

UNIVERSIDADE ESTADUAL PAULISTA
Campus de Marília
Faculdade de Filosofia e Ciências
Programa de Pós-Graduação em Ciência da Informação

FERNANDO ALVES DA GAMA

**AS CONTRIBUIÇÕES DAS LINGUAGENS DE MARCAÇÃO PARA A GESTÃO DA
INFOMAÇÃO ARQUIVÍSTICA DIGITAL**

MARÍLIA-SP
2011

FERNANDO ALVES DA GAMA

**AS CONTRIBUIÇÕES DAS LINGUAGENS DE MARCAÇÃO PARA A GESTÃO DA
INFOMAÇÃO ARQUIVÍSTICA DIGITAL**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências – Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP – Campus de Marília, como requisito para a obtenção do título de Mestre em Ciência da Informação.

Área de Concentração: Informação, Arquivologia, Tecnologia e Conhecimento.

Linha de Pesquisa: Informação e Tecnologia

Orientador: Prof. Dr. Edberto Ferneda

Financiamento: CAPES

MARÍLIA-SP
2011

G184c

Gama, Fernando Alves da

As contribuições das linguagens de marcação para a gestão da
informação arquivística digital / Fernando Alves da Gama – Marília,
2011.

165 f.; 30 cm.

Dissertação (Mestrado – Ciência da Informação) -

Universidade Estadual Paulista, Faculdade Filosofia e Ciências, 2011.

Bibliografia: f. 156-165

Orientador: Edberto Ferneda

1. Informação e tecnologia. 2. Gestão da Informação. 3. Gestão
de documentos. 4. Preservação digital. 5. Ambiente informacional
digital. I. Autor. II Título.

CDD 005.73

CDU 005.92

FERNANDO ALVES DA GAMA

**AS CONTRIBUIÇÕES DAS LINGUAGENS DE MARCAÇÃO PARA A GESTÃO DA
INFOMAÇÃO ARQUIVÍSTICA DIGITAL**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências – Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP – Campus de Marília, como requisito para a obtenção do título de Mestre em Ciência da Informação.

BANCA EXAMINADORA

Presidente e Orientador:

Prof. Dr. Edberto Ferneda
Departamento de Ciência da Informação
Faculdade de Filosofia e Ciências
Universidade Estadual Paulista – UNESP, Campus Marília

Membro Titular

Prof.^a Dra. Telma Campanha de Carvalho Madio
Departamento de Ciência da Informação
Faculdade de Filosofia e Ciências
Universidade Estadual Paulista – UNESP, Campus Marília

Membro Titular

Prof.^a Dra. Maria Elisabete Catarino
Departamento de Ciência da Informação
Universidade Estadual de Londrina – UEL

MARÍLIA-SP
2011

DEDICATÓRIA

*Dedico esse trabalho a todos os estudantes
e pesquisadores que trabalham em prol do
desenvolvimento da Ciência da Informação.*

AGRADECIMENTOS

A Deus, primeiramente, por prover vida, saúde e todos os recursos necessários ao desenvolvimento do trabalho. A Ele, pois, a glória eternamente.

Aos meus pais, Jeomar e Odete, por todo o apoio dado no desenvolvimento deste trabalho.

Aos meus irmãos, Inaldo A. Gama, Claudemir Zani e Antônia Zani, por todo o companheirismo e auxílio prestados.

Aos amigos Claudio Mainer e José Cavalcanti da Silva por todo o apoio e incentivo.

Aos professores da Universidade Estadual de Londrina, em especial, Maria Elisabete Catarino, Wilmara R. Calderon e Renata G. Curty, pelo incentivo dado aos alunos de graduação pelo ingresso na iniciação científica.

Aos colegas da Caixa Econômica Federal, Prefeitura de Londrina e do Ministério Público do Paraná, pelo auxílio dado, em especial: Elias K. Marutani, Luís H. Imai, João P. Silva, Eufrásio Valencia, Antônio C. Baraldi e Eliezer G. Silva.

Ao meu orientador de pesquisa no curso de mestrado, Prof. Dr. Edberto Ferneda, pela paciência, leituras, sugestões, competência, compreensão e confiança.

A todos os professores do Programa de Pós-Graduação, pelos quais tenho grande admiração e, em especial, às professoras Silvana A. B. G. Vidotti, Plácida L. V. A. da Costa Santos e Marta L. P. Valentim.

Aos colegas estudantes da UNESP por todo o companheirismo e incentivos prestados, em especial: Regis Garcia, Rodrigo Sales, Letícia G. Molina, e Mirian Mattos.

A todos os autores citados neste trabalho.

A todos os leitores deste trabalho.

GAMA, Fernando Alves da. **As contribuições das linguagens de marcação para a gestão da informação arquivística digital**. 2011. 165f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2011.

RESUMO

O presente trabalho desenvolve um diálogo entre a Ciência da Informação, Arquivologia e Tecnologia da Informação visando apresentar os benefícios que as linguagens de marcação podem trazer à gestão de documentos arquivísticos por meio da representação de ferramentas de classificação e a preservação de documentos digitais. A arquivologia visa ao tratamento da informação orgânica por meio do tratamento dado aos documentos de arquivo, que são produzidos por um indivíduo ou instituição no exercício de suas atribuições. Tais documentos são submetidos a dois grupos de usuário, cujo comportamento informacional está ligado ao uso do valor primário, de caráter administrativo, na fase corrente e intermediária e, posteriormente, ao valor secundário, de caráter informativo, na fase da guarda permanente. Este trabalho busca trazer um delineamento no conjunto de metadados necessários à representação do conteúdo, contexto e estrutura documentais voltadas às necessidades dos usuários. São apresentadas, na sequência, as linguagens de marcação e as modelagens da web semântica como estruturas para a representação de metadados. Uma vez delineadas estas questões, o estudo direciona-se ao comportamento informacional dos usuários da fase corrente e sua forma de interação com a informação orgânica. Apresenta o programa de gestão documental e todo o processo metodológico envolto em sua consolidação, seguido de uma explanação em torno da produção de ferramentas de classificação documental que promovem o controle de vocabulário na gestão de documentos, a saber, Plano de Classificação e Tesauro de Função Australiano. Estas ferramentas se complementam e permitem confiabilidade e precisão na recuperação da informação orgânica em todas as fases da gestão de documentos. Elas possibilitam, ainda, a representação do contexto de produção documental a ser analisado pelos usuários da fase de guarda permanente. Na sequência, são apresentadas as linguagens de marcação XML e as modelagens em RDF, RDFS, e SKOS na representação das ferramentas de classificação no ambiente digital. Isso possibilita que os sistemas computacionais trabalhem com a estrutura de controle de vocabulário provida pelas ferramentas de classificação. O estudo verificou, ainda, as contribuições que a linguagem de marcação traz para a preservação digital por meio da representação de formatos binários de documentos digitais em documentos XML.

Palavras-chave: Ciência da Informação, Arquivologia, Gestão de Documentos, Recuperação da Informação, Documento Digital, Preservação Digital, Linguagens de Marcação, XML.

GAMA, Fernando Alves da. **The markup languages' contributions for the management of business information**. 2011. 165f. Dissertation (Master Degree in Information Science) – College of Philosophy and Sciences, Universidade Estadual Paulista, Marília, 2011.

ABSTRACT

This Project develops a dialogue among the Information Science, Archival Science and Information of Technology to show the benefits that markup language can bring to the management of archival documents through the representation of classification tools and the preservation of digital records. The main objective of Archival science is the treatment of business information, through the treatment given to the records, which are produced by an individual or institution in the performance of their duties. Such documents are submitted to two user groups, whose behavior is linked to the informational use of the primary value of an administrative nature, in the current phase and intermediate and then the secondary value of its information, at the stage of permanent custody. This Project seeks to bring in a design set of metadata required for the representation of content, context and structure documentary directed to the needs of users. They will be presented in sequence, markup languages and the modeling of the semantic web as structures for the representation of metadata. After that the study will move into the users' information behavior on the current phase and its interaction with the form of organic information. It displays the document management program and all the methodological process on its consolidation, followed by an explanation surrounding the production of documentary classification tools that promote vocabulary control in the management of documents, namely, Classification Plan and Australian Function's Thesaurus. These tools are complementary and provide reliability and accuracy in information retrieval organic in all stages of document management. They also allow the representation of the context of document production to be analyzed by users of the phase of the standing guard. Then we will present the XML markup languages and modeling in RDF, RDFS and SKOS representation of the classification tools in the digital environment. It enables computer systems to work with the structure of vocabulary control provided by tools of classification. The study will also verify the contributions that will bring the markup language for digital preservation through the representation of the binary formats of digital documents in XML documents.

Keywords: Information Science, Archival Science, Document Management, Information Retrieval, Digital Paper, Digital Preservation, markup languages, XML.

LISTA DE SIGLAS

AIP	<i>Archival Information Package</i>
CI	Ciência da Informação
DBML	<i>Database Markup Language</i>
DIP	<i>Dissemination Information Package</i>
DTD	<i>Document Type Definition</i>
GML	<i>Generalized Markup Language</i>
HTML	<i>Hypertext Markup Language</i>
ISO	<i>International Organization for Standardization</i>
OAIS	<i>Open Archival Information System</i>
OWL	<i>Web Ontology Language</i>
PGD	Programa de Gestão Documental
RDF	<i>Resource Description Framework</i>
RDFS	<i>Resource Description Framework Schema</i>
SGML	<i>Standard Generalized Markup Language</i>
SIARD	<i>Software Independent Archival of Relational Database</i>
SIP	<i>Submission Information Package</i>
SKOS	<i>Simple Knowledge Organization System</i>
TE	Termo Específico
TG	Termo Geral
TI	Tecnologia da Informação
TR	Termo Relacionado
UP	Usado Para
URI	<i>Uniform Resource Identifier</i>
W3C	<i>World Wide Web Consortium</i>

LISTA DE FIGURAS

FIGURA 01: EXEMPLO DE UTILIZAÇÃO DA LINGUAGEM HTML: CATÁLOGO DE ENDEREÇO. ...	45
FIGURA 02: CAMADAS DA WEB SEMÂNTICA.	56
FIGURA 03: MODELO DE DADOS GRÁFICO EM RDF.....	58
FIGURA 04: RELACIONAMENTO ATIVIDADE E TIPO DOCUMENTAL.	58
FIGURA 05: EXEMPLOS DE DECLARAÇÕES EM RDF.	59
FIGURA 06: ELEMENTOS DE UM TIPO DOCUMENTAL.....	60
FIGURA 07: CONTÊINERES EM BAG E SEQ.....	60
FIGURA 08: RELACIONAMENTO ENTRE FUNÇÃO ATIVIDADE E CARGO.....	75
FIGURA 09: EXEMPLO DE CLASSIFICAÇÃO DOCUMENTAL.	94
FIGURA 10: EXEMPLIFICAÇÃO DO RELACIONAMENTO HIERÁRQUICO E ASSOCIATIVO.	104
FIGURA 11: RELACIONAMENTO DE EQUIVALÊNCIA ENTRE TERMINOLOGIAS PREFERÍVEIS E NÃO PREFERÍVEIS.....	105
FIGURA 12: SEGMENTO DE FUNÇÕES E ATIVIDADES DE UM MINISTÉRIO PÚBLICO.....	108
FIGURA 13: RELACIONAMENTOS TERMINOLÓGICOS EM UM TESAURO DE FUNÇÃO.....	109
FIGURA 14: FRAGMENTO DE PLANO DE CLASSIFICAÇÃO.....	112
FIGURA 15: EXEMPLIFICAÇÃO DE SUBFUNÇÃO EM TESAUROS.....	116
FIGURA 16: REPRESENTAÇÃO TERMINOLÓGICA EM SKOS.....	117
FIGURA 17: REPRESENTAÇÃO DE FRAGMENTO DE TESAURO DE FUNÇÃO EM XML, RDF E SKOS.	118
FIGURA 18: CADEIA DE INTERPRETAÇÃO DO NÍVEL FÍSICO AO CONCEITUAL.....	126
FIGURA 19: MODELO CONCEITUAL OAIS.	134
FIGURA 20: XENA - TELA DE NORMALIZAÇÃO DE DOCUMENTOS.	137
FIGURA 21: XENA TELA DE CONCLUSÃO DA NORMALIZAÇÃO.	138
FIGURA 22: VISUALIZADOR XENA MODO VISUALIZAÇÃO DE DOCUMENTOS.	139
FIGURA 23: VISUALIZADOR XENA DA ÁRVORE DE METADADOS.....	140
FIGURA 24: VISUALIZADOR XENA DO XML BRUTO.....	141
FIGURA 25: ARQUIVO XENA NORMALIZADO.	142
FIGURA 26: ESQUEMA DE DADOS EM DBML.....	146
FIGURA 27: EXEMPLO EM DBML.	146

LISTA DE QUADROS

QUADRO 01: METADADOS DE IDENTIFICAÇÃO OU REGISTRO.	38
QUADRO 02: METADADOS DE CONTEXTO.	38
QUADRO 03: METADADOS DE CONTEÚDO.	39
QUADRO 04: METADADOS DE TERMOS PARA CONDIÇÃO E USO.	39
QUADRO 06: METADADOS DE CONTEÚDO.	40
QUADRO 07: METADADOS DE HISTÓRICO DE PRESERVAÇÃO.	41
QUADRO 08: METADADOS DE HISTÓRICO DE USO.	41
QUADRO 09: PROPRIEDADES DA RDFS.	62
QUADRO 10: ALGUNS RECURSOS SKOS.	67
QUADRO 11: PROCEDIMENTOS PARA A CRIAÇÃO DE UM TESAURO DE FUNÇÕES.	107
QUADRO 12: ELEMENTOS DO TESAURO.	110
QUADRO 13: TESAUROS E SKOS.	115

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Problema.....	14
1.2	Objetivos	16
1.3	Justificativa.....	17
1.4	Metodologia.....	18
2	ARQUIVOLOGIA: CONCEITOS E DEFINIÇÕES DE UMA ÁREA DA CIÊNCIA DA INFORMÇÃO	20
2.1	A Arquivologia e a Ciência da Informação.....	20
2.2	O arquivo como sistema de informação	21
2.3	A informação orgânica e os documentos de arquivo	22
2.4	Conteúdo, Contexto e Estrutura documentais no âmbito arquivístico	23
2.5	Suporte analógico e digital: forma de registro da informação orgânica.	30
2.6	Os princípios arquivísticos.....	31
2.7	Considerações	34
3	OS METADADOS ARQUIVÍSTICOS E AS LINGUAGENS DE MARCAÇÃO: ALGUMAS CONSIDERAÇÕES.....	35
3.1	Os Metadados Arquivísticos: Instrumentos de representação do conteúdo, contexto e estrutura documental	35
3.2	As linguagens de marcação enquanto elementos de estruturação de metadados	42
3.3	Considerações	53
4	AS LINGUAGENS DA WEB SEMÂNTICA NO CONTEXTO DA INFORMAÇÃO ARQUIVÍSTICA.....	54
4.1	A internet e a recuperação da informação.....	54
4.2	A Web Semântica.....	54
4.3	Considerações	68
5	O COMPORTAMENTO INFORMACIONAL DOS USUÁRIOS DE ARQUIVOS CORRENTES E INTERMEDIÁRIOS	70
5.1	O Comportamento Informacional do usuário de arquivo na fase corrente.....	70
5.2	O gerenciamento estratégico e a relevância da informação no ambiente corporativo	78
5.3	O fluxo de informação na fase corrente e intermediária.....	81
5.4	Considerações	83
6	A GESTÃO DE DOCUMENTOS E DA INFORMAÇÃO ORGÂNICA.....	85
6.1	Gestão Documental: Definição e Elementos	85
6.2	O programa de Gestão Documental.....	89
6.3	Ferramentas da classificação: Instrumentos para a gestão estratégica da informação.	93
6.4	Considerações	95
7	AS CONTRIBUIÇÕES DAS LINGUAGENS DE MARCAÇÃO PARA A GESTÃO DOCUMENTAL: REPRESENTAÇÃO DE PLANOS DE CLASSIFICAÇÃO E TESAuros DE FUNÇÃO 97	
7.1	A recuperação da informação na fase corrente da guarda documental	97
7.2	O controle de vocabulário na gestão documental.....	98
7.3	A linguagem documentária enquanto elemento de interação informacional.....	99
7.4	O plano de classificação documental enquanto uma ferramenta de vocabulário controlado a serviço da recuperação da informação orgânica.....	101
7.5	Os tesauros e suas contribuições para a recuperação da informação orgânica.....	103
7.6	A interação das linguagens de marcação com planos de classificação e os tesauros de função: ferramentas de apoio à gestão da informação arquivística	112
7.7	Considerações	121
8	LINGUAGENS DE MARCAÇÃO: UMA ALTERNATIVA PARA A PRESERVAÇÃO DE DOCUMENTOS DIGITAIS	123
8.1	O advento das tecnologias da informação e suas implicações na produção documental	123
8.2	A preservação de registros digitais: desafios, princípios e estratégias abordadas pela arquivística	128

8.3	As contribuições das linguagens de marcação para a preservação de documentos	
	arquivísticos digitais	135
8.4	Considerações	147
	CONSIDERAÇÕES FINAIS	148
	REFERÊNCIAS	150

1 INTRODUÇÃO

A Ciência da Informação (CI) é caracterizada como a área que analisa as propriedades, o comportamento e fluxo de informações, bem como seu processamento, a fim de aperfeiçoar o seu uso e acessibilidade.

Um dos enfoques da CI é a informação orgânica, ou arquivística, a qual é produzida por um indivíduo ou uma instituição no exercício de suas funções e atividades administrativas. Essa informação encontra-se registrada em documentos de arquivo, que passam por um ciclo de vida e que correspondem a três fases integradas: corrente, intermediária e permanente. (RUSSEAU; COUTURE, 1998, p. 111-126)

A Tecnologia da Informação (TI) revolucionou os modos de produção documental. Documentos, que antes eram registrados em suportes analógicos, tais como papel e rolos de filme, estão sendo codificados em bits e armazenados em mídias digitais. O volume excessivo de documentos digitais traz novos desafios para a CI, no que tange ao acesso e a preservação da informação orgânica em todas as fases do ciclo de vida documental.

1.1 Problema

A acessibilidade e a preservação dos documentos digitais são de fundamental importância para garantir a eficácia dos sistemas de arquivo. Tais benefícios deverão ser perseguidos nos procedimentos que visem à gestão de documentos e a guarda permanente.

