informação considerados necessários para a exposição na rede das unidades de dados de descrição arquivística garantindo a integração com outros conjuntos de dados publicados também em formato de Linked Open Data. (RELOAD, 2014, tradução nossa).

A OAD, em versão beta (ou seja, ainda em revisão e recebendo contribuições), é expressa em OWL, a camada de semântica mais elaborada, ideal para ontologias pesadas

Leva em conta todos os elementos de descrição arquivística presentes na norma ISAD (G) e integra com outros elementos de informação não abrangidos pela norma descritiva - como as entradas de índice - e com

links para os criadores e depositantes. (RELOAD, 2014, tradução nossa).

A informação arquivística resultante da descrição arquivística estabelece um modelo de dados interoperáveis, padronizando seus elementos e é o que Moreiro Gonzáles (2011a, p. 80) considera uma ontologia de tarefa: aquela que descreve vocabulário de determinada atividade mediante especialização de termos definidos em alto nível. Tais elementos estruturam ontologias leves, principalmente para ser disponibilizada na camada específica para isso.

Os elementos podem ser descritos utilizando o OWL, ou como diz Hinz (2006, p. 26), tratar-se de uma ontologia de aplicação, que representa um domínio específico de uma área genérica de conhecimento. Por isso, acredita-se que as ontologias arquivísticas, resultado da descrição usando padrões internacionais são capazes de refletir a informação arquivística na Web 3.0 e, conseqüentemente, operar na preservação da informação digital arquivística em sistemas informatizados de arquivo.

3.9. Ontologias Digitais Arquivísticas

Aqui, aborda-se as características das ontologias que podem resultar do uso de padrões arquivísticos e como, na área da Ciência da Informação, isso pode representar um *Knowledge Organization System* (KOS). A discussão é apresentada a partir do ponto de vista da estruturação da informação.

Uma ontologia é uma conceituação formal de um domínio ou uma parcela da realidade, como afirma Moreiro Gonzales (2011a, p.76). Isso cria uma relação entre a ontologia e a arquivística, pois os arquivos são o resultado das ações das instituições e das

pessoas que nelas atuam. Trazem uma representação da realidade das instituições e pessoas e das estruturas que a geraram, expressadas por meio de suas relações orgânicas (o *Archival bond*), podendo ser considerada uma parcela da realidade. Uma ontologia, portanto, pode manifestar a informação arquivística, o que a torna uma ferramenta de representação da informação orgânica e da estrutura que a geraram e têm sua custódia. Assim, ao utilizar as estruturas propostas pelos padrões arquivísticos o EAD e a EAC, estes instrumentos de pesquisa digitais disponíveis na Web (guias, catálogos, índices, etc.) e sua totalidade de metadados poderia ser considerado uma Ontologias de Aplicação.

Uma Ontologia de Aplicação descreve os conceitos de um domínio com termos definidos dentro do domínio. Moreiro Gonzáles separa os tipos de ontologias por granularidade ou especialização:

Quadro 18 – Tipos de ontologia por granularidade ou especialização.

Alto Nível	Descrevem conceitos muito gerais como espaço, tempo, materiais, objetos. É independente de domínio.
Domínio ou tarefa	Descrevem vocabulário relacionado de um domínio genérico ou determinada atividade mediante especialização de termos definidos no alto nível
Aplicação	Descrevem conceitos que dependem tanto de ontologia de domínio ou de tarefa mediante especialização de termos definidos nele mesmo

Fonte: Adaptação de Moreiro Gonzáles (2011a).

Outra visão resumida por Hinz (2006, p. 26) trata da divisão de Tipos de Ontologia, relacionando-as à sua função, ao grau de formalismo de seu vocabulário, à sua aplicação e à estrutura e conteúdo da conceitualização:

Quadro 19 – Tipos de Ontologias relacionando à sua função.

Ontologias de Representação	Definem as primitivas de representação - como <i>frames</i> , axiomas, atributos e outros – de forma declarativa.
Ontologias Gerais (ou de alto nível):	Contém definições abstratas necessárias para a compreensão de aspectos do mundo, como tempo, processos, papéis, espaço, seres, coisas, etc.
Ontologias centrais ou genéricas de domínio (<i>core ontologies</i>):	Descrevem ramos de estudo de uma área e seus conceitos mais genéricos e abstratos.
Ontologias de domínio e de aplicação	Tratam de um domínio específico de uma área genérica de conhecimento, como direito tributário, microbiologia, etc.

Fonte: Adaptação de Hinz (2006, p. 26).

Considera-se, portanto, a ontologia gerada da aplicação do padrão arquivístico descritivo uma ontologia digital arquivística, pois trata-se de uma parcela da realidade disponibilizada na rede e representando um arquivo, objetivando sua interoperabilidade, pois esta informação traz o contexto orgânico da geração informacional intrínseco ao objeto informacional, permitindo assim sua migração entre sistemas o que pode garantir sua preservação em longo prazo. Esse sentido possibilita a aplicação da abordagem arquivística proposta para este trabalho. Com este vínculo orgânico, vê-se que a ontologia carrega em sua estrutura também o esclarecimento acerca deste organismo que acumula a informação arquivística.

A informação da fonte originadora, isto é, o contexto de geração orgânica, faz parte de ontologias tipificadas como de Aplicação, pois tratam dos termos e dos temas tratados, trazendo também informação sobre o gerador da informação e sobre o custodiador do acervo. Esse tipo de informação pode ser considerado como informação arquivística pois também carrega os dados sobre a estrutura dos arquivos e seus níveis (do fundo às séries, das coleções às unidades documentais). Isto potencializa, portanto, o uso de ontologias de aplicação, estruturadas com base em padrão descritivo da área arquivística, para a preservação da informação de arquivos em longo prazo.

Analogicamente é o que Thomaz (2007) identifica como informação de representação, ao se referir ao modelo OAIS, qualificando como estrutura de dados computacionais os agrupamentos desses tipos de dados e suas regras de mapeamento. São tratados, portanto, como informação estrutural, além das demais necessárias para a abertura dos pacotes de informação do modelo, que também trazem as informações semânticas. De tal modo, a informação de representação, validado pelo modelo OAIS, é composta por informação estrutural e mais a informação semântica.

Para a área da Ciência da Informação isto pode representar mais uma frente de atuação. Hodge (2000) afirma que ontologia é o mais novo rótulo ligado a alguma KOS ou sistema de organização do conhecimento. Afinal, as ontologias estão sendo desenvolvidas como modelos de conceitos específicos em comunidades de Gestão do Conhecimento, viabilizados por meios de tecnologias da informação. Eles podem representar relações complexas entre objetos e parcelas da realidade e incluem as regras e axiomas que faltam às redes semânticas.

Para Zeng e Hodge (2011), as ontologias que descrevem o conhecimento de uma área específica são frequentemente relacionadas com sistemas de mineração de dados e gestão do conhecimento. O conceito de sistema de organização do conhecimento pretende abranger todos os tipos de esquemas para organizar informações e promover a gestão do conhecimento. A sigla KOS (Knowledge Organization System) tem sido usada tanto como um singular (para representar uma espécie de recurso, que é diferente de outros tipos de recursos) com no plural (quando se refere a exemplos KOS). “Different families of KOS, including thesauri, classification schemes, subject heading systems and taxonomies, are widely recognized and applied in both modern and traditional information systems. (ZENG; HODGE, 2011, p. 30).