O acesso à informação na fase corrente é fundamental para procedimentos de tomada de decisão no ambiente corporativo. É necessário que os sistemas informatizados de gestão documental disponham de mecanismos que possam interpretar as informações registradas nos documentos arquivísticos, a fim de produzirem dados abrangentes e precisos que visem beneficiar o administrador.

Há outra questão que apresenta ainda maior relevância: a acessibilidade dos documentos digitais por médio e longo prazo.

A evolução constante dos equipamentos e do *software* ocasiona a obsolescência dos formatos de arquivos digitais proprietários, os quais são protegidos por leis de propriedade intelectual. Tão logo o uso de tais equipamentos perde evidência, os documentos a eles vinculados tornam-se órfãos. O desaparecimento dos instrumentos necessários para a interpretação das sequências binárias faz com que os registros arquivísticos digitais se tornem completamente inacessíveis. (NATIONAL ARCHIVES OF AUSTRALIA, 2002, 2006).

Os pesquisadores das áreas de CI e TI promovem importantes discussões visando solucionar os problemas supracitados. Eles percebem que as linguagens de marcação têm apontado contribuições relevantes para os desafios trazidos pelos documentos digitais por serem, dentre outras características, independentes de sistema específico de *hardware* e *software* e escritas em formato de texto ASCII.

As linguagens de marcação identificam os objetos de um texto por meio do uso de marcas, denominadas em inglês *tags*. Os principais tipos de linguagens de marcação serão apresentados a seguir.

A *Standard Generalized Markup Language* (SGML) foi reconhecida pela *International Organization for Standardization* (ISO) no ano de 1986. Dela surgiram a *Hypertext Markup Language* (HTML) e a *Extensible Markup Language* (XML)

A linguagem HTML, desenvolvida em 1989, é voltada para a apresentação de textos na internet. Devido ao fato de trazer *tags* predefinidas, apresenta poucas contribuições para a gestão arquivística de documentos. (W3C, 1998).

A *World Wide Web Consortium* (W3C) lançou no ano de 1997 a XML. Essa linguagem não apresenta predefinição de marcas, permitindo ao desenvolvedor rotular os dados da forma que julgar necessário. Nesse modelo, as informações são compreendidas também pelas máquinas e, portanto, compartilhadas, processadas e recuperadas em diferentes sistemas computacionais. (W3C, 2003).

O uso da XML é recomendado no desenvolvimento de ferramentas de *software* que visem à gestão e preservação da informação arquivística digital. Como exemplo poderia ser mencionado a *web* semântica na gestão documental e a normalização de formatos de arquivo aberto para a preservação do acesso. Uma breve explanação acerca de cada uma dessas abordagens será exposta a seguir.

A *Web Semântica* tem por objetivo tornar os dados disponíveis na *web* compreensíveis às máquinas. Dessa forma, as informações registradas nas páginas da internet poderão ser processadas e recuperadas de forma precisa. Para atingir esse fim, são utilizadas linguagens derivadas da XML, tais como: *Resource Description Framework Schema* (RDFS) e *Web Ontology Language* (OWL). (W3C, 2001).

O uso integrado dessas linguagens permite a construção de ontologias que descrevam as relações hierárquicas e os aspectos semânticos existentes entre termos de um tesouro ou determinada taxonomia. A transposição de um tesouro de funções para as referidas linguagens de marcação possibilitarão aos sistemas de gestão de documentos eletrônicos a integração entre as informações orgânicas dos documentos de arquivos com as atividades que os gerou.

Verifica-se também a aplicabilidade do uso da linguagem XML no que tange às estratégias desenvolvidas para contornar a obsolescência de formatos de arquivos digitais e assegurar a acessibilidade dos documentos digitais por longo prazo. O Arquivo Nacional Australiano tem desenvolvido, desde 2005, um modelo padronizado de formatos de arquivo Digital. O *XML Electronic Normalising of Archives* (XENA) converte formatos, tais como JPG, PDF, MP3 etc., em documentos XML em formato arquivístico de dados normalizado. Desta forma, é possível reconstituir a apresentação do formato original sem a necessidade de efetuar migrações constantes. (NATIONAL ARCHIVES, 2002, p. 17-21)

Mediante o exposto, é necessário definir que o principal objeto de pesquisa deste trabalho é a linguagem de marcação XML e suas contribuições trazidas à gestão e preservação da informação orgânica registrada em documentos arquivísticos. É importante ressaltar que o foco direcionar-se-á à gestão digital de documentos registrados em suporte analógico e digital, sendo a preservação voltada exclusivamente para documentos arquivísticos digitais. Buscou a pesquisa responder aos seguintes questionamentos:

- Quais seriam as possíveis contribuições que a linguagem de marcação XML e suas derivações poderiam trazer à arquivística?
- Quais são as possíveis relações existentes entre as linguagens de marcação utilizadas na *web* semântica e a gestão de documentos digitais e informações arquivísticas?
- Como se dá a padronização de formatos de arquivos digitais na linguagem XML?

1.2 Objetivos

1.2.1 Geral:

Analisar as contribuições que as linguagens de marcação possam trazer à informação arquivística, no que diz respeito à gestão e a preservação de documentos digitais.

1.2.2 Específicos:

- Analisar a padronização de formatos de arquivos digitais em XML, provida por *software* de preservação, bem como os reflexos trazidos à obsolescência de formatos de arquivos proprietários.

- Investigar as contribuições que a web semântica, por meio de linguagens de marcação, poderá trazer à representação de tesouros de função visando à recuperação de informações em sistemas de gestão eletrônica de documentos.

1.3 Justificativa

O conteúdo de informação existente no meio digital cresce exponencialmente. Estima-se que mais de 90% de toda a informação gerada no mundo esteja registrada no meio digital. (THOMAZ, 2004, p. 30).

Weinberger (2007) comenta que os documentos englobam uma grande variedade de gêneros documentais que antes apresentavam suportes distintos. Esta mistura de documentos na *web* inviabiliza a recuperação da informação. Os documentos de arquivos digitais apresentam desafios semelhantes aos recursos da *web*.

O crescimento do volume de informações orgânicas registradas no meio digital exige que as áreas de TI e CI unam-se na busca de meios que possam otimizar a preservação e o acesso à informação nos três estágios do ciclo de vida documental.

Ramalho, Vidotti e Fujita (2007) comentam que “atualmente, verifica-se uma tendência de aproximação entre as áreas de Ciência da Informação e Ciência da Computação, principalmente no que tange ao desenvolvimento de novos instrumentos de representação e recuperação de recursos informacionais.” Os autores ressaltam que, no entanto, as duas áreas abordam a informação de maneira distinta. Enquanto a CI trabalha a informação como um elemento dotado de significado, podendo ser transmitido a um ser consciente, a TI trabalha a informação sob o enfoque da Teoria Matemática “para a transmissão de mensagens e troca de sinais, a qual não se preocupa com a semântica dos dados, porém adéqua a construção de sistemas operacionais, onde a informação possa ser quantificada, processada e transmitida pelas máquinas”.

Neste contexto, a área da Ciência da Informação traz contribuições de grande relevância para a TI, uma vez que trabalha a informação sob um enfoque semântico. Surge a necessidade de se elaborar pesquisas que estabeleçam um diálogo entre essas duas áreas, a fim de promover um intercâmbio de benefícios.

Este projeto de pesquisa tem por objetivo relacionar as contribuições das linguagens de marcação trazidas às necessidades apresentadas pela literatura científica em torno da gestão e da preservação da informação arquivística desde os procedimentos de uso corrente à guarda permanente. Apresenta, portanto, relevância significativa para linhas de pesquisa que

trabalhem nos segmentos de informação e tecnologia visando à gestão, o acesso e a preservação da informação orgânica registrada nos documentos de arquivo digitais.

1.4 Metodologia

Para executar os objetivos desse trabalho optou-se pela pesquisa exploratória. Segundo Sampierri et al (1994, tradução nossa)

os estudos exploratórios se efetuam, normalmente, quando o objetivo é examinar um tema ou problema de investigação pouco estudado ou que não tenha sido abordado antes; ou seja, quando a revisão de literatura revelou que há somente guias não investigadas e ideias relacionadas vagamente com o problema de estudo.

A Pesquisa Exploratória foi escolhida, portanto, devido a possibilidade de promoção da investigação, bem como a articulação dos elementos dos objetivos específicos que se encontram dispersos na literatura científica.

O método a ser utilizado será a Análise de Conteúdo em sua função heurística, a qual “enriquece a tentativa exploratória, aumenta a propensão à descoberta”. (VALENTIM, 2005, p. 121).

A execução desse projeto se realizou por meio da coleta de dados provenientes da pesquisa bibliográfica, que correspondeu a um estudo exaustivo de publicações científicas que verssem a respeito da temática estudada. Tal pesquisa tem por finalidade a conscientização do pesquisador acerca das abordagens realizadas em torno do assunto investigado, quer sejam estas fontes transcritas ou gravadas. (LAKATOS; MARCONI, 1991, p. 183).

Os autores ainda afirmam que a pesquisa permite a resolução e a exploração de novas áreas onde os problemas não se cristalizaram, uma vez que “[...] a pesquisa bibliográfica não é mera repetição do que já foi dito ou escrito sobre certo assunto, mas propicia o exame de um tema sob novo enfoque ou abordagem, chegando a conclusões inovadoras”.

Gil (2008, p. 59-86) afirma que a execução de uma pesquisa bibliográfica é um processo que envolve uma série de etapas, sendo essas:

- Elaboração do plano provisório de assunto;
- Busca de fontes;
- Leitura do material;
- Fichamento;
- Organização lógica do assunto; e
- Redação do texto.

As etapas supracitadas são os procedimentos metodológicos adotados na coleta de dados da pesquisa que foram seguidas nesse projeto.

A fim de alcançar os objetivos propostos, os capítulos receberam a seguinte distribuição temática:

Para estabelecer definições da área de Ciência da Informação, Arquivologia, Metadados, Linguagens de Marcação e Web Semântica, desenvolveu-se os seguintes capítulos:

- Capítulo 02 – Arquivologia: conceitos e definições de uma área da Ciência da Informação;
- Capítulo 03 – Os metadados arquivísticos e as linguagens de marcação: Algumas considerações;
- Capítulo 04 – As linguagens da web semântica no contexto da informação arquivística.

Uma vez estabelecidas as definições, segue-se para as duas primeiras fases do ciclo de vida documental (corrente e intermediário), em que serão abordadas as contribuições das linguagens de marcação para a gestão de documentos e da informação orgânica, bem como da preservação de documentos digitais.

- Capítulo 05 – O comportamento informacional dos usuários de arquivos correntes e intermediários;
- Capítulo 06 – A gestão de documentos e da Informação Orgânica;
- Capítulo 07 – As contribuições das linguagens de marcação para a gestão documental: representação de planos de classificação e tesauros de função;
- Capítulo 08 – Linguagens de Marcação: uma alternativa para a preservação de documentos digitais;
- Considerações finais.

Uma vez definido o processo metodológico, o presente trabalho segue para o marco teórico.

2 ARQUIVOLOGIA: CONCEITOS E DEFINIÇÕES DE UMA ÁREA DA CIÊNCIA DA INFORMAÇÃO

O presente capítulo apresenta conceitos e definições essenciais para a compreensão da Arquivologia na Ciência da Informação. Tal abordagem é necessária para uma melhor compreensão da temática desenvolvida nos capítulos posteriores. Apresentar-se-á a inserção da arquivística na CI, ressaltando o seu papel no tratamento da informação orgânica, e discorrer-se-á sobre os elementos constituintes do acervo arquivístico, ou seja, quer seja o documento e suas partes, bem como dos princípios norteadores da área.

Desta forma o leitor poderá obter maior interação com as abordagens a serem explanadas posteriormente, as quais exigirão conhecimento prévio dos fundamentos teóricos da arquivologia.

2.1 A Arquivologia e a Ciência da Informação.

O Arquivo Nacional (2005, p. 37) define a arquivologia como “disciplina que estuda as funções do arquivo e os princípios e técnicas a serem observados na produção, organização, guarda, preservação e utilização de arquivos. Também chamada arquivística”. Para esta definição, o arquivo deverá ser compreendido como “instituição ou serviço que tem por finalidade a custódia, o processamento técnico, a conservação e o acesso a documentos”. (ARQUIVO NACIONAL, 2005, p.27)

A arquivologia ou arquivística encontra-se inserida na Ciência da Informação (CI) que, segundo Borko (1968), investiga o comportamento, as propriedades, bem como as forças que governam os fluxos da informação, visando promover a usabilidade e acessibilidade da mesma. Permite mapear e contextualizar elementos informacionais visando torná-los acessíveis a usuários potenciais.

Lecoadic (2004, p. 25) define, por objeto da CI, o estudo das propriedades gerais da informação (natureza, gênese, efeitos) e a análise de seus processos de construção, comunicação e uso. Esta informação caracteriza-se como um conhecimento inscrito em forma escrita, oral ou audiovisual em um suporte.

A informação registrada em um suporte pode representar os eventos que a produziu, dotando-os de evidência. Por este motivo, a informação registrada será a principal abordagem neste trabalho, uma vez que os sistemas informacionais materializam a informação a fim de disponibilizá-la aos usuários e somente este tipo de informação conseguem processar (BUCKLAND, 1991).

A informação registrada é, portanto, o único tipo informacional processado, armazenado e recuperado por um sistema de informação arquivística, e que será delineada a seguir.

2.2 O arquivo como sistema de informação

Por sistemas compreende-se ser um conjunto de elementos que interagem entre si para cumprir um determinado objetivo. O modelo de sistema é composto por três etapas, a saber: entrada, processamento e saída.

Beal (2008, p. 16) afirma que em um sistema de informação a entrada corresponde a dados capturados; o processamento envolve a conversão ou transformação dos dados; e a saída envolve a produção de informações úteis, muitas vezes na forma de documentos.

Silva et al (2002 p. 25-27) define informação como um conjunto de dados compreensíveis e registrados em um suporte físico e passível de comunicação. A informação é uma propriedade de dados produzidos por um processo, devendo ser vinculada tanto aos meios operacionais quanto à interação sistêmica inerente ao processo informacional.

Por sistema de informação o autor compreende ser um conjunto organizado de procedimentos para coleta, processamento, armazenamento e recuperação da informação visando à satisfação de necessidades informacionais.

O Arquivo Nacional (2009, p. 10) complementa afirmando que um sistema de informações pode ser definido como:

Conjunto organizado de políticas, procedimentos, pessoas, equipamentos e programas computacionais que produzem, processam, armazenam e provêm acesso à informação proveniente de fontes internas e externas para apoiar o desempenho das atividades de um órgão ou entidade.

A finalidade dos sistemas de informação é permitir e facilitar a tomada de decisão, proporcionando o suporte aos procedimentos operacionais. Contrariamente ao que rege o senso comum, um sistema de informação nem sempre está diretamente vinculado à Tecnologia da Informação (TI), mas poderá ser composto de procedimentos, manuais ou automatizados, que promovam o processamento da informação.

Nesse contexto, Silva et al (2002, p. 211) define arquivística como uma ciência da informação social que estuda o arquivo enquanto um sistema (semi-)fechado. A este respeito, o autor comenta que muitos outros autores abordam modalidades dos sistemas como aberto ou fechado. Um sistema aberto permite a interação com o ambiente externo, enquanto um

sistema fechado é endógeno, ou seja, os seus procedimentos não interagem com o ambiente externo.

Silva et al (2002) defende o conceito de arquivo como um sistema (semi-) fechado porque, em um primeiro momento, a informação registrada nos arquivos será de interesse exclusivo da entidade produtora e, posteriormente, será do interesse de toda a sociedade, na fase de recolhimento ao arquivo permanente.

Neste contexto, é possível observar que o arquivo é composto por um conjunto de registros documentais que oferecem suporte aos processos envolvidos em um sistema de informação em que a produção da informação pode ser vinculada ao procedimento de entrada; o armazenamento vincula-se ao processamento; e a busca com a recuperação da informação à saída.

O sistema arquivístico de informação é complexo e, portanto, exige tratamento informacional integrado desde a produção da informação pelo usuário primário, até a recuperação pelo usuário secundário, quando o acervo obtiver valor de guarda permanente.

Uma vez definido o conceito de arquivo como sistema de informação, é necessário trabalhar os conceitos relacionados à informação orgânica e aos documentos de arquivo.

2.3 A informação orgânica e os documentos de arquivo

Antes de apresentar a definição de informação orgânica, é necessário apresentar um esclarecimento terminológico. A presente dissertação traz o termo “informação arquivística” no título. Para este trabalho, “informação arquivística” e “informação orgânica” são termos sinônimos que poderão ter o uso alternado em capítulos posteriores. Uma vez feito este esclarecimento, apresenta-se a definição.

A informação registrada em arquivos é denominada “informação orgânica”, pois é produzida e acumulada por um indivíduo ou instituição no exercício de suas atribuições visando cumprir finalidades administrativas, fiscais ou legais. (ROUSSEAU; COUTURE, 1998 p. 65).

A expressão “orgânica” se dá devido à característica de produção informacional que ocorre atrelada aos trâmites existentes em procedimentos administrativos voltados ao cumprimento de atribuições peculiares de uma atividade inserida em uma função que, por sua vez, estará direcionada a uma missão institucional. Por esta razão, a informação orgânica encontra-se registrada em documentos.

O conceito de documento e documentação apresenta uma abordagem ampla. Briet (2006) comenta que o termo “documento” pode ser compreendido como qualquer sinal

concreto ou indexado, simbolicamente preservado ou registrado, com a finalidade de representar, reconstituir ou provar um fenômeno físico ou intelectual. A autora exemplifica afirmando que um animal vivente em seu habitat natural não é um documento, mas esse mesmo animal vivendo em um zoológico poderia ser considerado um documento.

Constata-se que diversos objetos podem ser considerados como documentos em um determinado contexto. A arquivologia, por sua vez, trabalha com foco direcionado aos documentos de arquivo que são definidos como documento produzido ou recebido por uma pessoa física ou jurídica, no exercício de suas atribuições. (ARQUIVO NACIONAL, 2005; DURANTI, 1996).

Por conseguinte, a especificidade existente no conceito de documentos arquivísticos remonta a produção orgânica deste instrumento informacional. Esta modalidade de documento surge naturalmente, vinculada à necessidade de registro fidedigno de um procedimento funcional, ou seja, de uma atividade. O documento de arquivo é composto, portanto, de um inter-relacionamento de conteúdo e estrutura documental visando à representação de um determinado contexto de produção.