Vários tipos de KOS têm sido cada vez mais adotados como serviços baseados na Web para facilitar a descoberta de recursos e a recuperação de informações que adotem princípios de interoperabilidade. Afinal, diferentes agentes, serviços e aplicações precisam se comunicar com dados na forma de KOS para transferência, troca, transformação, mediação, migração e integração.

Foi reconhecido que a informação sobre a KOS, incluindo o seu modelo de dados, tipo de protocolo, status, órgão responsável, formato disponível, e contexto de uso e geração, além de outros dados descritivos, fazem parte integrante de registros de terminologia, serviços de documentos e os utilizadores do vocabulário (se máquina ou humanos), além de sistemas de recuperação. (ZENG ; Hodge, 2011, p 30 tradução nossa).

Novamente, a interoperabilidade e a informação contextual fazem parte de KOS, e as ontologias também são

consideradas um tipo de KOS. Zeng e Hodge (2011, p. 33) propõem uma taxonomia para os instrumentos KOS, categorizando que havia três grupos básicos:

Quadro 20 – Taxonomia de instrumentos KOS.

Lista de termos	Registros de autoridade, glossários, dicionários
Categorização e Classificação	Cabeçalhos de assunto, sistemas de classificação, taxonomias e sistemas de categorias
Grupos de relacionamento	Tesouros, redes semânticas e ontologias

Fonte: Adaptação de Zeng e Hodge (2011, p. 33).

Com isso, é possível verificar que a maioria dos KOS servem para estruturar a informação, principalmente no tipo grupos de relacionamento que garantem sua representação temática e podem armazenar informações de seu contexto, além de garantir a interoperabilidade semântica.

Hinz (2006) demonstra que a necessidade de apresentar de forma estruturada e organizada um domínio e um vocabulário comum, usado em consenso por uma comunidade (princípios de uma ontologia), é de importância reconhecida pelos desenvolvedores de soluções informatizadas. Desse modo, são amplamente utilizadas em Inteligência Artificial, Engenharia do Conhecimento e Ciência da Computação, comércio eletrônico, processamento de linguagens naturais, recuperação da informação na Web, entre outros.

Além da rede Web, vê-se a uma das aplicações das ontologias também na área de Banco de Dados, que as utiliza para fazer a integração entre recursos e compatibilização de dados, provendo com padrões de formato e possibilitando a interoperabilidade de informações entre sistemas heterogêneos, porque a maioria de sistemas podem possuir diferentes modelos conceituais, mesmo quando tratam de um mesmo domínio (HINZ, 2006, p. 42). Aqui vê-se o uso das ontologias para a interoperabilidade.

Já, para Moreiro-Gonzáles (2011a, p. 75), as ontologias possuem camadas que contextualizam a informação em determinado cenário. Assim sendo, uma ontologia traz conceitos relevantes de conhecimento compartilhado por um domínio, as relações entre estes conceitos e os axiomas ali gerados. Logo, conclui-se que uma ontologia é a soma de uma série de conceitos relevantes, que representam o conhecimento compartilhado de em domínio de conhecimento.

O compartilhamento (ou a interoperabilidade) também é original às ontologias, e está destacada no conceito disposto por Smith (2008, p. 156) onde frisa que “definições aceitas e projetadas para produzir um quadro lexical ou taxonômico para a representação do conhecimento e que pode ser compartilhado por diferentes comunidades de sistemas de informação”. Nesse caso, a interoperabilidade traz em si a portabilidade e o compartilhamento como princípio.

E sobre os conceitos de arquivo específicos, quando aborda ferramentas de representação, Angjeli (2012, p. 21) afirma que estas são as ferramentas, os vocabulários e ontologias existentes são o suficiente para trazer corretamente os conjuntos de dados LOD

de arquivo da nuvem, principalmente para interconectar os recursos de forma adequada, e não distorcerem a história. Nas experiências com LOD, cada um dos projetos criou seu próprio modelo de dados, ou seja, sua ontologia.

Logo, uma Ontologia de Aplicação que represente um arquivo, estruturada com base em padrões descritivos da área arquivística, ou uma Ontologia Digital Arquivística, pode garantir a preservação da informação de arquivos em longo prazo, especialmente por conta da interoperabilidade que pressupõe a pré-disposição da informação para o intercâmbio e portabilidade. Consequentemente, é possível disponibilizar a informação arquivística a ser utilizada pela semântica computacional por meio dos instrumentos de pesquisa arquivísticos disponibilizados na Web, e para migração entre sistemas de gestão e repositórios confiáveis, garantindo a manutenção da cadeia de custódia.

3.10. Assim sendo: juntar os pontos

A realidade dos dados institucionalizados em rede e a maneira que a Tecnologia da Informação, e as formas de relação de acesso por meio de interface de interatividade impactam na organização dos objetos informacionais e refletem em suas relações com a arquivística. Tal fato será para a Ciência dos Arquivos representadas pelas funções arquivísticas, que engloba a descrição como uma função primordial ao permitir acesso a informação arquivística registrada.

O impacto do digital na arquivística parte do conhecimento da Diplomática Arquivística contemporânea e a tipologia documental como ferramenta de análise funcional. A partir disso, diferencia-se os documentos digitalizados e os digitais, e

vislumbrou-se de que forma pode ser tratada sua gestão e preservação. Ao falar de vínculo arquivístico (*Archival bond*), viu-se de que forma esta relação orgânica pode garantir a confiabilidade da informação, fornecendo para sistemas de arquivo e para a Web 3.0 os dados acerca do contexto de produção documental.

Os sistemas informatizados de arquivos, seus tipos conhecidos, indicados por normas e a forma que a informação é preservada e acessada por meio destes sistemas. Foram permeados os conceitos de SIGAD e e-ARQ, que são a realidade brasileira em relação a normatização e modelo tecnológico de sistema de informação de arquivo, e posteriormente, discutiu-se o papel de definidor de requisitos que o profissional da informação assume frente à tecnologia e a relação dessas interfaces de gestão com a preservação digital. Também foi analisada o Modelo *Record Continuum*, adotado pelos arquivos nacionais da Austrália e possui aporte teórico e metodológico importante.

A confiança de um repositório arquivístico digital, considerando seu contexto característico e requisitos que qualificam esse tipo de Sistema de Informação de Arquivo, assim como a relação da cadeia de custódia e do contexto arquivístico são essenciais para garantir, em longo prazo, a confiabilidade do ambiente de gênese documental. Pois considerando Thomassen (2006, p. 7) “arquivos confiáveis contêm evidência confiável a respeito de decisões tomadas, direitos adquiridos e compromissos assumidos”. A presunção de autenticidade deve ser resultado da confiança na cadeia de custódia.

O emprego do modelo OAIS como padrão para a preservação da informação arquivística. Afinal, a implantação de um repositório digital confiável é fundamental para assegurar a

preservação, o acesso e a autenticidade de longo prazo dos materiais digitais arquivísticos. Ao considerar os pressupostos tecnológicos para a interoperabilidade e sua relação com a organização da informação, relacionou também os conceitos de Tecnologia da Informação com os conceitos desenvolvidos na Ciência da Informação e na Ciência Arquivística.

Um modelo de padrão avaliado foi a Descrição Arquivística e seu potencial para a interoperabilidade e para evidenciar a informação de arquivos, além de representar uma das principais funções da Arquivologia, indica a forma de tratar o resultado da descrição como potencial Ontologia.