Observa-se que a informação orgânica, materializada em suportes documentais, produz o documento de arquivo que, por sua vez, a reveste de tangibilidade. Isso permite o tratamento e a difusão dos recursos informacionais aos usuários aferindo-se, ainda, a autenticidade e integridade dos registros arquivísticos recuperados.

Uma vez definido o conceito de documento arquivístico, passa-se a explicar a composição que o sustenta no que concerne o conteúdo, o contexto e a estrutura.

2.4 Conteúdo, Contexto e Estrutura documentais no âmbito arquivístico

A norma ISO 15489-1 de 2001 acrescenta que os documentos de arquivo são acometidos de conteúdo, contexto e estrutura suficiente para prover evidência para as funções e atividades que os gera. (BANTIN, 2008, p. 27).

O conteúdo consiste na mensagem primária do documento podendo ser em palavras, símbolos, imagem ou áudio. O contexto, por sua vez, define a circunstância organizacional, funcional e operacional envolvendo a produção, armazenamento, uso e a gestão dos documentos no decorrer do tempo. Já a estrutura define as características físicas do documento e a composição do conteúdo.

No que tange ao documento de arquivo, a sua estrutura poderá ser vinculada ao gênero documental, o contexto ao ciclo de vida documental e o conteúdo é determinado,

muitas vezes, pela tipologia documental. Estes elementos serão explanados nos parágrafos seguintes.

2.4.1 O gênero documental e a estrutura do registro arquivístico

Por gênero documental compreende-se:

Reunião de espécies documentais que se assemelham por seus caracteres essenciais, particularmente o suporte e o formato, e que exigem processamento técnico específico e, por vezes, mediação técnica para acesso, como documentos audiovisuais, documentos cartográficos, documentos eletrônicos, documentos filmográficos, documentos iconográficos, documentos micrográficos, documentos textuais. (ARQUIVO NACIONAL, 2005, p. 99).

Para uma melhor compreensão da definição acima, será realizada uma análise dos termos considerados mais relevantes: suporte e formato. A definição de espécie documental será abordada na sequência, quando for definido a natureza do conteúdo documental. Por ora resta dizer que a espécie reúne conjuntos de documentos por sua semelhança. São exemplos de espécie: certidão, memorando, atestado, boletim etc.

O suporte indica o material que recebe os registros informacionais como o papel, o pergaminho, o meio óptico e magnético, o filme etc.

O formato, por sua vez, é o conjunto das características físicas de apresentação das técnicas de registro e da estrutura da informação e o conteúdo de um documento (ARQUIVO NACIONAL, 2005). Desta forma, o formato apresenta a configuração do suporte e das técnicas de registro nele contidas podendo ser manuscrito, datilografado, desenho etc. São exemplos de formato: folha, caderno, mapa, fotografia etc.

O Gênero documental indica, portanto, a configuração da estrutura física do meio em que se configurará o registro da informação pela união do tipo de formato e do suporte documental. São exemplos de gênero os documentos:

- Audiovisuais - contem imagens fixas ou em movimento, acompanhadas de registro sonoro;
- Bibliográficos - impressos, tais como livros e periódicos;
- Cartográficos - documentos que registram representações gráficas da superfície terrestre, corpos celestes e desenhos técnicos, tais como mapas e plantas;
- Iconográficos - contem imagens fixas, impressas, desenhadas ou fotografadas, tais como fotografias e gravuras;
- Micrográficos - relativo aos documentos registrados em microformas, tais como o microfilme e o cartão janela;

- Textuais - são os documentos manuscritos, datilografados ou impressos, como, por exemplo, atas de reunião, cartas; e
- Eletrônicos - são os registros acessíveis somente por meio eletrônico, tais como cartão perfurado, discos magnéticos e óticos e documentos digitais. (ARQUIVO NACIONAL, 2005).

Embora o Arquivo Nacional mencione documentos digitais dentro do gênero eletrônico, é interessante observar que os documentos digitais comportam registros do gênero cartográfico, iconográfico, audiovisual, bibliográfico, filmográfico e textuais. Os documentos do gênero micrográfico também registram os gêneros bibliográficos, textuais, cartográficos e iconográficos.

Os gêneros documentais relacionados ao conteúdo podem se resumir, portanto, em: audiovisual, cartográfico, iconográfico e textual. Esses elementos apresentam um conjunto de signos que identificam a estrutura existente em documentos de arquivo. Uma vez analisada a estrutura, passa-se a verificar o relacionamento existente com o contexto de produção documental representado no ciclo de vida documental.

2.4.2 O Ciclo de vida e o contexto de produção documental

O ciclo de vida documental embasa-se na valoração documental. Schellenberg (2006, p. 179-188) divide a valoração documental em dois estágios: o valor primário e o valor secundário.

Na vigência do valor primário, os documentos são utilizados para atender as finalidades para as quais foram produzidos, no que tange o cumprimento de obrigações administrativas, fiscais e legais. Neste contexto informacional, conforme já mencionado, o usuário participa ativamente no processo de produção documental, sendo suas ações registradas no ambiente arquivístico.

O valor primário, portanto, está relacionado às fases:

- Corrente - momento em que ocorre a produção documental atrelada aos procedimentos de tomada de decisão por parte do administrador; e
- Intermediária - momento em que a documentação aguarda a prescrição do valor primário.

O contexto de produção documental é delineado durante o valor primário. Uma vez prescrito o valor primário, serão selecionados os documentos que adquiriram valor secundário para que possam ser mantidos em guarda permanente.

Na vigência do valor secundário, os documentos são utilizados em desígnios diferentes daqueles para os quais foram produzidos, o que diz respeito à comprovação de direitos adquiridos, desenvolvimento de pesquisas científicas, ou ainda aquisição de conhecimento cultural. O valor secundário relaciona-se, portanto, à fase de guarda permanente.

A interação dos usuários com o sistema de informação arquivística altera-se no decorrer do ciclo de vida documental e isso modifica a finalidade da recuperação informacional no decorrer do tempo. A característica de sistema de informação (semi-) fechado, apresentada por Silva et al (2002), indica a atuação de dois grupos de usuários com necessidades informacionais distintas.

Assim, o usuário do grupo primário recuperará a informação para atender demandas específicas produzidas em suas atribuições funcionais e produzirá novas informações que serão incorporadas no acervo.

Já o usuário do grupo secundário interagirá, posteriormente, com esse sistema a compreensão do acervo documental e de todo o contexto envolto em sua produção, o que corresponde à atuação do usuário primário durante a execução de suas atividades e os reflexos por elas produzidos.

O usuário primário, portanto, produz, armazena e busca informações para finalidades específicas, enquanto o usuário secundário recupera informações para finalidades diversificadas.

A abordagem da atuação dos dois grupos de usuários é muito complexa para ser abordada em uma dissertação. Por este motivo, o presente trabalho terá seu foco voltado às necessidades informacionais dos usuários pertencentes ao grupo primário, momento em que a intervenção do profissional da informação traz maior efeito para a atuação posterior do grupo secundário ao sistema de informação arquivística.

A atuação do profissional da Ciência da Informação é fundamental já na vigência do valor primário, pois permite maior eficácia na seleção dos registros informacionais que serão alvo da guarda permanente, o que garante maior riqueza no processo descritivo dos registros documentais e o correto tratamento ao acervo.

Uma vez delineado o ciclo documental dentro do contexto de produção, será explanado o conteúdo dos documentos de arquivo por meio das tipologias documentais.

2.4.3 O conteúdo por meio das tipologias documentais

Para uma explanação que permita ampla compreensão do conteúdo dos documentos de arquivo é necessária a exposição dos conceitos de tipologia, espécie e gênese documental, bem como da disciplina de diplomática, uma vez que estes elementos delimitam as diretrizes que norteiam a produção documental.

O Arquivo Nacional (2005, p. 70) define a diplomática como “disciplina que tem por objeto o estudo da estrutura formal e da autenticidade dos documentos”. Para Belloto (2008) a diplomática ocupa-se da estrutura formal dos atos escritos, sejam eles de origem notarial ou governamental. Tem por objeto a estrutura formal do documento, o qual deverá conter a mesma construção semântica do discurso para uma mesma problemática jurídica.

O documento diplomático configura-se no registro legitimado de atos jurídicos e administrativos, e apresenta as seguintes características:

- Suporte - veículo sustentador transmissor da mensagem;
- Texto - conteúdo expresso com linguagem, escrita, vocabulário e fórmulas peculiares em concordância com o tema e com a época;
- Finalidade e ideologia - estruturação e formalidades que lhe sirvam de garantia.

Neste contexto, a espécie documental destaca-se como o centro de estudos da diplomática moderna que se concentra na gênese, na constituição interna, na transmissão e na relação dos documentos entre o seu criador e seu próprio conteúdo, com a finalidade de identificar, avaliar e demonstrar a sua verdadeira natureza. Sua estrutura é estabelecida de forma a permitir certificar a autenticidade do documento arquivístico. Por este motivo, a espécie sofrerá uma construção semântica obrigatória na qual haverá:

- Dados fixos que existirão em todos os documentos integrantes da mesma espécie; e
- Dados variáveis que serão próprios de cada documento indivíduo dentro de uma mesma espécie. (BELLOTO, 2008).

Compreende-se por espécie documental a “divisão de gênero documental que reúne tipos documentais por seu formato. São exemplos de espécies documentais: ata, carta, decreto, disco, filme, folheto, fotografia, memorando, ofício, planta, relatório.” (ARQUIVO NACIONAL, 2005, p. 85).

No contexto da diplomática, a espécie documental pode ser compreendida como uma forma textual “padronizada” para o registro de ações administrativas distintas (BELLOTO, 2008).

Para uma maior elucidação, é importante explicar que utiliza-se uma certidão, por exemplo, como uma estrutura formalizada que dê suporte a certificação de um assentamento de um determinado registro em livro. O ofício, por sua vez, traz um conteúdo informacional que formaliza a correspondência oficial entre autoridades. Uma portaria traz o conteúdo de um ato normativo que regulamenta procedimentos a serem adotados em uma unidade administrativa de uma instituição pública.

Cada uma das espécies documentais mencionadas permite a configuração do conteúdo de um documento em conformidade com a finalidade da produção do registro documental.

É possível, portanto, compreender a espécie documental como um elemento da diplomática que define uma gama de diretrizes para a configuração de um grupo de tipos documentais subalternos, visando garantir a autenticidade dos atos neles registrados (BELLOTO, 2008).

Cada espécie reúne um conjunto de tipos documentais que herdarão a estrutura vigente e permitirão o registro de dados variáveis para garantir a especificidade do tipo documental. A tipologia documental, portanto, permite a subdivisão da espécie documental a uma finalidade específica. Essas subdivisões compreendem os tipos documentais.

O Arquivo Nacional (2005, p. 163) define tipo documental como:

Divisão de espécie documental que reúne documentos por suas características comuns no que diz respeito à fórmula diplomática, natureza de conteúdo ou técnica do registro. São exemplos de tipos documentais cartas precatórias, cartas régias, cartas-patentes, decretos sem número, decretos-leis, decretos legislativos[...]

O tipo documental traz maior especificidade à atividade que o produziu. Pode ser compreendido como a fusão da espécie documental com a função produtora, o que permite maior representatividade da organicidade do acervo (LOPEZ, 1999 p. 71). Isso se dá porque o formalismo existente nos procedimentos executados em uma função será refletido no conteúdo formalizado das espécies documentais vinculadas às atividades subalternas e, desse vínculo, surge o tipo documental.

Defende Belloto (2008) que a tipologia é a ampliação da diplomática em direção à gênese documental, ou seja, aos procedimentos que geraram os registros documentais.

Enquanto a diplomática busca trabalhar a configuração interna do documento, bem como realizar estudo jurídico dos seus elementos e caracteres para atingir a autenticidade, a tipologia, além disso, aborda o documento enquanto componente de um contexto orgânico, advindo da junção de documentos referentes à mesma atividade, tornando um conjunto homogêneo de atos expresso em um conjunto homogêneo de documentos sendo, portando,

uma configuração assumida pela espécie documental em conformidade com a atividade que a gerou. Visa à lógica orgânica dos conjuntos documentais. (BELLOTO, 2008).

Para efeito de exemplificação das tipologias documentais, analisa-se a espécie certidão, explanada anteriormente como a comprovação da existência de um registro. Para a espécie certidão temos os seguintes tipos: certidão de nascimento, certidão de casamento, certidão de óbito, certidão negativa de débitos e tributos etc. Cada uma dessas tipologias registra a especificidade de atividades distintas, com finalidades específicas. Observa-se a existência de uma finalidade específica, que é a certificação, para atividades distintas, o que configura a existência de dados fixos, a finalidade de certificação, bem como de dados variáveis, o direcionamento para atividades distintas.

O conteúdo das espécies e dos tipos documentais é construído, nos documentos textuais, com elementos de texto do discurso diplomático, composto da união de três partes distintas, a saber: protocolo inicial, texto e protocolo final. Cada um desses itens contém um conjunto de elementos que serão listados a seguir.

No protocolo inicial encontra-se:

- Invocação (*Invocatio*) - presente nos documentos mais antigos, sendo identificado pela expressão “Em nome de Deus”;
- Titulação (*Titulatio*) - traz o nome da autoridade de quem emana o ato e seus títulos;
- Direção ou endereço (*Inscriptio*) - nomeia a quem se dirige o ato;
- Saudação (*Salutatio*) - finaliza o protocolo inicial.

No texto encontra-se:

- Preâmbulo (*Prologus* ou *Exordim*) - no qual se justifica a criação do ato;
- Notificação (*Notificatio* ou *Promulgatio*) - compreendida na expressão “tenho a honra de comunicar a vós”;
- Exposição (*Narratio*) - explicita as causas do ato, o que o gerou e as necessidades que o tornaram necessário;
- Dispositivo (*Dispositio*) - é a substância do ato em que se determina o que se quer;
- Sanção (*Sanctio* ou *Minatio*) - penalidades no caso de não cumprimento do dispositivo;
- Corroboração ou Cláusulas Finais (*Valoratio* ou *Corroboratio*) - dispõe os meios morais ou materiais que assegurem a execução do dispositivo.

O protocolo final é composto de:

- Subscrição/assintura (*Subscriptio*) - a assinatura do emissor do documento;
- Datação (*Datatio*) - composto por data tópica e cronológica;
- Precação (*Apprecatio*) - composta de assinaturas de testemunhas e de sinais de validação como selos e carimbos a fim de reiterar-se a qualidade do documento.

As espécies documentais apresentam esses elementos em sua totalidade ou menor quantidade. Esta estrutura de conteúdo é, então, repassada ao conjunto de tipos documentais pertencentes à espécie. A tipologia documental vincula o documento arquivístico à atividade e ao ato que o produziu e, portanto, à estrutura do documento ao contexto de produção.

Em relação à diplomática, é necessário esclarecer que seu uso é voltado aos documentos de gênero textual, não sendo possível a sua aplicação em documentos audiovisuais, cartográficos e iconográficos, já que esses têm um conteúdo diferenciado. Por apresentar configuração puramente textual, a diplomática não alcança esses itens, uma vez que a estrutura envolta nessas modalidades de gênero não são passíveis de submissão a uma forma de “escrita” pré-definida.

Apesar de não estarem inclusos na diplomática, os documentos de gênero audiovisual iconográfico e cartográfico são caracterizados como documentos de arquivo e podem ser representados e tratados com métodos específicos à sua natureza. A delineação desta metodologia, no entanto, não será objeto de estudo neste trabalho.

Observa-se que a estrutura, que é o suporte e a forma que se configura o documento, alinha-se ao conteúdo, teor documental e ao contexto de produção na constituição do registro arquivístico, dotando-o de evidência.

São múltiplas as formas que poderão assumir os suportes documentais arquivísticos podendo as mesmas serem subdivididas em suporte analógico e digital. É necessário apresentar uma breve explanação dessas duas modalidades.

2.5 Suporte analógico e digital: forma de registro da informação orgânica.

Uma vez apresentado o suporte como parte da estrutura documental, surge a necessidade de apresentar os tipos de suportes em que os documentos arquivísticos podem ser registrados, tanto no ambiente analógico como no ambiente digital.

O documento de arquivo pode, portanto, ser registrado em suportes que possibilitam o registro de modo analógico e suportes em que os documentos de arquivo são registrados no

modo digital. Os primeiros serão denominados, neste trabalho, suporte analógico e o segundo será denominado suporte digital.

No suporte analógico, os documentos são registrados de modo a representar, como o próprio nome indica, uma analogia do objeto registrado. Os documentos textuais trazem registro em tinta dos caracteres componentes da mensagem. As fotografias em papel são compostas de pigmentos de tinta que fazem uma analogia às cores presentes na lente da câmera no momento da captura.

O suporte analógico permite ainda definir e especificar o gênero documental a ser registrado. O papel poderá receber documentos do gênero textual bibliográfico ou iconográfico, mas não poderá receber documentos micrográficos. Microformas deverão ser registradas em filmes de saís de prata.

Os documentos registrados no suporte digital, por sua vez, trazem em sua essência apenas um código binário, em 0 e 1, que submetido a um procedimento específico de interação de elementos de *software* e *hardware*, representará os sinais analógicos que registram. O suporte digital poderá registrar documentos dos mais diversificados gêneros documentais, conquanto que disponha do procedimento correto de *hardware* e *software* para a reconstrução da forma analógica que está codificada nos bits.

Apresentam uma série de benefícios aos arquivos na sua fase de uso corrente, uma vez que economizam espaço físico e podem ser transmitidos com facilidade.

Ao mesmo tempo em que propõe registrar e representar tudo, o meio digital pode resultar em tarefas dispendiosas para manter os procedimentos de *hardware* e *software* necessários à apresentação. Qualquer falha poderá resultar em perda da informação (NATIONAL ARCHIVES OF AUSTRALIA, 2002).

Por esta razão, os registros digitais trazem grandes desafios à guarda permanente. A preservação dos registros de suporte digital é um tema a ser refletido em capítulos posteriores. A fragilidade é outro ponto negativo, que aumenta o risco de adulteração e deterioração.

Uma vez definido os conceitos de Arquivologia, Ciência da informação, Informação Orgânica e Documento de Arquivo, passa-se a investigar o histórico de evolução dos princípios arquivísticos ao longo dos séculos.

2.6 Os princípios arquivísticos

Uma vez analisado o desenvolvimento da área ao longo do tempo, passemos a estudar os princípios que norteiam a produção da informação orgânica no ambiente de arquivo. São princípios da arquivologia:

- Proveniência: também conhecido como princípio do respeito aos fundos, é um dos princípios fundamentais da arquivologia. Determina que o produtor da documentação seja visto como unidade de referência para a constituição de um fundo arquivístico. Documentos de diferentes fundos não poderão ser misturados, a fim de preservar a individualidade e integridade do acervo.
- Unicidade: os documentos de arquivo, a despeito de sua forma ou tipo, conservam caráter único para a execução de uma função dentro do seu contexto de origem. Isto significa que, em um determinado contexto de produção, contendo caracteres internos e externos com os dados fixos ou variáveis dentro da estrutura da tipologia documental, o documento de arquivo é único para os seus efeitos não havendo nenhum outro com propósitos semelhantes.
- Indivisibilidade: também conhecido como princípio da integridade, caracteriza o fundo arquivístico como um todo indivisível. A informação orgânica é registrada em diferentes tipos documentais que ficam inter-relacionados de modo que, desvinculado do acervo, o documento de arquivo perderá o significado, já que não mais poderá representar o contexto de produção. Por esse motivo, são vedadas as práticas de dispersão, mutilação, alienação, destruição não autorizadas ou o acréscimo indevido. (BELLOTO, 2008; ROUSSEAU, COUTURE, 1998).

Uma vez apresentados os princípios arquivísticos, cabe uma explanação a respeito da relevância contida nos mesmos, principalmente no ambiente digital.

Atribuído a Natalis de Wailly, historiador francês e chefe da Seção Administrativa do Ministério do Interior francês em 1841, o princípio da proveniência configura-se como base teórica que rege todas as demais intervenções arquivísticas. Ele recomenda que os documentos sejam agregados por fundos, ou seja, deverão ser reunidos em referência a um organismo, quer seja pessoa física ou jurídica, que produz um acervo documental no exercício de suas atribuições. (ROUSSEAU; COUTURE, 1998).

O princípio da proveniência dota a documentação arquivística de uma característica peculiar. Como os documentos são produzidos de forma contínua e necessária, esse princípio faz que o acervo seja unificado em torno das atividades realizadas pela entidade produtora, selando a razão de existência da documentação. A peculiaridade reside no fato de que os tipos documentais são únicos no cumprimento das finalidades para os quais foram produzidos e os

mesmos se inter-relacionam para que seja provida a completa representação dos eventos envoltos na produção documental.

É possível observar que as relações orgânicas existentes em todos os procedimentos funcionais da entidade produtora decorrem na formação do acervo arquivístico e, por conseguinte, o acervo arquivístico representará e comprovará o conjunto de funções e atividades institucionais, o que caracteriza a organicidade. (BELLOTO, 2008).

Os documentos de arquivo, ao contrário do que ocorre com o acervo bibliográfico, não podem ser configurados como itens de coleção. Quando separados do seu contexto original de produção os documentos de arquivo perderão o significado peculiar, uma vez que as informações ali contidas estarão fragmentadas e não poderão ser acareadas com os demais itens do acervo.

Com base nesses princípios a arquivologia volta o seu tratamento documental às necessidades informacionais que os diferentes usuários possam ter durante o ciclo de vida documental; assegurando, portanto, a autenticidade e integridade do acervo para permitir a representação dos eventos administrativos que produziu a documentação.

Em tais fundamentos, a área deverá se embasar para a construção de seu referencial teórico e consolidação de sua atuação frente ao desenvolvimento da Ciência da Informação.

O advento das Tecnologias da Informação ampliou a necessidade de observação dos princípios arquivísticos para a obtenção da autenticidade nos registros informacionais orgânicos digitais. Isso ocorre porque este novo ambiente de produção documental pode eximir o usuário da construção de uma forma de representação que represente a organicidade dos registros produzidos no ambiente institucional.

Os registros documentais poderão ser produzidos e pulverizados nos mais diversificados meios de armazenamento digital, concernentes a mídias off-line, datacenters, e estações de trabalho que se localizam dentro e fora do ambiente institucional. Poderão ainda ser fragmentados em bases de dados ou registrados em variados tipos de formatos digitais, tais como: processadores de textos, *softwares* de planilha, áudio e vídeo etc.

Esta nova modalidade de produção documental amplia a necessidade da aplicabilidade dos princípios arquivísticos, uma vez que eles asseguram o correto tratamento da informação orgânica independentemente do tipo de suporte e asseguram a existência da informação orgânica dotada de autenticidade, integridade e evidência.

2.7 Considerações

O presente capítulo permitiu a compreensão do papel da Arquivologia dentro da Ciência da Informação no que diz respeito ao tratamento dado aos registros informacionais orgânicos, ou seja, os documentos de arquivo.

As definições e conceitos explanados elucidam a relevância trazida pelo tratamento arquivístico para a satisfação dos usuários em todo o ciclo de vida documental, asseverando-se a autenticidade e integridade do acervo.

Observou-se, ainda, que os princípios arquivísticos são de fundamental importância no ambiente digital, dada a necessidade de assegurar-se a integridade e a organicidade do acervo em todo o ciclo de vida documental.

O capítulo seguinte fará uma abordagem do uso de metadados no ambiente digital como uma alternativa para a aplicação dos princípios arquivísticos, seguido de uma explanação sobre as linguagens de marcação, que são utilizadas no registro de metadados. Essa explanação é necessária para a definição do objeto a ser trabalhado nesta dissertação.

3 OS METADADOS ARQUIVÍSTICOS E AS LINGUAGENS DE MARCAÇÃO: ALGUMAS CONSIDERAÇÕES

O capítulo anterior abordou a arquivologia, seus conceitos e seu papel na Ciência da Informação, enquanto campo voltado ao tratamento da informação orgânica.

Visando ainda apresentar definições, o presente capítulo realizará uma abordagem teórica dos conceitos de metadados e sua aplicabilidade no campo arquivístico. Este estudo também versará sobre as características das linguagens de marcação com ênfase na Linguagem XML, utilizada na estruturação de metadados.

No ambiente digital, os metadados e as linguagens de marcação têm apresentado suporte eficiente que permite a apresentação, representação e recuperação da informação orgânica registrada nos documentos arquivísticos.

A apresentação diz respeito ao objeto que contém a informação propriamente dita, isto é, o documento arquivístico que é o resultado final a ser requerido pelo usuário. Para permitir a apresentação no ambiente digital são necessários recursos que permitam a definição e o registro das estruturas de um documento.

A representação compreende a estrutura de elementos que provê a descrição detalhada de um determinado objeto, a fim de possibilitar ao usuário a completa compreensão de seu teor e dotando-o de singularidade. A representação pode ser compreendida, portanto, como uma construção descritiva que se interpõe entre o documento e a necessidade informacional do usuário, a fim de atender os seus interesses e indicar os objetos corretos para a devida recuperação. É necessário, portanto, representar para apresentar. (ARAÚJO JUNIOR, 2007; CERVANTES, 2007; FEITOSA, 2006).

A recuperação da informação ocorre, por sua vez, quando o usuário, por meio de um mecanismo de representação, consegue encontrar o objeto documental que satisfaça suas necessidades informacionais.

3.1 Os Metadados Arquivísticos: Instrumentos de representação do conteúdo, contexto e estrutura documental

Por metadados compreende-se que são dados codificados e estruturados que descrevem características de elementos informacionais para auxiliar na identificação, localização, avaliação e gestão das entidades descritas. Os metadados desenvolvem dois importantes papéis, sendo: inventário, com enfoque nas necessidades institucionais e recuperação, com foco no usuário. (SMIRAGLIA, 2005).

Smiraglia (2005, p. 5) comenta ainda que, na maioria dos casos, os metadados são dados transcritos, obtidos diretamente do recurso informacional e reagrupados, a fim de criar uma representação. Fornece termos estruturados que possibilitam o acesso aos recursos por meio dos sistemas de recuperação da informação. Alguns metadados são de caráter administrativo, voltados ao gerenciamento dos repositórios enquanto outros são estruturais, usados para controle de inventário ou para trabalhar a informação no meio eletrônico.

Para Rondinelli (2002, p. 60), os metadados são elementos que descrevem os atributos de um determinado recurso, bem como o contexto que permita a compreensão do documento de arquivo no decorrer do tempo, dotando o usuário de capacitação para a obtenção do conhecimento preciso acerca da existência e das características dos objetos representados. São, portanto, de fundamental importância para a garantia da capacidade testemunhal dos documentos de arquivo.

É importante ressaltar que a aplicação de metadados não se restringe exclusivamente a documentos digitais.

A composição de sistemas computacionais digitais para a gestão de documentos produzidos em documentos analógicos e digitais ampliam a importância deste recurso informacional para que os registros arquivísticos possam ser recuperados no momento oportuno.

Quanto mais ampla for a representação dos documentos de arquivo por meio da inserção de metadados nos sistemas computacionais, mais fácil será a localização de um item documental. Isso ocorre porque a correta utilização dos metadados dota os agentes de *software* de maior eficácia na recuperação da informação contida em documentos de suporte analógico. É enganosa, portanto, a afirmação de que os metadados somente podem ser aplicados nos documentos digitais.

Os metadados são essenciais ao tratamento da informação orgânica em todos os três estágios do ciclo de vida documental. Podem ser utilizados para descrever conteúdo, contexto e estrutura dos documentos de arquivo, tanto nos suportes analógicos, quanto nos digitais.

Na fase corrente e intermediária, eles trazem suporte aos procedimentos relacionados à gestão documental, uma vez que registram todo o contexto de produção, classificação, uso e avaliação documental. Já na fase permanente, permitem viabilizar a mediação da informação aos usuários por meio do registro dos elementos contidos na descrição documental, além de viabilizar a preservação dos documentos digitais por longo prazo.

Bantin (2008, p. 47) afirma que um dos maiores desafios dos arquivistas consiste em determinar quais tipos de metadados são necessários em um sistema de gestão documental. O autor alega que um sistema de gestão apresenta as seguintes categorias de metadados:

- Identificação ou Registro - identifica o documento arquivístico e registra a data de entrada do documento no sistema de gestão;
- Conteúdo - traz informação sobre o gênero documental predominante no documento dotando-o de caráter textual, iconográfico, cartográfico ou audiovisual;
- Auditoria - identifica as atividades que permearam peças documentais, séries ou arquivos no decorrer do tempo;
- Acesso e uso - define condições para o acesso e uso, inclusive as restrições quanto ao grau de sigilo e dos direitos autorais;
- Destinação - informa por quanto tempo o documento será retido e como a distinção será gerenciada ao longo do tempo;
- Preservação - define como o documento será preservado ao longo do tempo e o impacto das atividades de preservação ao longo do tempo;
- Estrutural - registra as características físicas do documento e sua estrutura interna; e
- Histórico de uso - registra acesso e uso do documento ao longo tempo.

Tais categorias permitem o registro de informações que representam o conteúdo, contexto e estrutura dos documentos de arquivo e permitem que usuários e profissionais da informação gerenciem esses documentos, a fim de satisfazer suas necessidades informacionais. Bantin apresenta um conjunto de metadados para cada uma das categorias mencionadas, os quais serão apresentados nos quadros a seguir:

Quadro 01: Metadados de Identificação ou Registro.**IDENTIFICAÇÃO OU REGISTRO**

METADADO	DEFINIÇÃO
Record Identifier Identificador do Documento	Identificador único do documento.
Date/Time Registered Data e Hora do registro	Data e hora em que o documento entrou no sistema de arquivos.
Location Localização	Localização (física ou lógica) do documento.

Fonte: Bantin (2008).**Quadro 02:** Metadados de Contexto.**CONTEXTO**

METADADO	DEFINIÇÃO
Agent – Originator Agente Originador	Entidade corporativa, unidade organizacional ou indivíduo responsável por desenvolver as ações registradas no documento.
Agent – Receiver Agente Receptor	Quando o documento for transmitido registra-se o agente receptor.
Date/Time – Creation of Record Data e Hora da Produção	Data e hora em que o documento foi criado no curso de sua atividade.
Date/Time – Transmission of Record Data e Hora da Transmissão	Quando o documento for transmitido, registra-se a data de envio pelo agente originador.
Date/Time – Receipt of Record Data e Hora do Recebimento	Quando ocorrer transmissão, registra-se a data e a hora do recebimento do documento.
Record Relationships – Functions and Activities Funções e Atividades Relacionadas	Registros de funções e atividades relacionadas com a produção do documento.
Record Relationships – Related Record(s) Documentos Relacionados	Outros documentos criados pelos mesmos processos orgânicos.

Fonte: Bantin (2008).

Quadro 03: Metadados de Conteúdo.**CONTEÚDO**

METADADO	DEFINIÇÃO
Title Título	O nome dado ao documento, que a representação de conteúdo e função. A tipologia documental cumpre este papel.
Subject Assunto	Assunto ou tópico documental que descreve o conteúdo documental para prover ponto de acesso com nível de detalhe aguçado.
Descripton Descrição	Descrição do propósito do documento, permitindo habilitar pontos de acesso adicionais por meio de palavras-chave.
Coverage Cobertura	Características jurídicas espaciais ou temporais do documento, voltadas à recuperação acerca da jurisdição lugar ou tempo.
Language Idioma	Registra o idioma do conteúdo documental permitindo buscas por idiomas específicos.

Fonte: Bantin (2008).

Quadro 04: Metadados de Termos para Condição e Uso.**TERMOS E CONDIÇÕES DE USO**

METADADO	DEFINIÇÃO
Mandate or Warrant Mandato ou Mandado	Citação e referências às leis políticas e melhores práticas que impõem requisitos para o controle de acesso, referente ao uso do documento.
Acces and Use Terms and Conditions Termos de Uso e Condições de Acesso	Termos sobre os quais os documentos deverão ser acessados e utilizados; a fim de noticiar os usuários das restrições do documento voltadas a proteger a privacidade dos indivíduos e os interesses corporativos.
Responsibility Responsabilidade	Identificação da autoridade responsável por permitir ou restringir o acesso aos documentos.

Fonte: Bantin (2008).

Quadro 05: Metadados de Eliminação.**ELIMINAÇÃO**

METADADO	DEFINIÇÃO
Disposal Authorization or Warrant Mandato ou Autorização de Eliminação	Citações ou referências às leis políticas e melhores práticas que gerenciam o descarte documental, a fim de se assegurar gestão eficiente do documento e cumprir a legislação em torno do gerenciamento arquivístico.
Disposal Date Data de Eliminação	Identifica o prazo de guarda documental e registra a data em que o documento será eliminado.
Disposal Action History Histórico de Eliminação	Identifica quando o documento foi destruído e por quem.

Fonte: Bantin (2008, p.50)

Quadro 06: Metadados de Conteúdo.**ESTRUTURA**

METADADO	DEFINIÇÃO
File Encoding Metadata – Data Fomat Formato de Codificação de Dados	O formato de dados, lógico ou de arquivo do registro de dados, para facilitar recuperação e uso nos procedimentos de preservação digital. Exemplo: ASCII, TIFF e etc.
File Encoding Metadata – Media Format Formato de Mídia	O tipo de representação de dados, a fim permitir a recuperação e o uso do registro no procedimento de preservação digital. Exemplo: texto, imagem, áudio, vídeo.
File Encoding Metadata – Compression /Encryption Compressão e encriptação	O método de compressão e os algoritmos para comprimir ou encriptografar o documento.
File redering Metadata – <i>Hardware/Sotware</i> Dependencies Dependências de <i>Hardware</i> e <i>Software</i>	Registra as dependências de elementos de <i>hardware</i> e <i>software</i> para a renderização correta do registro digital.
Aggregation Level Nível de Agregação	Nível em que o documento está sendo descrito e controlado, podendo ser documento ou séries de classificação.
Extent Extensão	O tamanho físico ou capacidade do documento, a fim de prover informações para ao gerenciamento do espaço físico ou lógico para armazenamento.

Fonte: Bantin (2008, p. 51).

Quadro 07: Metadados de Histórico de Preservação.**HISTÓRICO DE PRESERVAÇÃO**

METADADO	DEFINIÇÃO
Responsibility Responsabilidade	Identifica a autoridade responsável por autorizar procedimentos para a preservação documental.
Activity Date Data da Atividade	Identifica quando as ações de preservação foram finalizadas, a fim de registrar os procedimentos de preservação ao longo do tempo.
Activity Type Tipo de Atividade	Especifica quais procedimentos de preservação foram empregados no documento.
Impacto f The Preservation Process Impacto do procedimento de preservação	Identifica o efeito dos procedimentos de preservação sobre a forma, conteúdo, acessibilidade e uso do documento.
Next Action Próxima Ação	Identifica os procedimentos que deverão ser executados nas próximas etapas da preservação.
Medium Suporte	Identifica o meio físico onde o documento está armazenado, ou seja, o suporte documental.

Fonte: Bantin (2008, p. 54).

Quadro 08: Metadados de Histórico de Uso.**HISTÓRICO DE USO**

METADADO	DEFINIÇÃO
Access and Use History Histórico de Acesso e Uso	Identifica quem acessou e utilizou o documento e quando essa ação ocorreu, a fim de manter registros dos acessos mais significantes e dos usos feitos ao longo do tempo.

Fonte: Bantin (2008, p. 55).

O conjunto de metadados, apresentado nos quadros anteriores, permite a representação pormenorizada do conteúdo, contexto e estrutura documental de um acervo arquivístico, o que contribui para a ágil recuperação dos registros documentais arquivísticos.

É importante também apresentar uma iniciativa tomada pela Câmara Técnica de Documentos Eletrônicos do CONARQ que desenvolveu um documento intitulado: Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos e-ARQ Brasil.

O e-ARQ apresenta uma ampla gama de requisitos obrigatórios, desejáveis ou facultativos que deverão ser encontrados nos sistemas de gestão de documentos digitais. Em relação ao uso dos metadados, são recomendáveis os seguintes requisitos:

Tem que permitir a associação de metadados às unidades de arquivamento, devendo restringir a inclusão e alteração a usuários autorizados. [...] tem que associar os metadados das unidades de arquivamento conforme estabelecido no padrão de metadados. [...] tem que permitir que uma nova unidade de arquivamento herde, da

classe em que foi classificada, alguns metadados predefinidos. [...] deve relacionar os metadados herdados de forma que uma alteração no metadado de uma classe seja automaticamente incorporada à unidade de arquivamento que herdou esse metadado. Pode permitir a alteração conjunta de um determinado metadado em um grupo de unidades de arquivamento previamente selecionado. (ARQUIVO NACIONAL, 2009, p. 54-55).