A utilização do padrão arquivístico de descrição como ISAD-G (estruturado em EAD e EAC) aumenta a visibilidade da informação arquivística e pode se utilizar elementos EAD para a construção de ontologias estruturadas em RDF e OWL. O EAC, por tratar do contexto, também compõem uma ontologia, uma vez que, para representar o contexto precisa explicitar seu *Archival bond*. Do ponto de vista da preservação, portanto, descreve e valida a ideia do uso de ontologias para preservação.

Ao examinar as ontologias e seu papel na estruturação dos domínios de conhecimento e seu uso na tecnologia digital, foi verificado que as ontologias arquivísticas podem ser consideradas ontologias de tarefas, e se comparou à estrutura de especificação de ontologias com a estrutura de classificação funcional da arquivística. As características das ontologias que podem resultar do uso de padrões arquivísticos e o resultado disso como uma KOS.

A Informação Arquivística gerada nos processos de descrição, utilizando a ISAD-G e NOBRADE, pode ser uma forma de

estruturar ontologias na Web, utilizando os elementos da EAD, EAC e EAG, criando uma ontologia digital arquivística. A partir desse ponto, pode-se aferir que a formação de padrões descritivos dos arquivos, por meio do uso de normas, visa facilitar o compartilhamento da informação arquivística, criando significado aos links criados nos acervos digitais.

Aqui utilizamos o conceito de informação arquivística, que é a informação descritiva e extraída no decorrer da aplicação de normas, como a ISAD-G e a NOBRADE, e utilizando-se padrões como EAD e EAC. Menciona-se a Informação Arquivística a sua principal característica de ser a informação contida na descrição do conjunto de documentos arquivísticos. Assim, pode-se considerar o material de arquivo, aquele que é produto do funcionamento de sistemas informacionais de arquivos e os que apoiam a realização de atividades, como a Informação Arquivística.

O uso de padrões descritivos como a ISAD-G cria uma normalização de elementos informacionais comuns para todos arquivos – os metadados -, e se extrai deles os contextos orgânicos variados, gerados numa realidade única para cada arquivo, como também se cria um padrão de interoperabilidade e a possibilidade de aplicar a visão da Web 3.0 nos arquivos, isto é, que utiliza a Informação Arquivística, a partir dos instrumentos descritivos, como mais uma Ontologia de Descrição Arquivística.

Cria-se, também, um padrão de interoperabilidade e a possibilidade de aplicar as técnicas da Web 3.0 para o arquivo, formando uma abordagem de uma Arquivologia contemporânea, isto é, que identifica a Informação Arquivística a partir dos instrumentos descritivos, como mais uma Ontologia de Descrição Arquivística na realidade da Web Semântica.

No caso brasileiro, iniciativas de interoperabilidade de dados são praticamente inexistentes, todavia já existem boas práticas internacionais que podem ser aplicadas quando projetos deste tipo de forem realizados. Além disso, a expectativa é que se aumente os processos de digitalização da informação, o que facilita na estruturação da interoperabilidade, já que os conteúdos indexados geram mais *corpus* de pesquisa e, conseqüentemente, podem melhorar a informação a ser oferecida ao usuário.

Mesmo que desbravem o emaranhado digital atualmente existente, o processo de organização da informação é de fato necessário. O procedimento primordial é a criação de metadados descrevendo a informação arquivística, principalmente para que os computadores possam processar seu significado e possibilitar o relacionamento entre dados.

Hoje, o relacionamento entre dados é realizado pelo usuário diretamente nas interfaces das plataformas digitais. Cabe, deste modo, ao profissional da informação, estabelecer os padrões de descrição da informação e de estruturação tecnológica, uma vez que a informação melhor descrita vai ser mais relevante ao usuário numa plataforma digital.

A forma de busca e uso da informação ainda é prioritariamente sintática, ou seja, o sistema entende caractere por caractere o que está escrito. Os sistemas e os mecanismos de busca não fazem deduções das associações que é preciso fazer com os termos utilizados nas buscas, desconsiderando variabilidade de nuances das línguas humanas. Um exemplo desse fato é uma sequência de números qualquer, que pode significar uma coordenada geográfica, um código postal, uma parte de um número telefônico, uma senha, um CPF, entre outros elementos.

Pode-se constatar que, no contexto da Web Semântica, que é o contexto dos links com significado para os agentes computacionais e o da arquivística, com suas informações resultado dos padrões de descrição arquivísticos, a existência de Ontologias Arquivísticas, o que é um indicativo da intersecção dessas áreas e, ainda, que podem haver iniciativas de representação dessa informação por meio de metadados, utilizando os padrões adotados universalmente e garantindo, desta forma, a interoperabilidade. Isto é muito importante, visto que o acesso de longo prazo para as instituições representa acumulação de arquivos digitais originados em períodos diferentes, em diversos formatos e tecnologias, mas que devem ser recolhidos e preservados em uma forma fixa e acessível para ser fonte de prova confiável.

A autenticidade e o acesso a esse acervo só se tornam possível com a manutenção da cadeia de custódia do documento e, nesse aspecto, as ontologias podem ser elementos constitutivos de sistemas informatizados e da rede, contribuindo com o registro do contexto arquivístico. Assegura-se, desta forma, que a informação arquivística (documento + contexto) evolui entre plataformas e sistemas até chegar a uma versão de Repositório Digital Confiável (RDC).

Por meio da visão arquivística, ao avaliar a Web Semântica, buscou-se identificar a convergência da informação orgânica na estrutura da rede, como forma de verificar pontos de atenção num processo de planejamento de seu acervo digital. Dessa forma, criou-se o diálogo entre as literaturas das áreas de Tecnologia da Informação, Ciência da Informação e Arquivologia, abordando elementos técnicos de cada campo.

Por isso é cada vez mais necessário aprofundarmos os estudos sobre a criação de políticas de descrição para a Web Semântica. Para tanto, o caminho de buscar padrões é o mais indicado. O compartilhamento de informação arquivística em meio digital pode ser uma realidade na Web Semântica, contudo é necessário aprimorar a gestão da informação nos serviços de informação e arquivos, considerando essa nova dimensão e a utilização destas técnicas para a preservação é outra frente de pesquisa observada a partir deste livro.

Posfaciar sobre ontologias

O papel das ontologias é abastecer o cérebro das máquinas. Ontologia é a ferramenta que faz a ponte entre a realidade dos homens e a realidade das máquinas, cada vez mais inteligentes com a internet das coisas. Mais do que uma moda na área de organização da informação, a ontologia surge como a ferramenta que faz a ponte entre a realidade dos homens e a realidade das máquinas.

Neste caso, as máquinas que começam a ter mais inteligência por conta da internet das coisas e dos algoritmos de inteligência artificial, que cada vez mais operam a informação humana digital. Foi Alan Turing que previu que “a inteligência das máquinas se tornaria tão penetrante, tão confortável e tão bem integrada à nossa economia baseada em informação que as pessoas sequer conseguiriam se dar conta disso.” (KURZWEIL, 2007, p. 108).

Porém, para que as máquinas fiquem cada vez mais inteligentes é que surgem ferramentas como as ontologias. A ontologia pode ser considerada como uma declaração de realidade, na qual é verificada a existência e a relação entre conceitos em determinados campos de conhecimento. Isto é, podemos criar, por exemplo, uma ontologia sobre o futebol, onde as regras (do pênalti ao impedimento), os envolvidos (do bandeirinha ao goleiro), o espaço físico (do campo ao vestiário), além dos objetivos do jogo são relacionados, formando laços entre si e mostrando sua interação formando sentidos neste pequeno mundo que é o futebol.