Os metadados podem, ainda, ser agrupados para formar um esquema. Um esquema de metadados provê estrutura formal para identificar a estrutura do conhecimento de uma dada disciplina por meio da criação de um sistema de informações que auxilia a identificação, localização e uso da informação contida na disciplina. Constitui uma combinação ordenada e sistemática de elementos e regras definidos para um propósito específico, podendo ser:

- Uma coleção de elementos de metadados reunidos para dar suporte a uma função ou a um grupo de funções;
- Uma coleção de metadados formando um contêiner estruturado ao qual são adicionados valores;
- Uma coleção de elementos de dados cujos atributos formalizados em uma especificação (GREENBERG, 2005, p. 24).

Os metadados arquivísticos são registrados em ambiente analógico, por meio de instrumentos de pesquisa impressos, ou em ambiente digital, por meio de registro em linguagens de marcação nos documentos eletrônicos. Os tipos e a estruturação das linguagens de marcação serão apresentados no item a seguir.

3.2 As linguagens de marcação enquanto elementos de estruturação de metadados

Para permitir a apresentação de um documento digital, as linguagens de marcação fornecem instrumental metodológico para a formatação da estrutura documental, podendo estender-se desde os elementos textuais e imagéticos inseridos no documento, até a representação de um formato de arquivo digital, a fim de ser interpretado e apresentado por uma ferramenta de *software*.

Uma linguagem de marcação pode ser compreendida como uma estrutura de *software* que permite a criação de registros textuais assinalados com o uso de etiquetas, do inglês *tags*, que permitem a marcação do texto. Assim, os documentos digitais são redigidos com marcas que delimitam os conteúdos ali inseridos (DACONTA, OBRST, SMITH, 2003; DAUM, 2002; FARIA, 2005).

As linguagens de marcação permitem a emissão de registros de metadados contextualizados que configuram a apresentação e trazem elementos e representação do conteúdo, contexto e estrutura documental.

As etiquetas são configuradas com sinais de maior (“>”) e menor (“<”) e/ou barra (“/”) declarando abertura ou encerramento de um determinado elemento. Temos, portanto, a configuração “<nome etiqueta>” para abertura e “</nome etiqueta>” para encerramento de um elemento. O texto inserido entre as etiquetas de abertura e encerramento é definido como o conteúdo a ser marcado, podendo conter outros elementos inseridos.

O desenvolvimento das linguagens de marcação se deu por volta da década de 1960, quando os profissionais de tecnologia da informação perceberam a necessidade de se elaborar uma linguagem computacional que permitisse a troca de informações entre diferentes sistemas computacionais. Com esse intuito, engenheiros da IBM criaram, em 1969, a *Generalized Markup Language* (GML), linguagem de marcação generalizada.

Em 1986, a *International Standards Organization* (ISO) adotou uma versão aprimorada da GML, denominando-a *Standard Generalized for Markup Language* (SGML). A SGML passa a ser o padrão para a construção de qualquer linguagem de marcação.

Em 1991, Tim Berners-Lee desenvolveu a linguagem *Hyper Text Markup Language* (HTML) que permitiu a criação de regras para a apresentação de textos na internet. O grande volume de textos trouxe a necessidade de se elaborar mecanismos de metadados que recuperassem informações contidas no hipertexto (DACONTA, OBRST, SMITH, 2003; DAUM, 2002; FARIA, 2005).

Dessa necessidade surgiu, em 1996, a linguagem de marcação *Extensible Markup Language* (XML) que permitiu a estruturação de metadados. Estes eventos constituem verdadeiros marcos no desenvolvimento das linguagens de marcação. (YOTT, 2005, p. 214).

Uma vez visto o histórico das principais linguagens de marcação, passa-se a delinear a estrutura envolta em cada uma das linguagens de marcação apresentada, com ênfase em XML.

3.2.1 A linguagem Standard Generalized for Markup Language SGML

A linguagem de Marcação SGML é a precursora de todas as demais e tem por objetivo descrever a estrutura do documento, bem como os atributos que sejam de relevância por meio de uma marcação que apresente critérios rigorosos de definição que assegurem a integridade dos dados contidos nos documentos. Para alcançar essa finalidade, a SGML

apresenta, por característica, marcações descritivas que podem obter tipos estruturados e que são independentes de qualquer plataforma computacional.

Por marcações descritivas, compreende-se a utilização de códigos com nomes facilmente identificáveis para marcar a estrutura de um documento de modo a descrever determinadas partes de um documento. A estrutura utilizada na construção das marcas permite distinguir as marcações dos dados permitindo que diferentes sistemas computacionais naveguem pela estrutura e recupere os conteúdos assinalados (DACONTA, OBRST, SMITH, 2003; DAUM, 2002; FARIA, 2005).

Os registros provenientes da SGML são, portanto, independentes de plataformas de *software* ou *hardware*. A SGML não possui um conjunto pré-definido de *tags* e nem uma semântica pré-definida. Assim, a SGML pode ser usada como uma metalinguagem, servindo de padrão de referência para a criação e descrição de outras linguagens de marcação. (FARIA, 2005, p. 8-10).

3.2.2 A Linguagem HTML

A HTML é uma linguagem que traz um conjunto predefinido de etiquetas que permitem a configuração da apresentação das páginas web. A tag HTML é o elemento raiz e dele advém toda a estrutura dos elementos da página que serão apresentados a seguir:

- HEAD, que delimita os elementos de cabeçalho da página web, tais como título e metadados para inserção de algumas palavras-chave;
- BODY, que permite a inserção de elementos que formatarão a página a ser apresentada no navegador, tais como textos e imagens.

O componente HEAD, que registra informações de cabeçalho, traz subelementos que registram metadados que permitem registrar pontos de acesso para a busca e recuperação do texto registrado no corpo da página.

Dentre os elementos de cabeçalho, destacam-se o elemento META que permite o registro de metadados fixados fora do corpo da página que informam palavras-chave, nome do autor e etc., e são invisíveis ao usuário. Outro componente de cabeçalho a ser mencionado é o TITLE, que define o título a ser apresentado na barra superior do navegador.

O componente BODY registra a informação do corpo da página a ser apresentada no navegador. Os elementos ali contidos permitem:

- Configuração de textos, o que inclui a delimitação e configuração de parágrafos, bem como dos caracteres;

- Configuração de tabelas, com a descrição de linhas, colunas e bordas;
- Inserção de componentes iconográficos e audiovisuais;
- Configuração de elementos para o preenchimento de formulários;
- Etc. (DACONTA, OBRST, SMITH, 2003; DAUM, 2002; FARIA, 2005).

O exemplo a seguir mostra o resultado da apresentação no browser da página HTML descrita abaixo:

Figura 01: Exemplo de utilização da linguagem HTML: catálogo de endereço.



Fonte: O autor.

O elemento HTML abrigará todos os elementos da página. O elemento de cabeçalho, *head*, apresenta os metadados para a descrição do documento. O elemento *title* define a mensagem a ser exibida na aba e na barra de título do navegador.

O elemento de corpo da página, *body*, e o local que abriga o conteúdo e suas instruções para a formatação. O Elemento <P> indica a abertura de um novo parágrafo no texto, enquanto o elemento <u> torna o texto sublinhado. O elemento deixa o texto configurado como negrito, enquanto o elemento
 gera uma quebra de linha.

Ao identificar essas marcas, o navegador configura a formatação do texto existente entre os elementos de acordo com as especificações sugeridas. É importante observar que a

linguagem HTML traz recursos consolidados para a apresentação do conteúdo que é bem compreensível aos seres humanos.

É possível também observar que as máquinas, por sua vez, ficam completamente impossibilitadas de estabelecer relação semântica entre os componentes textuais, uma vez que a HTML reduzirá a informação a um conjunto de caracteres que receberão a descrição da uma forma de apresentação.

Consequentemente, a informação pode ser apresentada, mas não poderá ser representada. Os dados contidos na formatação de apresentação da HTML são manipulados com grande dificuldade por sistemas computacionais de busca e recuperação da informação, haja vista que os dados ali apresentados não são passíveis de serem relacionados e recuperados de modo preciso.

Por esse motivo quando a internet se popularizou, em meados da década de 90, e houve um grande crescimento exponencial de registros informacionais na *web*, surgiu uma grande dificuldade em prover a recuperação da informação, já que muitas vezes a informação, embora estivesse disponível, não estava acessível.

Preocupada com essa questão, a *World Wide Web Consortium (W3C)* lançou, em 1996, a linguagem XML que não traz nenhuma predefinição de etiquetas para a estruturação do conteúdo a ser demarcado. Isso permite definição de elementos de metadados para qualquer conteúdo que se queira representar.

3.2.3 A linguagem Extensible Markup Language XML

A partir de 1998, foi divulgada a recomendação da XML pelo W3C que, desde então, começou ter um progresso de uso. Inicialmente, isso se deu de forma lenta, porque os profissionais de TI pre-supunham, de forma errônea, que a XML seria uma sucessora da HTML. Uma vez percebido que o foco da linguagem não estava voltado à apresentação de páginas web, o uso da linguagem cresceu rapidamente. A XML manteve cerca de 95% da funcionalidade da SGML com complexidade bem menos acentuada.

Para Daum (2002, p. 24), uma das maiores finalidades da linguagem XML é a criação de outras linguagens de marcação que se estendem para todos os tipos de aplicações. Por este motivo, recebe o nome de linguagem de marcação extensível, uma vez que permite a criação de quaisquer estruturas de etiquetas para a rotulagem dos metadados.

A linguagem permite o armazenamento de diversos tipos de dados de modo autodescritivo, em documentos textuais, permitindo a compreensão humana e o mapeamento dos dados. Devido ao fato das linguagens derivadas seguirem uma mesma estrutura, o

aprendizado e o desenvolvimento de aplicações são beneficiados. A estrutura da XML apresenta algumas regras que serão delineadas a seguir.

Os documentos XML iniciam com uma declaração que será apresentada a seguir:

```
<? XML version= "1.0" standalone = "yes" encoding = "iso-8859-1">
```

O atributo *version* é obrigatório e indica a versão XML a ser trabalhada no documento para que o *software* possa identificar e aplicar os recursos corretos de interpretação dos metadados.

O atributo *standalone* é opcional, podendo ser preenchido com “yes” ou “no”, e indica que serão utilizados recursos que não estão contidos na própria estrutura e deverão ser buscados em outros documentos. Por padrão, recebe o valor yes, mesmo quando não declarado.

O atributo *encoding* é também opcional e indica quais serão os tipos de caracteres utilizados no documento. Poderá ser preenchido com os seguintes valores: UTF-8, UTF-16, ISO-10646-UCS-2, ISO-10646-UCS-4, ISO-8859-1. Em português, o padrão indicado é o ISO-8859-1. (FARIA, 2005, p. 20).

A linguagem XML apresenta os seguintes atributos reservados, a saber: o *language* para definir idioma e o *space* para dar significância aos espaços em branco. O atributo *language* traz profundas contribuições para a interoperabilidade, porque permite a identificação do idioma utilizado no teor documental e permite a configuração de caracteres especiais a serem apresentados na linguagem. O atributo *space*, por sua vez, permite definir a configuração dos espaços em branco.

A estrutura da XML traz ao desenvolvedor a possibilidade de criar, além das etiquetas, os atributos para a inserção de dados nestas *tags*. Desta forma, surge a possibilidade de criação de vários tipos de representação de um mesmo conteúdo. Uma data, por exemplo, poderia receber as seguintes representações:

```
<data> 25/07/2010 </data>
<data dia= "25" mês= "07" ano= "2010" />
<data> <dia> 25 </dia> <mês> 07 </mês> <ano> 2010 </ano> </data>
```

Observa-se que a linguagem não apresenta um padrão definido para a inserção de dados em atributos ou elementos. Alguns autores recomendam a utilização de elementos

quando houver objetos que possam ser distribuídos em categorias, e a utilização de atributos, quando os dados forem característica de um determinado objeto. (FARIA, 2005, p. 23).

Considerando os recursos da XML para a estruturação dos dados poder-se-ia representar os dados contidos na Figura 01, mencionada anteriormente, conforme o exemplo apresentado a seguir:

```
<?xml version="1.0" encoding="iso-8859-1"?>
<instituicoes_arquivisticas>
  <catalogo_endereco>
    <organizacao>
      <nome>Arquivo Nacional</nome>
      <endereco_postal>
        <logradouro>
          <logradouro_tipo>Praca </logradouro_tipo>
          <logradouro_nome> Republica </logradouro_nome>
        </logradouro>
        <numero>173</numero>
        <cidade> Rio de Janeiro </cidade>
        <uf>RJ</uf>
        <cep>20211-350</cep>
      </endereco_postal>
      <endereco_eletronico>
        <web_site></web_site>
        <e-mail>arquivonacional@arquivonacional.gov.br</e-mail>
      </endereco_eletronico>
    </organizacao>
  </catalogo_endereco>
  <telefones_contato>
    <telefone>
      <tipo>fixo</tipo>
      <cod_pais>055</cod_pais>
      <cod_area>21</cod_area>
      <numero>2179-1228</numero>
    </telefone>
    <telefone>
      <tipo>fixo</tipo>
      <cod_pais>055</cod_pais>
      <cod_area>21</cod_area>
      <numero>2179-1273</numero>
    </telefone>
  </telefones_contato>
</instituicoes_arquivisticas>
```

É possível observar que na XML os mesmos dados existentes na Figura 01 são representados envoltos em uma estrutura descritiva que permite tanto aos humanos, quanto às aplicações de *software* contextualizar e mapear as informações ali presentes. As etiquetas ali registradas configuram pontos de acesso que permitem a recuperação da informação em vários níveis de granularidade, ou seja, de detalhamento.

Uma das principais diferenças entre HTML e XML implica na finalidade empregada em cada uma das respectivas linguagens. A HTML está voltada à apresentação de hipertextos

na web, enquanto XML está voltada à estruturação de dados e, ainda, ao desenvolvimento de outras linguagens de marcação. O foco da XML volta-se à representação enquanto a HTML direciona-se à apresentação.

A extensibilidade da XML permite que vários indivíduos criem etiquetas (*tags*) com nomes semelhantes para diferentes finalidades. Embora as aplicações funcionem bem em documentos separados, alguns problemas poderão surgir quando registros em XML criados para diferentes finalidades são mesclados. Para contornar problemas relacionados a colisão de nomes, desenvolveu-se os *namespaces*.

Daum (2002, p. 28) exemplifica os *namespaces* definindo-os como ruas de nomes semelhantes em cidades diferentes em que a cidade representa um domínio que dota a rua de singularidade. A construção dos *namespaces* é pautada no uso de *Uniform Resource Identifier* (URI) ou por um *Uniform Resource Name* (URN). Tais recursos identificadores são construídos a partir de um domínio registrado e apresentam o caminho para a localização do registro que contenha os nomes dos elementos.

Os *namespaces* podem, ainda, ser utilizados de maneira *default* ou prefixa. O uso em modo *default* permite que o domínio apresentado reflita-se em todos os elementos filhos a serem apresentados, enquanto o modo prefixo aponta o domínio a ser utilizado em um elemento específico, sendo os demais pertencentes ao *namespace* default. O exemplo a seguir apresenta a declaração de um *namespace* default e prefixo:

```
<serie_documental xmlns=
'http://www.arquivonacional.gov.br/serie.xml'>
<nome> Recursos Humanos </nome>
<a: tipologia xmlns ='
http://www.arquivonacional.gov.br/tipo_documental.xml'>
...
</a:tipologia>
</serie_documental>
```

No exemplo acima o domínio apresentado junto ao elemento “serie_documental” seria aplicado em todos os elementos declarados no documento. Quando a declaração é feita utilizando como apelido a letra “a:” configura-se o uso do elemento prefixo no qual apenas o elemento “tipologia” obterá o domínio do “tipo_documental”, sendo todos os elementos ainda submetidos ao elemento *default*.

3.2.4 Regras com Document Type Definition DTD

A estrutura ou tipos de marcação são definidos em SGML por meio da *Document Type Definition* (DTD). A DTD define uma gramática para a linguagem de marcação que permite determinar os tipos de elementos que um documento deverá comportar.

É utilizada para definir regras de inserção dos elementos a fim de que um documento seja validado, permitindo ao *software* verificar se o documento está de acordo com os tipos de dados que deveria conter e, caso não haja consenso, gerar um erro informando a inconsistência dos dados.

A declaração de Elementos em uma DTD configura-se da seguinte maneira:

```
<! ELEMENT tipologia ANY>
```

A declaração é aberta com a tag “<! Element”, seguida da declaração da estrutura de cada elemento, podendo ser:

- ANY, que permite conteúdo misto de caracteres e outros elementos;
- EMPTY, que permite declarar um elemento vazio;
- #PCDATA, que contém dados de caracteres analisados não podendo estes apresentar os caracteres “&” ou “<”.

É possível também declarar grupos de subelementos dentro de um elemento com o uso de parênteses () sendo os subelementos separados por vírgula, como no exemplo a seguir:

```
<! ELEMENT tipologia (protocoloinicial, texto, protocolo final)>
<! ELEMENT protocoloinicial #PCDATA>
<! ELEMENT texto #PCDATA>
<! ELEMENT protocolofinal #PCDATA>
```

Os elementos poderão ainda receber caracteres de restrições. O caractere “ponto de interrogação” (“?”) indica a existência de uma ocorrência opcional. O caractere “mais” (“+”) indica a obrigatoriedade de um ou vários elementos. Por fim, o caractere “asterisco” (“*”) indica um conjunto opcional de ocorrências. A inexistência de caracteres de restrição é interpretada como uma ocorrência obrigatória (DAUM, 2002, p. 33).

As DTDs também permitem a definição de atributos que é feita por meio de uma declaração por nome de ATTLIST, que consiste em nome, tipo do atributo, e um valor padrão.

Daum (2002, p. 34-35) afirma que os tipos de atributo receberão uma das seguintes declarações:

- CDATA, que corresponde os dados de caractere;

- NMTOKEN, correspondente a nomes válidos consistidos em letras e caracteres “.”, “-”, “_” ou “:”;
- NMTOKENS, correspondente aos conjuntos de NTOKEN separados por um espaço em branco, por exemplo: (produtor |destinatário |receptor);
- NOTATION, que corresponde à enumeração de símbolos para a notação;
- ID, correspondente a identificação do elemento dotando-o de um nome válido e exclusivo no contexto do documento;
- IDREF, que faz uma referência ao ID de um documento;
- IDREFS, correspondente a uma lista de IDREF;

Para a definição de atributos poderão ser utilizados os seguintes valores padrão:

- #IMPLIED, utilizado quando o atributo não é obrigatório e nem possui um valor padrão;
- #REQUIRED, indicando que o valor é obrigatório e precisa ser indicado;
- “valor entre aspas” é utilizado quando o atributo não está especificado; e
- #FIXED, “valor” quando especificado, as instâncias precisam combinar com esse valor. (DAUM, 2002, p. 34).