Uma forma de definir ontologia é [...] uma conceitualização formal de um domínio ou de uma parcela de realidade, com a qual podem operar diferentes aplicações de software. Os conceitos ou termos utilizados para a descrição servem como vocabulário comum (sintático e semântico) que favorece a comunicação e a interoperabilidade de recursos. Dão sentido pleno à informação ao situá-la dentro de um contexto. (MOREIRO GONZÁLES, 2011. p. 151).

Logo, é formal porque deve seguir determinados padrões e trazer sentidos comuns daquele determinado campo.

É nesse cenário que entra o ambiente de aplicação das ontologias que é a web semântica. Este é o caminho que segue nossa rede concebida em ambiente gráfico e estruturada pelos sistemas de organização do conhecimento existentes (taxonomias presentes no universo sintático).

No ambiente da web semântica, as taxonomias são complementadas por ontologias e, por isso, cabe destacar o papel da organização da informação, tais como as técnicas de descrição arquivísticas, taxonomias navegacionais e outras que são realizadas e disponibilizadas em ambientes digitais.

O caminho da tecnologia da informação é incorporar cada vez mais as ontologias nas estruturas de bancos de dados e ambientes digitais, que abastecem os “cérebros” computacionais. É neste momento que humanizamos as máquinas, mostrando um pedaço de como é complexo nosso universo humano, mesmo que seja uma fatia da realidade ou um campo de conhecimento específico.

REFERÊNCIAS

- ANGJELI, A. *Linking Data Across Libraries, Archives, and Museums*, 2012. #SAA ANNUAL MEETING12, session 401, 2012. Disponível em: <<http://files.archivists.org/conference/sandiego2012/401-Angjeli.pdf>>. Acesso em: 04 fev. 2016.
- 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, set. 2014. Disponível em: <<http://dx.doi.org/10.1590/1981-5344/1736>>. Acesso em: 22 fev. 2015.
- ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, v. 32, n. 3, p. 7-20, 2003. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652003000300002>. Acesso em: 22 fev. 2015.
- ALVES, R. C. V. *Metadados como elementos do processo de catalogação*. 2010. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, 2010. Disponível em: <https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/alves_rachel.pdf>. Acesso em: 23 mar. 2015.
- AN, X. An integrated approach to records management: the Records Continuum model's purpose-oriented approach to records management changes the role of recordkeeping from reactive to proactive. *The information Management Journal*, v. 37, n. 4, jun./ago. 2013. Disponível em: <https://www.arma.org/bookstore/files/xiaomi_0703.pdf>. Acesso em: 18 abr. 2015.
- ANDRADE, L. V.; BRUNA, D.; SALES, W. N. **Classificação: uma análise comparativa entre a Classificação Decimal Universal - CDU e a Classificação Decimal de Dewey - CDD. Biblos**, [S.l.], v. 25, n. 2, p. 31-42, ago. 2012. Disponível em: <<http://www.seer.furg.br/biblos/article/view/2088/1497>>. Acesso em: 22 fev. 2015.
- ANDRADE, R. S.; SILVA, R. R. G. Uma nova geração de instrumentos arquivísticos de referência: a publicação dos produtos das descrições arquivísticas em meio eletrônico. In: SIMPÓSIO BAIANO DE ARQUIVOLOGIA, 2., 2009 Salvador, BA. *Anais...* Salvador, BA: 2009. Disponível em: <http://www.twiki.ufba.br/twiki/pub/Cridi/Publica%E7%F5es/3cna-ricardoandrade_01.pdf>. Acesso em: 22 nov. 2014.

ARAÚJO, C. J. A. **Um modelo para interoperabilidade entre instituições heterogêneas**. Tese (Doutorado em Ciência da Computação) - Universidade de São Paulo, Instituto de Matemática e Estatística, 2012. São Paulo: USP, 2012. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/45/45134/tde-08022013-111002/>>. Acesso em: 06 mar. 2015.

ARQUIVO NACIONAL. **Dicionário brasileiro de terminologia arquivística**. Rio de Janeiro: Arquivo Nacional, 2005. 232 p. Publicações Técnicas, 41. Disponível em: <<http://www.arquivonacional.gov.br/Media/Dicion%20Term%20Arquiv.pdf>>. Acesso em: 12 jan. 2015.

ALONSO, S. S.; URBÁN, M. Á. S.; LEGUINA, G. R. Sobre la interoperabilidad semántica en las descripciones archivísticas digitales. **Revista Española de Documentación Científica**, v. 31, n. 1, p. 18-31, jan.-mar. 2008. Disponível em: <<http://redc.revistas.csic.es/index.php/redc/article/download/410/422>>. Acesso em: 18 fev. 2015.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, maio 2001. Disponível em: <http://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf>. Acesso em: 26 nov. 2015.

BELLOTTO, H. L. **Arquivos permanentes**: tratamento documental. São Paulo: T. A. Queiróz, 1991.

BRAGA, K. S. Aspectos relevantes para a seleção de metodologia adequada à pesquisa social em Ciência da Informação. In: MUELLER, S. P. M. **Métodos para a pesquisa em Ciência da Informação**. Brasília: Thesaurus, 2007. p. 17-38.

BRASIL. **Lei n. 8.159, de 8 de janeiro de 1991**. Dispõe sobre a política nacional de arquivos públicos e privados e dá outras providências. Diário Oficial [da] República Federativa do Brasil. Brasília, DF, 9 de janeiro de 1991. Seção 1. p. 455. Disponível em: <http://www.planalto.gov.br/ccivil_03/LEIS/L8159.htm>. Acesso em: 08 abr. 2015.

BORGES, M. A.; MORALES, A. B. T. Representação ontológica da memória organizacional da mediação da interação educacional. In: COLÓQUIO LUSO-BRASILEIRO DE EDUCAÇÃO A DISTÂNCIA E ELEARNING, 3., Lisboa, Portugal. 2013. **Atas...** Lisboa: 2013. Disponível em: <<http://lead.uab.pt/OCS/index.php/CLB/club/paper/view/43>>. Acesso em: 09 maio 2015.

BUFREM, L. S. Relações construídas no campo de conhecimento da Ciência da Informação no Brasil: a literatura periódica científica em foco. *Inf. Inf.*, Londrina, v. 18, n. 3, p. 68–97, set./dez. 2013. Disponível em:

<http://www.uel.br/revistas/uel/index.php/informacao/article/download/16705/pdf_6>.

Acesso em: 06 out. 2014.

CAFÉ, L.; SALES, R. Organização da informação: conceitos básicos e breve fundamentação teórica. In: ROBREDO, J.; BRÄSCHER, M. (Orgs.). (2010). *Passeios pelo bosque da informação*: estudos sobre representação e organização da informação e do conhecimento. Brasília, DF: IBICT, 2010. pp. 115-129. Disponível em:

<<http://repositorio.ibict.br/bitstream/123456789/36/1/eroic.pdf>>. Acesso em: 15 nov. 2014.

CAMARGO, A. M. A.; BELLOTTO, H. L. *Dicionário de Terminologia Arquivística*. São Paulo: Secretaria da Cultura, 1996.

CHOO, C. W. *A organização do conhecimento*: como as organizações usam a informação para criar conhecimento, construir conhecimento e tomar decisões. São Paulo: SENAC, 2003.

CONSELHO INTERNACIONAL DE ARQUIVOS (CIA). *ISAD-G: Norma geral internacional de descrição arquivística*: segunda edição, adotada pelo Comitê de Normas de Descrição, Estocolmo, Suécia, 19-22 de setembro de 1999, versão final aprovada pelo CIA. Rio de Janeiro: Arquivo Nacional, 2000. Disponível em:

<http://www.conarq.arquivonacional.gov.br/Media/publicacoes/isad_g_2001.pdf>. Acesso em: 15 nov. 2014.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). *NOBRADE: Norma Brasileira de Descrição Arquivística*. Rio de Janeiro: Arquivo Nacional, 2006. Disponível em:

<<http://www.conarq.arquivonacional.gov.br/Media/publicacoes/nobrade.pdf>>. Acesso em: 15 nov. 2014.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de Documentos Eletrônicos (CTDE). *e-ARQ Brasil: modelo de requisitos para sistemas informatizados de gestão arquivística de documentos*. Rio de Janeiro: CONARQ-CTDE, 2011. Disponível em: <http://www.conarq.gov.br/images/publicacoes_textos/earqbrasil_model_requisitos_2009.pdf>. Acesso em: 12 jul. 2014.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de Documentos Eletrônicos (CTDE). *Glossário*. Rio de Janeiro: CONARQ-CTDE, 2009. Disponível em:

<http://www.conarq.gov.br/images/ctde/Glossario/2009glossario_v5.0_final.pdf>. Acesso em: 01 nov. 2015.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de Documentos Eletrônicos (CTDE). **Perguntas mais frequentes**. Web site, 2016. Disponível em: <<http://www.conarq.arquivonacional.gov.br/perguntas-mais-frequentes.html>>. Acesso em: 18 ago. 2014.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de Documentos Eletrônicos (CTDE). **Diretrizes para a implementação de repositórios arquivísticos digitais confiáveis: RDC-Arq** 2015. Rio de Janeiro: CONARQ-CTDE, 2015. Disponível em: <http://www.conarq.arquivonacional.gov.br/images/publicacoes_textos/diretrizes_rdc_arq.pdf>. Acesso em: 29 abr. 2014.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Resolução nº 39**, de 29 de abril de 2014. Disponível em: <<http://www.conarq.gov.br/legisla%C3%A7%C3%A3o/resolu%C3%A7%C3%B5es-do-conarq/281-resolu%C3%A7%C3%A3o-n%C2%BA-39,-de-29-de-abril-de-2014.html>>. Acesso em: 12 mai. 2015.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Resolução nº 20**, de 16 de julho de 2004. Disponível em: <<http://www.conarq.gov.br/legisla%C3%A7%C3%A3o/resolu%C3%A7%C3%B5es-do-conarq/262-resolu%C3%A7%C3%A3o-n%C2%BA-20,-de-16-de-julho-de-2004.html>>. Acesso em: 12 mai. 2015.

CORCHO, O.; FERNÁNDEZ-LÓPES, M.; GÓMEZ-PÉREZ, A. Methodologies, tools and languages for building ontologies. Where is their meeting point? IN: Data & Knowledge Engineering, n. 46, Elsevier Science, p. 41-64.

CRUZ, J. R. M. **Manual de Archivística**. Madrid: Editorial Pirâmide, 1996.

DAVENPORT, T. **Ecologia da informação**. São Paulo: Futura, 1998.

DEL CASTILLO GUEVARA, J.; MENA MUGICA, M. M. La gestión de documentos de archivo en el actual contexto organizacional y la introducción a la Norma ISO 15489. **ACIMED**, Ciudad de La Habana, v. 22, n. 1, mar. 2011. Disponível em: <http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352011000100005&lng=es&nrm=iso>. Acesso em: 23 fev. 2014.

DUBLIN CORE METADATA APPLICATIONS INICIATIVE (DCMI). **DCMI Specifications**, 2014. Disponível em: <<http://dublincore.org/specifications/>>. Acesso em: 24 nov. 2014.

DURANTI, L. Registros documentais contemporâneos como provas de ação. **Estudos Históricos**, Rio de Janeiro, v. 7, n. 13, p. 49-64, 1994. Disponível em: <<http://bibliotecadigital.fgv.br/ojs/index.php/reh/article/view/1976>>. Acesso em: 02 dez. 2014.

EVANS, J.; ROUCHE, N. Utilizing systems development methods in archival systems research: building a metadata schema registry. **Archival Science**, v. 4, n. 3-4, p. 315-334, 2004. Disponível em: <<http://link.springer.com/article/10.1007/s10502-005-2598-4>>. Acesso em: 12 mar. 2014.

FEITOZA, L. A.; SILVA, T. E. Regime de informação dos sistemas de arquivos das universidades públicas do Brasil. **Inf. Prof.**, Londrina, v. 1, n. 1/2, p. 40-55, jul./dez. 2012. Disponível em: <<http://www.uel.br/revistas/uel/index.php/infoprof/article/view/14587>>. Acesso em: 18 fev. 2014.

FLORES, D. **Ambiente autêntico de preservação e acesso em longo prazo de documentos arquivísticos digitais - RDC-Arq**. Apresentação, UNB, dez. 2014. Disponível em: <<http://pt.slideshare.net/dfloresbr/unb-nov2014-ambiente-autntico-de-preservao-e-acesso-em-longo-prazo-de-documentos-arquivsticos-digitais-rdcarq-prof-dr-daniel-flores-ufsm>>. Acesso em: 04 out. 2014.

GILCHRIST, A. Thesauri, taxonomies and ontologies: an etymological note. **Journal of Documentation**, v. 59, n. 1, p. 7-18, 2003. Disponível em: <<http://dx.doi.org/10.1108/00220410310457984>>. Acesso em: 15 nov. 2014.

GONÇALVES, J. **Como classificar e ordenar documentos de arquivos**. São Paulo: Divisão de Arquivo do Estado de São Paulo, 1998. Disponível em: <http://www.arqsp.org.br/arquivos/oficinas_colecao_como_fazer/cf2.pdf>. Acesso em: fev. 2015.

HASSO PLATTNER INSTITUTE. **The Linking Open Data cloud diagram**, 2014. Disponível em: <<http://lod-cloud.net/versions/2014-08-30/lod-cloud.svg>>. Acesso em: 14 dez. 2015.

HAGEN, A. M. M. Algumas considerações a partir do processo de padronização da descrição arquivística. **Ci. Inf.**, Brasília, v. 27, n. 3, set. 1998. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651998000300007&lng=en&nrm=iso>. Acesso em: 21 mar. 2015.

HEATH, T.; HEPP, M.; BIZER, C. Special Issue on Linked Data. ***International Journal on Semantic Web and Information Systems (IJSWIS)***, jan. 2008. Disponível em: <<http://linkeddata.org/docs/ijswis-special-issue>>. Acesso em: 22 nov. 2014.

HEREDIA HERRERA, A. ***Arquivística general: teoría y práctica***. Sevilla, Espanha: Diputación Provincial de Sevilla, 1991.

HINZ, V. T. ***Proposta de Criação de uma Ontologia de Ontologias***. 2006. Trabalho Individual I - Escola de Informática, Universidade Católica de Pelotas. Pelotas: UCPel, 2006. Disponível em: <<http://ppginf.ucpel.tche.br/TI-arquivos/2006/VerlaniHinz/PPGINF-UCPel-TI-2006-2-10.pdf>>. Acesso em: 20 maio 2014.