O exemplo a seguir apresenta uma aplicação do ATTLIST:

```
<! ATTLIST Documento
  código ID #REQUIRED
  fase (corrente|intermediario|permanente) "intermediario">
```

Nesse exemplo, definiu-se o elemento “documento” que deverá conter um atributo código de identificação obrigatório (“código”) do tipo ID e um atributo opcional “fase” do ciclo de vida documental. Neste caso, se não for especificado, o valor padrão (“intermediário”) será aplicado.

A validação dos elementos provida pela DTD gera uma padronização na inserção de dados em diferentes registros, o que permite a interoperabilidade entre diferentes sistemas computacionais.

A SGML proveu estrutura para a construção de linguagens de marcação posteriores que são utilizadas para uma infinidade de recursos. Dentre essas linguagens, podem ser mencionadas a HTML para apresentação de páginas na *web* e a XML que, dentre outros benefícios, permite a estruturação de metadados voltados à disseminação e recuperação da informação.

3.2.5 Regras com XML Schema

A XML, possui, ainda, um esquema para a validação do conteúdo expresso nos documentos XML, denominado XML Schema (XMLS). Lançada em maio de 1999 e recomendada pela w3c em maio de 2001, a linguagem XMLS permite a definição de uma gramática que apresente uma série de regras para a validação de elementos de um documento XML tendo, portanto, a mesma função de uma DTD.

A respeito da XML Schema, Ferneda (2003, p. 107) salienta que:

A linguagem XML Schema, apesar de ter a mesma função da DTD, possui muitas características que a torna mais poderosa (e mais complexa) do que a DTD. Com a XML Schema é possível não apenas especificar a sintaxe de um documento XML, mas também especificar os tipos de dados de cada elemento desse documento [...] De fato, as definições feitas em XML Schema são elas próprias documentos XML.

São muitos os diferenciais entre XMLS e DTD, sendo um dos principais a explicitação dos tipos de dados a serem preenchidos nos elementos dos documentos XML, conforme será apresentado no exemplo a seguir:

Em uma situação hipotética, em que se encontra um registro em XML da relação de determinados tipos documentais, em fase de guarda intermediária aguardando a destinação final, que tenha uma data de arquivamento nome da tipologia e prazo de guarda em meses ter-se-ia a seguinte estrutura em DTD:

```
<! ELEMENT tipologia (#PCDATA>)>
<! ELEMENT prazo_arquivamento(#PCDATA>)>
<! ELEMENT data_transferencia (#PCDATA>)>
```

Neste exemplo todos os conteúdos são declarados com um tipo genérico de dados, que poderá apresentar dificuldades no tratamento das informações ali registradas.

Já em XMLS ter-se-ia:

```
<xs: element name= "tipologia" type= "xs:string"/>
<xs: element name= "prazo_arquivamento" type= "xs:integer"/>
<xs: element name= "data_transferencia" type= "xs:date"/>
```

Comparando-se os dois exemplos é possível observar que, enquanto em DTD ter-se-ia um lançamento padrão de caracteres que impedisse o tratamento de dados como datas de forma específica, no XMLS o nível de detalhamento permite maior eficiência no tratamento da informação registrada nos diferentes elementos.

O uso de DTD e XMLS está fortemente relacionado ao uso de esquema de metadados para a descrição e representação de acervos de guarda permanente. Verifica-se a existência de esquemas para a representação de itens documentais, produtores, e instituições arquivísticas. Como o foco deste trabalho está direcionado à gestão documental, o uso dos esquemas não será aprofundado.

3.3 Considerações

Conforme foi possível observar neste capítulo, os documentos de arquivo produzem um vasto conjunto de metadados que precisam ser devidamente tratados para que a organicidade imbricada na documentação possa ser representada e apresentada ao usuário. As linguagens de marcação, em especial a XML, permitem o registro e processamento desses metadados de modo a satisfazer as necessidades informacionais do usuário.

O capítulo seguinte fará uma extensão dos dois primeiros capítulos e promoverá as definições de web semântica, demonstrando ligeiramente suas contribuições para a arquivologia.

4 AS LINGUAGENS DA WEB SEMÂNTICA NO CONTEXTO DA INFORMAÇÃO ARQUIVÍSTICA

Uma vez abordado os fundamentos teóricos da arquivologia, seguido da definição dos metadados e das linguagens de marcação, surge a necessidade de aprofundar a explanação dos elementos envolvidos na abordagem deste trabalho com um estudo dos elementos que compõem a web semântica e sua potencialidade de uso nos acervos arquivísticos. Os estudos correlatos à web semântica são extensos e abrangentes, de modo que uma dissertação não seria suficiente para esgotar a temática.

O presente capítulo, portanto, trará uma breve explanação dos conceitos e ferramentas relacionados à web semântica, e será essencial à compreensão das contribuições das linguagens de marcação para a gestão da informação arquivística nos capítulos posteriores.

Os recursos oriundos das ferramentas da web semântica ampliam a eficiência das linguagens de marcação no tratamento de metadados em prol da representação e recuperação da informação orgânica registrada nos acervos arquivísticos

4.1 A internet e a recuperação da informação

O advento da linguagem HTML possibilitou que conteúdos informacionais fossem difundidos na rede mundial de computadores, fazendo que a web crescesse exponencialmente desde sua criação.

A linguagem HTML fornece diversos recursos relacionados à formatação e apresentação das páginas Web, porém observa-se uma imensa dificuldade na recuperação precisa da informação disponibilizada armazenada nas páginas.

A criação da linguagem XML proveu a infraestrutura necessária à criação de recursos para o desenvolvimento de ferramentas voltadas à representação da informação de modo inteligível tanto aos seres humanos como aos sistemas computacionais. Pautado nesse objetivo, Tim Berners-Lee divulgou em 2001 a web semântica.

4.2 A Web Semântica

Para este trabalho, a web semântica é compreendida como uma extensão da internet atual em que os computadores serão capazes de ‘compreender’, ou seja, processar, relacionar

e dotar as informações com semântica, além de apresentá-las. Isso é viabilizado por meio do uso de ferramentas, estruturadas em linguagens de marcação, que permitirão que as informações compartilhadas na rede sejam dotadas de significado, de modo que, tanto computadores, como indivíduos trabalhem em cooperação pela completa integração de recursos informacionais. (FEITOSA, 2006).

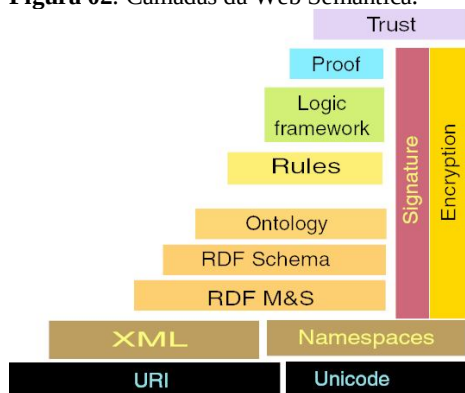
O contexto da web atual requer não somente recursos para disponibilizar a informação na rede. É necessário prover meios automatizados para tornar os recursos acessíveis aos usuários que deles necessitem. Para isso, ocorrer o conteúdo existente na rede deverá ser complementado com dados que permitam às máquinas o relacionamento de informações, sendo os registros efetuados em documentos que possam ser interpretados também por agentes de *software*. Por esse motivo, o termo “Web Semântica” é utilizado para referenciar a internet como uma rede de dados vinculados, a fim de promover a transição da clássica “web de documentos” para a “web dos dados”. (FEITOSA, 2006; LEUF, 2006; W3C, 2010).

Neste contexto, a web semântica é constituída de uma malha de informações relacionadas, em modelos eficientes de representação de dados na rede, em que agentes de *software* coletam conteúdos de fontes diferenciadas, processam essas informações e as disponibilizam para outros elementos de *software* distintos. (FEITOSA, 2006).

4.2.1 Os elementos da web semântica

A web semântica é constituída de um conjunto de metodologias e ferramentas que visam prover sistemas computacionais mais eficientes, bem como o suporte para o desenvolvimento de sistemas que permitam interações confiáveis na rede. Permite, portanto, que indivíduos armazenem dados na rede, que sobre esses construam controles de vocabulário para a recuperação da informação e definam regras para a troca de dados entre sistemas.

São as tecnologias distribuídas em camadas que serão apresentadas na figura a seguir:

Figura 02: Camadas da Web Semântica.

Fonte: W3C (2002).

Cada uma dessas camadas utiliza-se das tecnologias propostas pela camada diretamente inferior e agregam novas funcionalidades para o tratamento da informação. Estes elementos serão apresentados a seguir.

4.2.1.1 Identificadores de Recursos

A camada *Uniform Resource Identifier* (URI) permite o fornecimento de um identificador para um determinado recurso, trazendo-lhe uma referência e permitindo-lhe a obtenção de uma identidade. Essa identificação poderá ser um local, um nome ou ambos.

O padrão *Unicode* também deverá ser utilizado para que todos os documentos sejam registrados em caracteres textuais interpretáveis em todas as plataformas computacionais.

4.2.1.2 Linguagem XML

A linguagem de marcação extensível XML traz uma estrutura que permite a materialização da modelagem dos demais elementos da web semântica. Conforme já mencionado em capítulos anteriores, ela é uma linguagem escrita em texto simples que permite a estruturação de informações possibilitando a interoperabilidade entre diferentes sistemas computacionais. Sobre a XML se enraizarão todas as demais camadas da web semântica. Embora seja uma linguagem de marcação universal, a XML não provê semântica aos dados existentes em seus elementos. Essa atribuição é desenvolvida nas camadas posteriores.

4.2.1.3 O uso de *Namespace*

É importante relembrar que na mesma camada da linguagem XML encontra-se a camada de *Namespace*. Conforme visto anteriormente, essa camada provê meios para se

evitar conflitos entre nomenclaturas similares em domínios de recursos distintos. Provê, portanto, desambiguação dos elementos declarados na linguagem XML (DACONTA, OBRST, SMITH, 2003; DAUM, 2002; FARIA, 2005).

4.2.1.4 Modelagem de Dados com RDF

Por semântica compreende-se ser uma estrutura de relacionamento que se estabelece entre os elementos, a fim de dotá-los de significado. No intuito de prover esses requisitos, elaborou-se a modelagem de dados *Resource Description Framework* (RDF), que será apresentada a seguir.

A camada RDF encontra-se acima da XML e caracteriza-se por ser uma forma de modelagem que permite a rotulação de metadados para descrever as relações semânticas existentes entre recursos.

Feitosa (2006, p. 101) afirma que essa modelagem foi projetada de modo a permitir que um agente de *software* inteligente consiga descobrir, catalogar e “compreender” o recurso de um web site. Afirma, ainda, que:

O RDF é um esforço para a criação de convenções que controlem como a semântica, a sintaxe e a estrutura de metadados são formuladas em um domínio de modo que, quando tais estruturas forem misturadas por um segundo ou terceiro domínio, as declarações originais mantenham sua clareza e legibilidade.

Para o W3C (2004) o desenvolvimento do RDF se pautou em:

- Prover metadados com informações dos recursos web, bem como dos sistemas que os utilizam;
- Permitir que a informação seja processável também pelas máquinas, inclusive fora do ambiente em que foi criada, para favorecer o ambiente da internet;
- Promover interconexões entre aplicações web, de modo a combinar informações existentes em diversos ambientes computacionais, a fim de criar novas informações;
- Processamento automatizado das informações existentes na web pelos agentes de *software*.

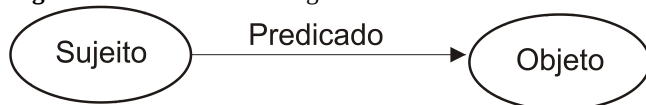
Para codificar o significado dos recursos da web, a RDF trabalha com os elementos apresentados a seguir:

- Recurso - qualquer informação a ser incorporada dentro de um URI, podendo ser um nome ou uma localização na web;

- Propriedade - um recurso nomeado com a finalidade de ser utilizado como propriedade em uma declaração;
- Declaração - é uma afirmação de relacionamento entre um recurso, uma propriedade e um valor;
- Citações - são a realização de uma declaração sobre outra declaração. (LEUF, 2006, p. 131).

As declarações de recursos em RDF podem ser representadas em um modelo de dado gráfico apresentado na imagem a seguir:

Figura 03: Modelo de dados gráfico em RDF.



Fonte: O autor.

A figura anterior apresenta os elementos existentes em uma declaração em RDF, a saber: sujeito, predicado e objeto.

Sujeito é o recurso principal de uma declaração sendo vinculado a ele um predicado, que se constituirá de um nome referente a uma propriedade, e um objeto que será o valor de uma propriedade. (DAUM, 2002; LEUF, 2006).

Considerando o contexto de produção documental onde uma atividade institucional gera uma tipologia seria possível determinar uma relação em que a atividade institucional figura-se como sujeito, o verbo “gera” denomina a propriedade, e a tipologia documental, por fim, seria o objeto que figura o valor da propriedade. Essa exemplificação poderia, ainda, ser convertida na seguinte frase: Pedido de Venda gera Nota Fiscal. Os modelos de dados gráficos são apresentados na figura a seguir:

Figura 04: Relacionamento Atividade e Tipo Documental.



Fonte: O autor.

Esse relacionamento pode ser declarado em RDF. Estas declarações recebem o nome de Tripla.

Os exemplos a seguir apresentam uma declaração em RDF do relacionamento existente entre atividade e tipologia documental:

Figura 05: Exemplos de declarações em RDF.

<pre><?xml version="1.0" ?> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:tipodoc="http://www.arquivos.com/tipologia-ns"> <rdf:Description rdf:about="atividade"> <tipodoc:gera> tipologia </tipodoc:gera> </rdf:Description> </rdf:RDF></pre>	<pre><?xml version="1.0" ?> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:tipodoc="http://www.arquivos.com/tipologia-ns"> <rdf:Description rdf:about="pedidoavenda"> <tipodoc:gera> NotaFiscal </tipodoc:gera> </rdf:Description> </rdf:RDF></pre>
---	---

Fonte: O autor.

Como é possível observar, os documentos RDF iniciam com a tag `<?XML>` indicando que a linguagem de marcação utilizada na modelagem é a XML. Em seguida, surge a etiqueta `<rdf:RDF>` que dá abertura à modelagem RDF.

Essa etiqueta apresenta, ainda, namespaces para a indicação dos recursos em que se estruturam alguns dos elementos que constituirão a tripla, ou seja, a declaração. Nesse exemplo específico encontram-se as referências da RDF pela W3C e um caminho fictício para o namespace tipodoc. O endereço apresentado trará, portanto, URIs que definirão o predicado “gera”, a atividade e a tipologia, bem como o pedido e a Nota Fiscal. Tais elementos são denominados prefixos. Uma vez definidos os prefixos em um domínio torna-se possível o uso dos recursos ali descritos em uma declaração RDF.

Na sequência encontra-se o elemento `<rdf:Description>` o qual estabelece as declarações das triplas em RDF. Esse documento apresenta o recurso “rdf:about” que define o sujeito da declaração.

Em seguida, se estabelece a relação existente entre os elementos embasando-se nos recursos definidos nos namespaces anteriores, no exemplo, tipodoc. Neste exemplo a declaração relaciona o recurso “atividade” com o objeto “tipologia” por meio da propriedade “gera”. É relacionado, ainda, o recurso “PedidoVenda” ao objeto “NotaFiscal” por meio da propriedade “gera”.

Com essa modelagem, tanto computadores como seres humanos podem relacionar a tipologia documental com a atividade produtora.

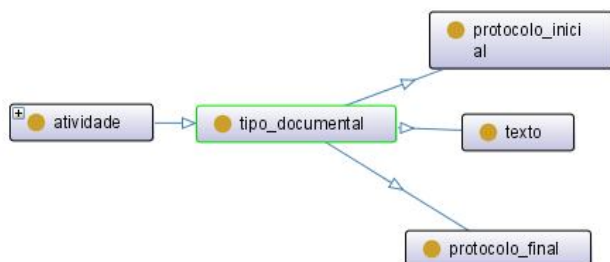
A RDF traz a possibilidade e reunir um vasto grupo de recursos dentro de um contêiner. São três os elementos que pode defini-lo, a saber:

- `rdf:Bag` - permite contêiner de recursos sem nenhum critério de ordenação;
- `rdf:Seq` - permite a criação de um contêiner em que os recursos são agrupados em uma sequência de ordenação; e

- `rdf:Alt` - permite um grupo de alternativas a um recurso preferível, por exemplo, um documento e sua tradução em diversas linguagens. (ANTONIOU, HARMELEN, 2004; DAUM, 2002).

Conforme visto em capítulos anteriores uma determinada tipologia documental pode ser dividida de protocolo inicial, texto e protocolo final. A figura a seguir apresenta um grafo em que esses elementos são relacionados.

Figura 06: Elementos de um tipo documental.



Fonte: O autor.

Os elementos poderão ser representados em um contêiner a ser relacionado à tipologia, conforme o exemplo a seguir.

Figura 07: Contêineres em Bag e Seq.

<pre> <?xml version="1.0" ?> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:tipodoc="http://www.arquivos.com/tipologia-ns"> <rdf:Description rdf:about="http://www.arquivos.com/tipologia-ns"> <tipodoc:elemen_intrins> <rdf:Bag> <rdf:li> Protocolo_Inicial</rdf:li> <rdf:li> Texto</rdf:li> <rdf:li> Protocolo_Final</rdf:li> </rdf:Bag> </tipodoc:elemen_intrins> </rdf:Description> </rdf:RDF> </pre>	<pre> <?xml version="1.0" ?> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:tipodoc="http://www.arquivos.com/tipologia-ns"> <rdf:Description rdf:about="http://www.arquivos.com/tipologia-ns"> <tipodoc:elemen_intrins> <rdf:Seq> <rdf:li> Protocolo_Inicial</rdf:li> <rdf:li> Texto</rdf:li> <rdf:li> Protocolo_Final</rdf:li> </rdf:Seq> </tipodoc:elemen_intrins> </rdf:Description> </rdf:RDF> </pre>
--	--

Fonte: O autor.

Nesses exemplos é possível observar a existência da propriedade `<rdf:li>` que descreve elementos inseridos nos contêineres.