HODGE, G. ***Systems of knowledge organization for digital libraries: beyond traditional authority files***. Washington, DC: Council on Library and Information Resources, 2000. 37 p. Disponível em: <<http://www.clir.org/pubs/reports/reports/pub91/pub91.pdf>>. Acesso em: 22 mar. 2015.

INNARELLI, H. C. ***Gestão da preservação de documentos arquivísticos digitais***: proposta de um modelo conceitual. 2015. 348 f. Tese (Doutorado) - Escola de Comunicações e Artes, Universidade de São Paulo. São Paulo, 2015. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/27/27151/tde-27052015-101628/pt-br.php>>. Acesso em: 08 nov. 2015.

INNARELLI, H. C. Preservação digital: a gestão e a preservação do conhecimento explícito digital em instituições arquivísticas. ***InCID: Revista de Ciência da Informação e Documentação***, v. 3, n. 2, p. 48-63, dez. 2012. Disponível em: <<http://www.revistas.usp.br/incid/article/view/48653/52724>>. Acesso em: 06 fev. 2014.

INTERNATIONAL RESEARCH ON PERMANENT AUTHENTIC RECORDS IN ELECTRONIC SYSTEMS (INTERPARES). ***InterPARES 3 Project. General Study 01 – Past and Present Digital Preservation Projects Devoted to the Long-term Preservation of Digital Records and Digital Information***. The InterPARES 3 Project, TEAM Canada, nov. 2007. Disponível em: <http://www.interpares.org/ip3/display_file.cfm?doc=ip3_gs01_digital_preservation_project_v1-4p.pdf>. Acesso em: 02 ago. 2015.

INTERNATIONAL RESEARCH ON PERMANENT AUTHENTIC RECORDS IN ELECTRONIC SYSTEMS (INTERPARES). ***InterPARES 2 Project - Diretrizes do produtor. A elaboração e a manutenção de materiais digitais: diretrizes para indivíduos***. The InterPARES 2 Project, 2002-2007. Disponível em: <http://www.interpares.org/ip2/display_file.cfm?doc=ip2_creator_guidelines_booklet--portuguese.pdf>. Acesso em: 07 dez. 2015.

KURZWEIL, Ray. **A era das máquinas espirituais**. São Paulo: Aleph, 2007.

LARA, M. L. G. **Linguística documentária**: seleção de conceitos. 2009. 1746 p. Tese (Livre Docência) - Departamento de Biblioteconomia e Documentação, Escola de Comunicações e Artes, Universidade de São Paulo. São Paulo, 2009.

LÉVY, P. **A inteligência coletiva**: por uma antropologia do ciberespaço. 4. ed. São Paulo: Loyola, 2003.

LIMA, G. A. B. O. < **MHTX** > **Modelagem hipertextual para organização de documentos**: princípios e aplicação. Rio de Janeiro: Interciência, 2015. 208 p.

LOPES, L. C. **A informação e os arquivos**: teorias e práticas. Niterói, RJ: EDUFF; São Carlos: EDUFSCar, 1996.

LOPES, L. C. **A nova arquivística na modernização administrativa**. Brasília: Projecto Editorial, 2009.

LOPEZ, A. P. A. **Como descrever documentos de arquivo**: elaboração de instrumentos de elaboração de instrumentos de pesquisa. São Paulo: Arquivo do Estado, Imprensa Oficial do Estado, 2002.

LUMSDEN, J; HALL, H; CRUICKSHANK, P. Ontology definition and construction, and epistemological adequacy for systems interoperability: a practitioner analysis. **Journal of Information Science**, v.37, p. 246, 2011. Disponível em: < <http://jis.sagepub.com/content/37/3/246.abstract> >. Acesso em: 12 jun. 2015.

LUZ, C. **Arquivologia 2.0**: a informação digital humana. Excertos de um arquivista 2.0 no mundo digital. Florianópolis: Bookess, 2010.

LINKED OPEN VOCABULARIES (LOV). **Web site**, 2015. Disponível em: <<http://lov.okfn.org/dataset/lov/>>. Acesso em: fev. 2015.

MARCONDES, C. H. Metadados: descrição e recuperação na Web. In: MARCONDES, C. H. et al. (Org.). **Bibliotecas digitais: saberes e práticas**. Salvador, BA: EDUFBA; Brasília, DF: IBICT, 2005. p. 77-143. Disponível em: <<http://www.brapci.ufpr.br/download.php?dd0=11228>>. Acesso em: 22 nov. 2014.

MARCONDES, C. H. Linked Data: dados interligados e interoperabilidade entre arquivos, bibliotecas e museus na Web. *Encontros Bibli*, v. 17, n. 34, p. 171-192, maio/ago. 2012. Disponível em: <10.5007/1518-2924.2012v17n34p171>. Acesso em: 22 nov. 2014.

MARTÍN-BARBERO, J. *Dos meios às mediações*: comunicação, cultura e hegemonia. Rio de Janeiro: UFRJ, 2001.

MAZZINI, S.; RICCI, F. EAC-CPF Ontology and Linked Archival Data. Semantic Digital Archives (SDA). INTERNATIONAL WORKSHOP ON SEMANTIC DIGITAL ARCHIVES, 1., SDA, 2011. *Proceedings....* Disponível em: <<http://ceur-ws.org/Vol-801/paper6.pdf>>. Acesso em: 18 mar. 2015.

MOREIRA, W.; SANTOS NETO, M. F. A formação do conceito de ontologia na Ciência da Informação: uma análise nos periódicos Scire e Ibersid. *Scire*, v. 20, n. 2, p. 49-54, jul./dez. 2014. Disponível em: <<http://www.ibersid.eu/ojs/index.php/scire/article/viewFile/4173/3774>>. Acesso em: 22 nov. 2014.

MOREIRO GONZÁLEZ, J. A. *Linguagens documentárias e vocabulários semânticos para a web*: elementos conceituais. Salvador: EDUFBA, 2011a.

MOREIRO GONZÁLEZ, J. A. Evolução ontológica das linguagens documentárias. Relato de uma experiência de curso organizado conjuntamente para o DT/SIBI-USP e o PPGCI/ECA. *InCID: Revista de Ciência da Informação e Documentação*, v. 2, n. 1, p. 143-164, jun. 2011b. Disponível em: <<http://www.revistas.usp.br/incid/article/view/42339>>. Acesso em: 22 nov. 2014.

NEGREIROS, L. R. *Sistemas eletrônicos de gerenciamento de documentos arquivísticos: um questionário para escolha, aplicação e avaliação*. 2007. 246 f. Dissertação (Mestrado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Escola de Ciência da Informação, UFMG, 2007.

NATIONAL INFORMATION STANDARDS ORGANIZATION (NISO). *Understanding Metadata*. Bethesda: NISO Press, 2004. Disponível em: <<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>>. Acesso em: 21 set. 15.

PATRICIO, H. A Europeana e a agregação de metadados na Web: análise dos esquemas ESE/EDM e da aplicação de standards da Web Semântica a dados de bibliotecas. CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 11., 2012. *Actas....* Lisboa: Fundação Calouste Gulbenkian, out. 2012. Disponível em: <

<http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/458>>. Acesso em: 22 out. 2014.

PEDRAZA-JIMÉNEZ, R.; CODINA, L.; ROVIRA, C. Web semántica y ontologías en el procesamiento de la información documental. *El profesional de la información*, v. 16, n. 6, p. 569-578, 2007. Disponível em: <<http://www.elprofesionaldelainformacion.com/contenidos/2007/noviembre/04.pdf>>. Acesso em: 22 out. 2014.

PIRES, D. Uso do *Dublin Core* na descrição de obras raras na Web: a coleção da Biblioteca Brasileira Digital. Fundação Biblioteca Nacional. In: ENCONTRO NACIONAL DE CATALOGADORES, 1.; ENCONTRO DE ESTUDOS E PESQUISAS EM CATALOGAÇÃO, 3., Rio de Janeiro, set. 2012. *Anais...* Rio de Janeiro: 2012. Disponível em: <<http://www.producao.usp.br/handle/BDPI/43413>>. Acesso em: 22 out. 2014.

QUISBERT, H. *A framework for the development of archival information systems*. 2006. Tese (Doutorado) - Department of Business Administration and Social Sciences, Luleå University of Technology, Luleå – Suécia, 2006. Disponível em: <<http://epubl.ltu.se/1402-1757/2006/62/LTU-LIC-0662-SE.pdf>>. Acesso em: 22 out. 2014.

RAMALHO, R. A. S.; VIDOTTI, S. A. B. G.; FUJITA, M. S. L. Web Semântica: uma investigação sob o olhar da Ciência da Informação. *DataGramaZero - Revista de Ciência da Informação*, v. 8, n. 6, dez. 2007. Disponível em: <http://www.dgz.org.br/dez07/Art_04.htm>. Acesso em: 26 nov. 2014.

REPOSITORY FOR LINKED OPEN ARCHIVAL DATA (RELOAD). *ReLoad Repository for linked open archival data Istituto per i beni artistici culturali e naturali Regione Emilia-Romagna*, 2014. Disponível em: <<http://labs.regesta.com/progettoReload/>>. Acesso em: 27 nov. 2014.

RIBEIRO, C. F. A. Organizar e representar informação: apenas um meio para viabilizar o acesso. *Revista da Faculdade de Letras, Ciências e Técnicas do Patrimônio*, Faculdade de Letras da Universidade do Porto, v. 4, p. 83-100, 2005. Disponível em: <<http://ler.letras.up.pt/uploads/ficheiros/1239.pdf>>. Acesso em: 22 mar. 2015.

RIBEIRO, C. F. A. *O acesso à informação nos arquivos*. Tese (Doutorado) - Faculdade de Letras da Universidade do Porto, Porto, 1998. Disponível em: <<http://repositorio-aberto.up.pt/bitstream/10216/7058/4/fribeirovol02000061437.pdf>>. Acesso em: 19 mar. 2015.

ROBREDO, J. *Da Ciência da Informação revisitada aos sistemas humanos de informação*. Brasília, DF: Thesaurus; SSRR Informações, 2003.

ROBREDO, J. Ciência da Informação e Web Semântica: linhas convergentes ou linhas paralelas? In: ROBREDO, J.; BRÄSCHER, M. (Orgs.). **Passeios pelo bosque da informação**: estudos sobre representação e organização da informação e do conhecimento. Brasília, DF: IBICT, 2010. p. 12-47. Disponível em: <<http://www.ibict.br/publicacoes/eroic.pdf>>. Acesso em: 22 out. 2014.

ROCHA, C. L. Glossário multilíngue do Projeto InterPARES 3. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, n. esp., 1º sem. 2011. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2011v16nesp1p76>>. Acesso em: 18 abr. 2015.

ROUSSEAU, J.; COUTURE, C. **Os fundamentos da disciplina arquivística**: glossário. Lisboa: Dom Quixote, 1998.

RUESTA, C. B. **Série ISO 30300: Sistema de gestão para documentos de arquivos**. Tradução do Grupo de Trabalho de Gestão de Documentos de Arquivo. Lisboa: Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas, 2012. Disponível em: <http://www.bad.pt/publicacoes/Serie_ISO_30300.pdf>. Acesso em: 23 set. 2014.

SANTOS, V. B. **A teoria arquivística a partir de 1898**: em busca da consolidação, da reafirmação e da atualização de seus fundamentos. 2011. 279 f. Tese (Doutorado) – Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2011. Disponível em: <http://repositorio.unb.br/bitstream/10482/10323/3/2011_VanderleiBatistaSantos.pdf>. Acesso em: 23 set. 2014.

SANTOS, H. M.; FLORES, D. Repositórios digitais confiáveis para documentos arquivísticos: ponderações sobre a preservação em longo prazo. **Perspectivas em Ciência da Informação**, v. 20, n. 2, p. 198-218, abr./jun. 2015. Disponível em: <<http://dx.doi.org/10.1590/1981-5344/2341>>. Acesso em: 14 dez. 2014.

SAYAO, L. F.; MARCONDES, C. H. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **TransInformação**, Campinas, v. 20, n. 2, p. 133-148, maio/ago., 2008. Disponível em: <<http://dx.doi.org/10.1590/S0103-37862008000200002>>. Acesso em: 23 set. 2014.

SCHELLEMBERG, T. R. **Arquivos modernos: princípios e técnicas**. Rio de Janeiro: Fundação Getúlio Vargas, 2004.

SCHMIDT, C. M. S. Entre o documento de arquivo e a informação arquivística: reflexões acerca do objeto científico da Arquivologia. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 14., Santa Catarina, 2013. **Anais....** Disponível em: <

<http://enancib2013.ufsc.br/index.php/enancib2013/XIVenancib/paper/viewFile/408/242> >. Acesso em: 22 out. 2014.

SILVA, E. P. A trajetória da Arquivologia: três visões sobre os arquivos. **Revista Eletrônica Documento Monumento**, Mato Grosso, v. 5, n. 1, dez. 2011. Disponível em: <<http://pt.scribd.com/doc/81374141/A-TRAJETORIA-DA-ARQUIVOLOGIA#scribd>>. Acesso em: 14 jun. 2014.

SMIT, J. A informação na Ciência da Informação. **IncID: R. Ci. Inf. e Doc.**, Ribeirão Preto, v. 3, n. 2, p. 84-101, jul./dez. 2012. Disponível em: <www.revistas.usp.br/incid/article/download/48655/52726>. Acesso em: 14 maio 2015.

SMIT, J. W.; BARRETO, A. A. Ciência da Informação: base conceitual para a formação do profissional. In: VALENTIM, M. L. P. (Org.). **Formação do profissional da informação**. São Paulo: Polis, 2002.

SMITH, B. Ontology, in The Blackwell Guide to the Philosophy of Computing and Information, **Blackwell Publishing**, Oxford, UK., 153-166, 2008. Disponível em: <[10.1002/9780470757017.ch11](https://doi.org/10.1002/9780470757017.ch11)>. Acesso em: 22 jan. 2015.

SOCIETY OF AMERICAN ARCHIVISTS (SAA). Technical Subcommittee on EAC-CPF. **Encoded Archival Context: Corporate bodies, Persons, and Families (EAC-CPF)**. Chicago: SAA, 2011. Disponível em: <<http://www2.archivists.org/groups/technical-subcommittee-on-eac-cpf/encoded-archival-context-corporate-bodies-persons-and-families-eac-cpf>>. Acesso em: 25 jan. 2016.

SOUZA, R. R.; ALVARENGA, L. A Web Semântica e suas contribuições para a Ciência da Informação. **Ciência da Informação**, [S.l.], v. 33, n. 1, jun. 2004. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/50/50>>. Acesso em: 21 set. 2015.