A linguagem RDF permite ainda a construção de reificações, ou seja, declarações sobre outras declarações. Isso é possível porque o sujeito de uma declaração é um recurso e, conseqüentemente, quando uma declaração é convertida em um recurso ela poderá ser utilizada como sujeito de outra declaração. Daum (2002, p. 137) comenta que, dentre as diversas possibilidades de reificação, a mais fácil de utilizar é o atributo *bag:ID* dentro do elemento *Description*.

Para exemplificar a reificação, considera-se a seguinte declaração: uma função é atribuída a uma atividade que gera uma tipologia.

```
<rdf:RDF>
<rdf:Description bagID= "SobreAtividade" about= "Atividade">
<tipodoc:gera>tipologia</tipodoc:gera>
</rdf:Description>
```

O elemento bagID reifica a declaração, dando-lhe o nome de “SobreAtividade”, que poderá ser usado nas declarações posteriores, conforme apresentado no exemplo a seguir:

```
<rdf:RDF>
<rdf:Description aboutEach= "#Sobre Atividade">
<tipodoc:atribuidopara> Funcao</tipodoc:atribuidopara>
</rdf:Description>
```

Nesse exemplo, o atributo aboutEach relaciona o recurso “SobreAtividade” com o termo “Função”, formando uma reificação sobre a declaração anterior. A modelagem RDF trabalha com recursos individuais. É necessário, no entanto, acrescentar uma descrição de vocabulários que permita o mapeamento de classes e subclasses de recursos. Esta finalidade é alcançada com a Resource Description Framework Schema (RDFS).

A RDFS consiste em um conjunto de recursos e propriedades RDF que permite aos usuários criarem seus próprios vocabulários. (ANTONOU; HARMELEN, 2004).

Sobre essa linguagem, Feitosa (2006, p. 108) afirma que a modelagem RDF pode ser pensada como propriedades condizentes a atributos de recursos. Afirma, ainda, que “[...] é papel da RDFS Schema descrever tais propriedades e suas relações com propriedades de outros recursos, o que é feito por meio da definição de classes e propriedades que podem ser utilizadas para descrever classes, propriedades e outros recursos”.

A RDFS se assemelha com as linguagens de programação orientada a objetos, permitindo a definição de recursos como instâncias de classes e subclasses. Essa linguagem, portanto, provê recursos que permitam a construção de taxonomias dotadas de semântica aplicada nos esquemas de classificação e, por conseguinte, o relacionamento de terminologias.

Esta modelagem provê estrutura para a construção de vocabulários controlados, em sistemas computacionais, sendo ferramenta indispensável ao tratamento e disseminação da informação aos usuários que dela necessitarem. O quadro a seguir apresenta uma listagem de propriedades da RDFS.

Quadro 09: Propriedades da RDFS.

PROPRIEDADE	DESCRIÇÃO
rdfs:Class	Define uma classe que compartilhará uma série de propriedade.
rdfs:label	Atributo que define uma etiqueta para a classe.
rdfs:subClassOf	Especifica que a classe será subordinada a outra classe já existente.
rdf:Property	Elemento que define uma propriedade para uma classe e um conjunto de valores que isso possa representar.
rdfs:domain	Define em que classe se enquadrará uma determinada propriedade.
rdfs:range	Define um conjunto de valores a serem aplicados em uma propriedade.
rdfs:seeAlso	Faz referência a um recurso que provê informação adicional.
rdfs:isDefinedBy	Define o namespace para um assunto.
rdfs:comment	Traz informação descritiva adicional a respeito das classes e propriedades do schema.
rdfs:Literal	Representa um valor constante representado.

Fonte: (DACONTA; OBRST; SMITH, 2003, p. 108-109).

Por meio desses elementos, é possível definir, portanto, quaisquer taxonomias ou instrumentos de classificação que sejam necessários.

Considere-se a classificação trabalhada em arquivologia em que a classificação segue uma estrutura hierárquica entre Função, Atividade e Tipologia. É possível desenvolver meios para representar as respectivas classes mencionadas em RDFS de acordo com o exemplo a seguir:

```
<?xml version="1.0"?>

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="http://www.arquivos.com/classificacao#">

  <rdf:Description rdf:ID="Funcao">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:comment>
      Funções institucionais são executadas para que a organização alcance
      sua missão institucional.
    </rdfs:comment>
  </rdf:Description>

  <rdf:Description rdf:ID="Atividade">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:subClassOf rdf:resource="#Funcao"/>
    <rdfs:comment>
      Esta é a classe das tipologias documentais. Ela é uma subclasse da
      atividade
    </rdfs:comment>
  </rdf:Description>
</rdf:RDF>
```

```

</rdf:Description>

<rdf:Description rdf:ID="Tipologia">
  <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
  <rdfs:subClassOf rdf:resource="#Atividade"/>
  <rdfs:comment>
    Esta é a classe das tipologias documentais. Ela é uma subclasse da
    atividade
  </rdfs:comment>
</rdf:Description>

</rdf:RDF>

```

O exemplo mencionado trabalha com as definições de classes que vão desde a tipologia documental, até a função institucional. Nesse exemplo foi utilizado o atributo `rdf:ID` para a definição das classes; o `rdf:comment` para especificar um comentário da finalidade da classe; e do elemento `rdf:subClassOf` para indicar a classe superior relacionada.

Por meio da estrutura oferecida por RDFS, é possível prover mecanismos de *software* inteligentes para gerenciar a construção de planos de classificação de documentos arquivísticos e disseminar a informação orgânica na fase corrente, conforme será demonstrado nos capítulos posteriores.

4.2.1.5 Representações de Ontologias com OWL

Os recursos trazidos pela RDF e RDFS à web semântica, embora expressivos, são limitados. As descrições limitam-se a predicados binários e hierarquias de subclasses. Não é possível, por exemplo, promover negação entre os atributos ou relacionar elementos específicos das partes. Para aprimorar e ampliar a descrição de recursos na web desenvolveu-se a camada ontológica da web semântica, que se embasa na linguagem de marcação Web Ontology Language (OWL).

Não é o foco do presente capítulo aprofundar-se e esgotar a temática envolvendo o uso das ontologias, uma vez que essas são por demais extensas aos limites impostos pela pesquisa. É necessário, no entanto, apresentar uma breve introdução ao conceito de ontologias e de OWL.

O conceito de ontologias advém da filosofia grega em que Aristóteles (384-322 a.C.) afirmava que o “EU” consistia em um modelo informal de lógica fazendo declarações sobre o mundo dos objetos, o “Não-EU”. Esta corrente foi incorporada pela Igreja Católica por Tomás de Aquino (1225 – 1274) e influenciou o pensamento do mundo ocidental durante a Idade Média. A partir do século XVI, os reformadores começam a questionar toda a base ideológica da igreja, o que trouxe rupturas na ideologia dominante e monolítica da Europa Medieval. Por

volta do fim do sec. XVIII, Georg Wilhelm Friederich Hegel (1770 – 1831) defendeu a existência de mais de uma visão possível do mundo, havendo a possibilidade de existir múltiplas ontologias. No século XX, Gotthard Gunther combinou, no ano de 1979, um sistema de ontologias múltiplas com cálculo lógico multinivelado. Isso veio a substituir, finalmente, a visão aristoteliana do mundo por uma sociedade de vários observadores. (DAUM, 2002, p. 168-169).

Stewart (2008, p. 163) afirma que, essencialmente, a ontologia é um quadro que define um grupo específico de conceitos. As relações existentes entre esses conceitos, bem como a natureza envolvendo estes relacionamentos. Configura-se, portanto, em uma especificação explícita e formal de conceitos compartilhados. Essa explicitação, segundo o autor, define e detalha todos os conceitos, propriedades e restrições em um domínio específico.

As ontologias consistem em vocabulários classificados em taxonomias que são inseridos em um sistema relacional. Podem as ontologias ser dotadas de restrições. São, portanto, difundidas nos seguintes níveis: Léxico, Taxonomias, Thesaurus, Modelo Relacional, Teoria totalmente axiomada (restrita). (DAUM, 2002).

São quatro os tipos de ontologias, a saber:

- Ontologias de nível superior, que definem termos generalizados e são compartilhados por uma comunidade;
- Ontologias de domínio, que se aplica a um domínio específico da área do conhecimento;
- Ontologias de tarefa, que são aplicáveis a procedimentos específicos; e
- Ontologias de aplicação, que se especializa em tarefas e domínios específicos.

Além das restrições existentes pela classificação de taxonomias ou mesmo tesauros, as ontologias possuem recursos pautados em lógica e outros elementos matemáticos que lhe permitem promover a representação do conhecimento, tamanha a riqueza dos recursos de descrição por ela providos.

No contexto computacional, as ontologias permitem a definição de maiores elementos de vocabulário para a descrição de propriedades e classes, dotando-as de maior riqueza. Por exemplo, permitem a expressão de cardinalidade relacionando grupos de uma taxonomia com elementos específicos de outras classes. Propriedades são definidas com tamanha precisão que permitem estabelecer simetria entre as classes, sendo também possível descrever relações de negação que uma classe possa ter com outra. (STEWART, 2008).

A linguagem OWL é utilizada para prover os recursos e estruturas mencionados no parágrafo anterior, que subsidiem a construção de ontologias no ambiente web. Para cumprir esse fim, a linguagem owl se utilizará de toda a estrutura trabalhada nas camadas inferiores, a fim de que possa se estabelecer. A linguagem define-se em: Owl Lite, Owl DL, e Owl Full. Essas modalidades serão apresentadas a seguir:

A OWL Lite é direcionada aos usuários que necessitam de uma hierarquia de classificação com vínculos simplificados. É uma linguagem de baixa complexibilidade, mas que apresenta uma quantidade limitada de recursos. É menos complexa que OWL-DL e a OWL-Lite, é, naturalmente, um modelo de ontologia OWL-DL.

A OWL-DL, por sua vez, oferece suporte aos usuários que desejam desenvolver ontologias que permitam o máximo de expressividade e, ao mesmo tempo, manter a plenitude computacional. Ela permite o trabalho em conjunto de declarações de linguagem descritiva que traz maior riqueza aos relacionamentos entre as classes. O resultado é que permite aos computadores maior flexibilidade para automatizar tarefas de conclusão e decisão. Permite a criação automática de relacionamentos entre classes e vocabulários. Apresenta, no entanto, algumas restrições na estrutura RDF. A OWL-DL é um modelo de OWL FULL.

A OWL FULL permite que os usuários tenham o máximo de expressividade com abertura total às classes RDF. Não permitem, no entanto, uma garantia de processamento computacional automatizado dos relacionamentos existentes.

Será apresentado a seguir um fragmento de código referente a uma descrição do relacionamento entre Função, Atividade e Tipologia Documental na linguagem OWL:

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xml:base="http://www.arquivos.com/#">
  <owl:Class rdf:ID="Funcao">
  <rdfs:comment>

  "Funcao" é o elemento que comporta um conjunto de atividades exercidas
  por um indivíduo ou organização.

  </rdfs:comment>
  </owl:Class>

  <owl:Class rdf:ID="Atividade">
  <rdfs:comment>

  "Atividade" comporta um conjunto de transações que geram as tipologias
  documentais.

  </rdfs:comment>
  <rdfs:subClassOf rdf:resource="#Funcao"/>
```

```

</owl:Class>

<owl:Class rdf:ID="Tipologia">
<rdfs:comment>
tipo documental
</rdfs:comment>
<rdfs:subClassOf rdf:resource="#Atividade"/>
</owl:Class>