SOUZA, R. T. B. Alguns apontamentos sobre a classificação de documentos de arquivo. **Brazilian Journal of Information Science: Research Trends**, v. 8, n. 1/2, 2014. Disponível em: <<http://dx.doi.org/10.5016/10.5016/1981-16>>. Acesso em: 12 out. 2015.

STANDARDS AUSTRALIA INTERNATIONAL. **Australian Standard AS 4390, Records Management**, 1996.

STANDARDS AUSTRALIA INTERNATIONAL. **Australian Standard, Records Management: Part 1: General** - AS ISO 15489.1-2002. Disponível me: <<https://www.saiglobal.com/pdftemp/previews/osh/as/as10000/15000/154891.pdf>>. Acesso em: 12 ago. 2015.

THE LIBRARY OF CONGRESS (LOC). **Encoded Archival Description Tag Library, Version 2002. Appendix A: EAD Crosswalks**. Disponível em: <http://www.loc.gov/ead/tglib/appendix_a.html>. Acesso em: 12 ago. 2015.

THE LIBRARY OF CONGRESS (LOC). Standards. **MARC SGML Archive**. Network Development and MARC Standards Office Library of Congress. maio 2003. Disponível em: <<https://www.loc.gov/marc/marcsgmlarchive.html>>. Acesso em: 12 ago. 2015.

THE LIBRARY OF CONGRESS (LOC). Standards. **MADS Metadata Authority Description Schema**. jun. 2013. Disponível em: <<http://www.loc.gov/standards/mads/>>. Acesso em: 12 ago. 2015.

THE LIBRARY OF CONGRESS (LOC). Standards. **EAD Encoded Archival Description**. jan. 2016a. Disponível em: <<https://www.loc.gov/ead/>>. Acesso em: jan. 2016.

THE LIBRARY OF CONGRESS (LOC). Standards. **ALTO Technical Metadata for Optical Character Recognition**. jan. 2016b. Disponível em: <<https://www.loc.gov/standards/alto/>>. Acesso em: jan. 2016.

THE LIBRARY OF CONGRESS (LOC). Standards. **PREMIS Preservation Metadata Maintenance Activity**. jan. 2016c. Disponível em: <<http://www.loc.gov/standards/premis/>>. Acesso em: jan. 2016.

THE LIBRARY OF CONGRESS (LOC). Standards. **METS - Metadata encoding & transmission Standard**. fev. 2016d. Disponível em: <<http://www.loc.gov/standards/mets/>>. Acesso em: jan. 2016.

THE LIBRARY OF CONGRESS (LOC). Standards. **PREMIS OWL Ontology**. jul. 2015a. Disponível em: <<http://www.loc.gov/standards/premis/ontology/index.html>>. Acesso em: jan. 2016.

THE LIBRARY OF CONGRESS (LOC). Standards. **NISO Metadata for Images in XML Schema**. Nov. 2015b. Disponível em: <<http://www.loc.gov/standards/mix/>>. Acesso em: 12 dez. 2015.

THE OPEN ARCHIVES INITIATIVE PROTOCOL FOR METADATA HARVESTING (OAI-PMH). **Protocol Version 2.0**, 2002. Disponível em: <<http://www.openarchives.org/OAI/openarchivesprotocol.html>>. Acesso em: 22 nov. 2014.

THOMASSEM, T. Uma primeira introdução à Arquivologia. **Arquivo & Administração**, Rio de Janeiro, v. 5, n. 1, p. 5-16, jan./jun. 2006.

TOMAÉL, M. I. Redes de informação: o ponto de contato dos serviços e unidades de informação no Brasil. *Inf. & Inf.*, Londrina, v. 1-2, jan./ dez. 2005. Disponível em: < <http://www.uel.br/revistas/uel/index.php/informacao/article/viewFile/1611/1366>>. Acesso em: 14 fev. 2015.

TRZESNIAK, P.; KOLLER, S. Tornar-se semântico: adequando nossa revista ao tempo de informação por toda parte. *Revista Interamericana de Psicología/Interamerican Journal of Psychology*, v. 39, n. 2, p. 185-188, 2005. Disponível em: <<http://www.psicorip.org/Resumos/PerP/RIP/RIP036a0/RIP03920.pdf>>. Acesso em: 12 ago. 2014.

UNIRIO. *Curso de Arquivologia*. Web site, 2015. Disponível em: <<http://www2.unirio.br/unirio/cchs/arquivologia/a-arquivologia>>. Acesso em: 22 nov. 2015.

UPWARD, F. Structuring the records continuum (Series of two parts) Part 1: post custodial principles and properties. *Archives and Manuscripts*, v. 24, n. 2, p. 268–285, 1996. Disponível em: <https://www.zotero.org/groups/records_continuum_research_group_publications/items/it emKey/RRBMNBKU>. Acesso em: 24 out. 2015.

VICKERY, B.C. Knowledge Representation: a brief review. *Journal of Documentation*, n. 42, 1986. Disponível em: < <http://dx.doi.org/10.1108/eb026790> >. Acesso em: 24 out. 2015.

WORLD WIDE WEB CONSORTIUM (W3C). *Website*, 2015. Disponível em: < <http://www.w3.org/>>. Acesso em: 12 ago. 2015.

WERSIG, G.; NEVELING, U. *Terminology of documentatio =: Terminologie de la documentation = Terminologie der Dokumentation = Terminologiia v oblasti dokumentatsii = Terminología de la documentación*: a selection of 1200 basic terms published in English, French, German, Russian, and Spanish. Paris: The Unesco Press, 1976.

YAKEL, Elizabeth. KIM, Jihyun. Adoption and diffusion of Encoded Archival Description. *Journal of the American Society for Information Science and Technology*, v. 56, n. 13, p. 1427-1437, 2005. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.20236/abstract>>. Acesso em: 12 ago 2015.

ZENG, M. L.; HODGE, G. Developing a *Dublin Core* Application Profile for the knowledge organization systems (KOS) resources. *Bulletin of the American Society for Information Science and Technology*, v. 37, n. 4, p. 30-34, 2011. Disponível em: < <http://onlinelibrary.wiley.com/doi/10.1002/bult.2011.1720370409/abstract>>. Acesso em: 12 ago 2015.



Ex-libris nostrum quotidianum

Ontologia Digital Arquivística

Na tecnologia digital, é possível que a relação dos padrões de descrição arquivísticos, representados por meio de metadados, estimulam a criação de ontologias leves com potencial uso no processamento e na interoperabilidade?

O livro mostra este caminho, onde a descrição arquivística permite acesso a informação arquivística registrada e é capaz de fornecer dados acerca do contexto de produção documental para sistemas de arquivo e para a Web 3.0.

Para o estudo que culmina nesta publicação, foram utilizados os padrões dos sistemas informatizados de arquivos, como o Sistema Informatizado de Gestão Arquivística de Documentos (SIGAD) e Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos (e-ARQ), além dos Repositórios Digitais Confiáveis Arquivísticos (RDCArq) e o Modelo Record Continuum.

As ontologias têm seu papel na estruturação dos domínios de conhecimento e seu uso na tecnologia digital. Por isso falamos sobre as características das ontologias que podem resultar do uso de padrões arquivísticos e o resultado disso como uma Knowledge Organization System (KOS).

Para a arquivística, é possível a aplicação de ontologias para auxiliar na manutenção da cadeia de custódia e do contexto da informação arquivística e do documento arquivístico, além de seu uso na difusão e na preservação digital em longo prazo, tanto em sistemas de arquivo quanto na rede mundial de computadores.

Charlley Luz