</rdf:RDF>

```

A partir da construção das classes, a OWL permite desenvolver uma série de operações lógicas que relacionam os conjuntos inseridos na declaração.

A linguagem permite, por exemplo, afirmar que um elemento não pode ser membro de duas classes simultâneas, por meio da declaração *disjointof*.

É possível também especificar o número exato de elementos a ser utilizado em uma relação, ou seja, a cardinalidade, com o elemento *owl:cardinality*

Por meio da declaração *equivalentClass*, é possível relacionar equivalências entre classes de ontologias distintas.

É possível, por exemplo, criar relacionamentos de união, interseção, e complemento por meio dos elementos: *intersectionOf*, *unionOf*, *complementOf*.

Essas são apenas algumas das muitas propriedades a serem trabalhadas dentro das ontologias descritas em linguagem OWL.

Recentemente, em outubro de 2009, a W3C recomendou a linguagem OWL 2. A linguagem manteve relacionamentos com XML e RDF. A estrutura da OWL é perfeitamente compreensível e validável na OWL 2, e esta nova versão acrescentou alguns recursos, a saber:

- Chaves;
- Cadeias de propriedades;
- Intervalos e tipos de dados enriquecidos (W3C, 2009b);
- Restrições de cardinalidade qualificadas;
- Propriedade para disjunção, reflexão e assimetria; e
- Aprimoramento nos meios para anotações.

Não é objetivo deste trabalho aprofundar na descrição da OWL2. Resta dizer que a linguagem trabalha com três perfis independentes entre si denominados:

- OWL 2 EL - é indicado para aplicações que empreguem ontologias com um número de classes e propriedades muito amplo. A sigla EL está relacionada a um grupo da descrição lógica que se volta a quantificação Existencial;

- OWL 2 QL - é indicada para aplicações cujas consultas a vastos conjuntos de dados seja a atividade predominante. Esta linguagem pode ser integrada a sistemas de bancos de dados relacionais. A sigla QL tem por significado *Query Language*, linguagem de consultas.
- OWL 2 RL - é recomendável para aplicações que requerem raciocínio escalável sem prejudicar a força de expressão. Relaciona a eficiência das linguagens de regras com a expressividade da modelagem RDF. A sigla RL tem por significado *Rule Language*, linguagens de regras. (W3C, 2009b).

A OWL1 e OWL2 são linguagens que permitem representar com grande expressividade os recursos informacionais de um domínio do conhecimento de modo a mapear e recuperar a informação registrada na internet.

4.2.1.6 A Representação de Tesouros e Taxonomias com SKOS

Ainda na camada ontológica, é necessário mencionar outra linguagem de marcação, que recebeu recomendação da W3C em junho de 2009, denominada *Simple Knowledge Organization System* (SKOS). Essa linguagem é direcionada à modelagem de tesouros, taxonomias, esquemas de classificação e cabeçalhos de assunto. Estes instrumentos compartilham uma estrutura em comum e são utilizados em aplicações similares. A linguagem permite integrar ferramentas de representação do conhecimento à web semântica, realçando o uso das linguagens RDF e OWL. (W3C, 2010b).

No que diz respeito à modelagem de tesouros, a SKOS traz uma série de recursos que permite definir termos e relacioná-los com outros termos dentro da estrutura da RDF e OWL, trazendo maior formalidade a estrutura descritiva dos relacionamentos.

Alguns dos atributos são apresentados no quadro a seguir:

Quadro 10: Alguns recursos SKOS.

PROPRIEDADE	DESCRIÇÃO
skos:prefLabel	Termo adotado.
skos:related	Termo relacionado.
skos:altLabel	Termo alternativo.
skos:scopeNote	Nota explicativa.
skos:narrower	Termo genérico.
skos:broader	Termo específico.

Fonte: W3C (2009).

O presente quadro não traz todos os recursos apresentados na recomendação da W3C, mas apresenta alguns elementos que poderão ser combinados às linguagens RDF e OWL para dar maior formalidade à representação dos termos de um tesouro.

As linguagens RDF, RDFS, OWL, e SKOS podem ser combinadas para prover a construção de tesouros com a representação formal dos relacionamentos existentes entre terminologias de um determinado domínio da área do conhecimento, inclusive com integração de termos da linguagem natural com os termos de linguagem controlada. Em capítulos posteriores será apresentada a integração das linguagens da web semântica com os tesouros de funções utilizados na arquivologia.

4.2.1.7 Camada rule

A camada *rule* define regras lógicas que se relacionam com recursos informacionais, possibilitando uma espécie de introdução lógica a ser aprimorada na camada superior. A camada *logic framework* incorpora lógicas avançadas com regras mais abrangentes utilizando-se de informações descritas nos níveis anteriores, o que possibilita a agentes de *software* fazerem inferências automatizadas.

4.2.1.8 Camadas *proof* e *trust*

A camada *proof* tem por objetivo o desenvolvimento de modelagens que comprovem a coerência lógica dos recursos descritos nas camadas inferiores de modo a assegurar que os aspectos semânticos da informação estejam descritos de maneira adequada. E, por fim, a camada *trust* certifica a coerência existente nas descrições das camadas inferiores dotando a semântica de grande índice de confiabilidade. (RAMALHO; VIDOTTI; FUJITA, 2007).

4.3 Considerações

Durante a exposição desse capítulo, foi possível observar que a W3C promove um esforço contínuo no desenvolvimento de ferramentas que assegurem a recuperação dos recursos informacionais da web. Embora estejam voltadas para informações publicadas na internet, as ferramentas da web semântica podem trazer inúmeros benefícios no gerenciamento de recursos informacionais arquivísticos em sistemas de informações semifechados. Neste contexto, usuários em uma determinada instituição poderão automatizar a recuperação da informação orgânica concernente aos seus interesses, por meio do

relacionamento existente entre os recursos informacionais orgânicos e as necessidades informacionais geradas por suas atividades.

O uso integrado das linguagens RDF, OWL e SKOS parece prover os sistemas digitais de gestão de documentos de recursos significativos para a representação e recuperação da informação orgânica durante as fases corrente e intermediária do ciclo de vida documental. Isso favorece a indexação e a construção de taxonomias com vocabulários controlados que podem ser empregados nos registros documentais.

Os capítulos posteriores versarão a respeito do comportamento informacional dos usuários do arquivo corrente e intermediário, bem como do tratamento informacional necessário ao suprimento das demandas. Em seguida, será abordada a Gestão Documental e suas ferramentas de classificação, para que possam ser apresentadas sugestões de uso das linguagens da web-semântica integradas à recuperação da informação orgânica em sistemas de gestão de documentos arquivísticos.

5 O COMPORTAMENTO INFORMACIONAL DOS USUÁRIOS DE ARQUIVOS CORRENTES E INTERMEDIÁRIOS

Os capítulos anteriores possibilitaram obter uma ampla compreensão das temáticas a serem trabalhadas no cumprimento dos objetivos desta dissertação.

Surge, neste momento, a necessidade de direcionar a discussão à gestão digital da informação arquivística, iniciando com o direcionamento do foco ao usuário do grupo primário, o qual atua nos documentos arquivísticos da fase corrente e intermediária.

Nesse contexto, o presente capítulo descreve o comportamento informacional dos usuários de arquivo na fase corrente e intermediária, que corresponde à produção e utilização da documentação nas atribuições institucionais.

Conforme já mencionado, os documentos serão produzidos e gerenciados pelo produtor durante as duas primeiras fases do ciclo de vida documental. Ocorre, neste período, a vigência do valor primário, em que o documento é utilizado pelo administrador no cumprimento de atribuições específicas da missão institucional. Nesta mesma fase, a informação orgânica atende às finalidades de caráter administrativo, fiscal e legal, sendo o comportamento do usuário voltado à satisfação destas necessidades.

5.1 O Comportamento Informacional do usuário de arquivo na fase corrente

O registro documental é realizado para dar suporte aos eventos ocorridos em âmbito institucional, dotando-os de evidência. A documentação permite registrar o trâmite existente na produção documental, materializando-os no suporte documental. Isso ocorre quando o usuário registra o endereçamento do destinatário no documento, bem como quando o destinatário registra a ciência do recebimento com os encaminhamentos executados.

Neste contexto, o comportamento informacional dos usuários interage com a informação orgânica registrada nos documentos arquivísticos, a fim de produzir novos registros informacionais orgânicos na documentação.

Wilson (2000, p. 1) define comportamento informacional como a totalidade do comportamento humano em relação a fontes informacionais, no que tange a busca e o uso da informação. Por comportamento informacional compreende-se a busca realizada por um usuário em fontes informacionais.

O autor ainda defende que a interação do usuário poderá ser ativa ou passiva. A interação ativa ocorre quando o usuário busca a informação para satisfazer uma necessidade

específica. Já a interação passiva ocorre quando o usuário recebe a informação sem ter realizado uma busca, por exemplo, ao passar em frente a uma televisão ligada. O foco deste trabalho será voltado à interação ativa.

O estudo do comportamento do usuário permite aprimorar os sistemas de informação, a fim de que os recursos informacionais possam ser recuperados e disponibilizados em conformidade com a solicitação demandada.

Teruel (2005, p. 67) afirma que um procedimento de busca envolve o usuário, em primeiro plano, seguido, em segundo plano, de suas necessidades e desejos informacionais em conjunto com a demanda realizada a um sistema e o uso da informação obtida.

Ainda segundo o autor, o usuário pode ser compreendido como um indivíduo ou organização que necessita de uma informação especializada em um sistema de informação, podendo ser potenciais ou reais. Usuários potenciais são aqueles que fazem parte de um grupo que possa beneficiar-se dos recursos informacionais de um sistema de informação; enquanto os usuários reais são aqueles que buscam e apropriam-se da informação contida em um sistema.

A necessidade de informação é compreendida como os recursos informacionais que um determinado indivíduo deverá possuir, a fim de executar um procedimento no trabalho, uma pesquisa, ou ainda para obter realização pessoal. É, portanto, qualquer experiência de um usuário em relação à busca de informação.

O autor comenta ainda que na necessidade informacional compreendem-se quatro níveis sequenciais, que evoluem de um nível inconsciente de necessidade informacional à formulação de questões específica para a busca em um sistema, a saber:

- Necessidade visceral, em que há uma necessidade de informação não expressada presente no inconsciente;
- Necessidade consciente, em que surge uma descrição consciente de uma área de indecisão ainda mal definida;
- Necessidade formalizada, em que o indivíduo constrói uma declaração formal de necessidade de informação; e
- Necessidade comprometida, em que o indivíduo formula questões de recuperação e as apresenta a um sistema de informações (TERUEL, 2005).

Esses quatro níveis, portanto, permitem delinear todo o processo que gera no usuário a necessidade de informação e a constituição do procedimento de busca empregado para obter a satisfação.

O desejo de informação pode ser compreendido como a percepção do próprio usuário de suas necessidades informacionais traduzidas em interesse de obtenção de recursos informacionais.

A demanda consiste na solicitação que um usuário apresenta a um sistema de informação que, por sua vez, desencadeará em um procedimento de estratégias para a busca e recuperação da informação.

De posse das informações que satisfaçam o seu comportamento, o usuário se apropriará da informação e atenderá sua demanda informacional inicial.

O comportamento informacional dos usuários de arquivos correntes e intermediários ocorre em um contexto de produção documental regido pelo relacionamento existente entre funções e atividades dentro de uma missão institucional.

As funções são compostas por um grupo de atividades voltadas a uma finalidade específica. A atividade poderá ser compreendida como um procedimento formalizado que é executado em atendimento às atribuições específicas de cada uma das funções. Esses procedimentos são compostos por um conjunto de transações.

As funções podem ainda ser classificadas em áreas funcionais meio e fim. Funções da área fim relacionam-se diretamente com o cumprimento da missão institucional. Em outras palavras, relacionam-se com a finalidade para a qual a instituição fora criada. A área funcional meio compreende as funções executadas para dar o devido suporte às funções da área fim.

É importante frisar que funções e atividades exercem relacionamentos hierárquicos e associativos. Uma função maior pode estar subdividida em um grupo de subfunções. Cada uma das subfunções contemplará um conjunto de atividades. É possível observar a hierarquia existente em cada um desses elementos funcionais.

Diferentes funções e atividades podem estar associadas umas às outras, produzindo registros informacionais que atendam ao comportamento informacional dos usuários de ambas. Para fins de exemplificação, é possível mencionar as funções de Gestão de Pessoal com Finanças. Neste contexto, as atividades referentes ao pagamento de pessoal associam-se com as atividades de prestação de contas do departamento financeiro. As informações registradas pelas transações de ambas as atividades serão importantes ao suprimento do comportamento informacional dos usuários que executam cargos funcionais em ambas as funções.

Os níveis de hierarquia existentes entre funções e atividades permitem que a informação orgânica seja gerenciada em níveis diferenciados. Isso possibilita delimitar os

níveis de interação informacional exigidos por um determinado cargo funcional para o correto exercício das atribuições. São pelo menos três os níveis informacionais a serem trabalhados, a saber:

- Operacional;
- Tático ou gerencial; e
- Estratégico (BEAL, 2009).

O nível operacional trabalha com os procedimentos corriqueiros da organização. Usuários que exerçam atribuições desse patamar apresentarão comportamento informacional voltado à busca da produção documental das transações rotineiras da instituição, a fim de executar as suas atribuições de maneira eficiente. Observa-se que o enfoque maior do comportamento informacionais desses usuários é direcionado aos trâmites decorridos dentro das atividades institucionais correlatos ao cargo. As informações geradas nas operações das atividades deverão estar disponibilizadas detalhadamente a esses usuários, a fim de que possam delas apropriar-se e executar com melhor eficiência os trâmites institucionais.

O nível gerencial, por sua vez, condensa os dados trabalhados no nível operacional para atender o comportamento informacional de usuários que exercem atribuições no corpo gestor. Neste contexto, os dados gerados pelo nível operacional serão agrupados e sintetizados, a fim de garantir ao usuário uma visão ampliada das operações regulares de uma determinada função institucional. A partir dessas informações, o usuário atua na coordenação das atividades inerentes a uma determinada função no intuito de aprimorar a execução dos trâmites inerentes e alcançar melhores resultados.

O nível estratégico, por fim, procura integrar e sintetizar as informações do nível operacional e gerencial para obter uma visão pormenorizada das funções institucionais e identificar se elas estão em harmonia com a missão institucional. Essas informações são ainda comparadas com informações existentes no ambiente externo da instituição. A partir de análises de informações do ambiente interno e externo é possível efetuar tomadas de decisões que nortearão o gerenciamento de estratégias para que a organização possa sobreviver em um ambiente turbulento e obter eficácia no cumprimento de sua missão institucional. Essas estratégias determinam o rumo a ser seguido pela organização. São integrantes deste grupo os diretores e presidentes da instituição. (BEAL, 2009; OLIVEIRA, 2009).

A delimitação destes níveis informacionais permite especificar as peculiaridades envolvidas no comportamento informacional dos diferentes usuários de arquivo, na fase de valor primário, e desenvolver estratégias para atender as especificidades inerentes aos usuários de cada um dos níveis.

Beal (2009, p. 36) afirma que:

Identificar as necessidades informacionais [...] dos grupos e indivíduos que integram a organização e de seus públicos externos é um passo fundamental para que possam ser desenvolvidos produtos informacionais orientados especificamente para cada grupo e necessidade, o que tende a ampliar significativamente tanto a necessidade de informação quanto a propensão dos usuários de aplicá-la em benefício da organização.

As necessidades informacionais variam substancialmente entre os variados grupos de usuários, sendo necessário conhecer e mapear cada uma das necessidades informacionais desses grupos a fim de que se possa direcionar a informação de modo a atender as exigências e expectativas de cada público alvo.

É necessário, portanto, conhecer os tipos de recursos informacionais existentes e direcioná-los aos grupos que apresentem comportamento informacional compatível.

No âmbito do arquivo corrente e intermediário, as atribuições inerentes a cada um dos cargos funcionais geram necessidades informacionais que precisam ser sanadas a fim de efetuar a realização de tarefas, a tomada de decisão, ou mesmo para compreender o ambiente organizacional em que se situa. Essas necessidades fazem o usuário interagir com o sistema de arquivos existentes na organização a fim de obter as orientações necessárias e executar os trâmites funcionais referentes ao cargo.

É interessante observar, neste contexto, que há um fluxo constante de informação tramitando no interior da instituição. Esse fluxo informacional é alimentado por usuários no exercício das atividades institucionais. As atividades trazem ao usuário demandas informacionais, ou seja, exigem a obtenção de informações institucionais que deverão ser sanadas e registradas em documentos para que seja possível dar continuidade ao fluxo.

As demandas advindas das atribuições de um cargo gerarão a necessidade informacional no usuário, que evoluirá do nível visceral à necessidade comprometida. Neste último estágio, o usuário consegue desenvolver uma demanda a ser direcionada a um sistema de informação arquivística e começa a lançar sobre este os questionamentos de busca.

Por meio da estrutura de classificação e dos instrumentos de recuperação da informação, o usuário consegue encontrar as fontes informacionais orgânicas, produzidas por outros usuários que possam complementar ou sanar a necessidade informacional.

De posse dessas informações, o usuário produzirá o registro documental das mesmas, a fim de atender às exigências das atividades institucionais que sejam de âmbito administrativo, fiscal, e jurídico.

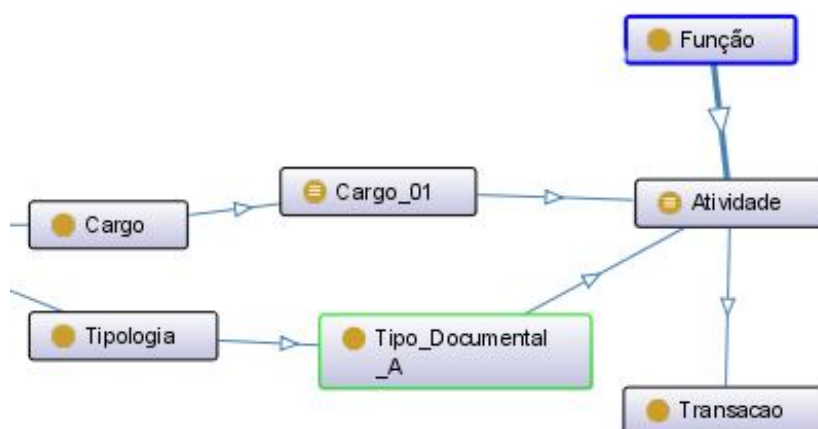
Esses registros deverão estar disponibilizados a outros usuários que, no exercício de suas atribuições, se apropriarão das informações e produzirão novos registros orgânicos a serem disponibilizados a outros usuários correlatos.

Verifica-se que, em um contexto informacional orgânico, o usuário apropria-se da informação e produz novos registros informacionais que atenderão ao comportamento informacional de outros usuários que, por sua vez, produzirão outros registros documentais. Isso torna visível o princípio da indivisibilidade, já mencionado em capítulos anteriores, em que um documento de arquivo reflete as relações orgânicas funcionais existentes em uma instituição e seu sentido só é completo quando os documentos são interpretados dentro do contexto de produção.

Tendo em consideração que os registros de informação orgânica são produzidos por um indivíduo ou corporação no exercício e em razão de funções e atividades; e que os documentos de arquivo, por conseguinte, apresentam forte relação com os trâmites que os produziu, é possível estabelecer um mapeamento dos recursos informacionais a partir da relação que documentos e usuários mantêm com o conjunto de funções e atividades inerentes a uma organização.

A figura a seguir apresenta uma relação existente entre função, atividade, tipo documental, e cargo.

Figura 08: Relacionamento entre função atividade e cargo.



Fonte: O autor.

A figura anterior apresenta o relacionamento orgânico e o possível mapeamento existente entre documentos e usuários. Em um primeiro nível, apresenta-se a função

institucional que deverá ser executada para ser alcançada a missão. Essa função agrupa um conjunto de procedimentos que são denominados atividades. Cada atividade terá um conjunto de transações que produzirão as tipologias documentais. Cada uma dessas atividades é executada por um grupo de usuários, que buscarão satisfazer suas necessidades informacionais nos tipos documentais inerentes às suas atribuições, ou ainda presentes em atividades correlatas.

O cargo exercido pelo indivíduo traz certa previsibilidade da interação que um determinado usuário constituirá com a informação orgânica, uma vez que delimitam a quantidade e a profundidade de interação frente ao corpo de processos funcionais de uma determinada instituição.

Com base no que já foi explanado, é possível afirmar que os registros tipológicos inerentes a uma atividade institucional específica deverá relacionar-se com os usuários que a exercem, ou ainda com usuários que exercem cargos em atividades correlatas.

Os relacionamentos existentes entre funções e atividades formam uma teia de fluxos informacionais que inter-relaciona documentos de arquivo, atividades e cargos funcionais.

Por esta razão, a gênese documental deverá ser tomada como base para a estruturação de um acervo arquivístico, já que permite identificar a atividade institucional que produz cada um dos tipos documentais da instituição. Uma vez estando os tipos documentais fortemente vinculados às atividades que os gerou, torna-se possível disponibilizar o conteúdo informacional para atividades associadas. Como um determinado cargo traz consigo forte associação com determinado grupo de atividades, torna-se possível o desenvolvimento de estratégias para que a informação orgânica seja pré-disponibilizada aos usuários correlatos.

Assim, é possível explanar sobre a relação envolta entre gênese documental e demandas que usuários possam fazer em relação aos registros de informação orgânica, no âmbito do gerenciamento informacional. É importante frisar esta questão, porque parecem existir algumas dificuldades no que tange às relações interdisciplinares envolvendo a Arquivologia e a Ciência da Informação.

A respeito dessa questão, Rodrigues (2006, p. 115):

Buscando-se entender essa dificuldade, aponta-se, inicialmente, que cada uma delas tem questões próprias, de fundo, a serem resolvidas. Por exemplo, o seu objeto. Mas isso não tem impedido que elas se desenvolvam enquanto áreas de conhecimento. Vale, então, verificar a principal diferença de enfoque para a gestão. Tal como foi exposto, a arquivologia parte do motivo da criação do documento para elaborar os instrumentos de acesso à informação. Já para ciência da informação, como se encontra em Marchiori (2002), o ponto de partida para a gestão da informação é a demanda (p. 75). Avaliar como essas duas perspectivas podem se integrar parece crucial para se promover a cooperação desejada.

A autora apresenta, portanto, diferentes focos de gestão que poderiam trazer algumas dificuldades entre a integração das áreas. Enquanto na Arquivologia o enfoque é dado à gênese documental, na Ciência da Informação o enfoque é feito a partir da demanda do usuário.

Surge, aparentemente, um embate a ser travado entre a gênese documental e a demanda informacional. Vale frisar que a gênese documental, compreendida como a geração do documento em uma atividade, é fundamental para estruturar um acervo documental orgânico; e a demanda informacional, por sua vez, faz com que a estrutura informacional do acervo seja voltada às necessidades dos usuários.

Em vez de abordar uma situação conflitante, é necessário apresentar uma relação integrada destes enfoques. A integração destas duas perspectivas é perfeitamente benéfica e necessária à gestão da informação orgânica, permitindo à integridade do acervo e satisfação do usuário. O enfoque voltado à gênese documental garante que os princípios arquivísticos de proveniência, organicidade e unicidade sejam respeitados, o que garante a integridade e autenticidade do acervo.

Por meio do enfoque voltado à gênese documental, o usuário poderá certificar a confiabilidade da representação que a documentação traz acerca dos trâmites envolvidos nos eventos funcionais que produziram o acervo documental. Permite, ainda, a produção de instrumentos de recuperação mais precisos para satisfazer a demanda de informações referentes aos níveis estratégico, tático e operacional já mencionados.

A demanda informacional, por sua vez, estará intimamente relacionada à estrutura de recuperação provida pela gênese, uma vez que o comportamento informacional se dará em torno das atribuições institucionais exercidas pelo usuário. Por esse motivo, a satisfação da necessidade do usuário estará diretamente relacionada à interação que os documentos de um determinado acervo apresentam com a atividade que os produziram.

A busca informacional em um arquivo deverá seguir a teia de relações funcionais, mesmo quando o usuário não interagir diretamente com as atribuições correlatas à informação desejada. Quando um professor universitário, por exemplo, desejar obter informações referentes ao seu adicional de férias, é bem provável que irá procurar sanar sua necessidade informacional junto aos documentos produzidos pelo departamento de gestão de pessoas, preferencialmente na atividade da folha de pagamento, ou ainda, em departamento que exercem atividades correlatas como o Financeiro. Dificilmente esse usuário iria procurar essa informação em documentos que versam a respeito dos procedimentos de manutenção e limpeza.

Isso ocorre porque o usuário, que recupera os documentos na vigência do valor primário, encontra pouca dificuldade em entender o contexto de produção, uma vez que já está inserido no mesmo. O usuário, portanto, consegue identificar a relação orgânica existente entre o registro informacional e a atividade que o produziu. A demanda informacional gerada pelo mesmo será direcionada, portanto, de acordo com a estrutura provida pela gênese documental.

Como o comportamento informacional dos usuários de arquivo corrente está intimamente relacionado às especificidades das atribuições dos cargos funcionais que exercem, torna-se perfeitamente possível a elaboração de estratégias para prever possíveis interações que o usuário possa fazer com os recursos informacionais e pré-disponibilizar mecanismos para a recuperação de informações que atendam o seu comportamento informacional.

5.2 O gerenciamento estratégico e a relevância da informação no ambiente corporativo

As organizações operam suas atribuições em uma atmosfera de extrema competitividade em que a sobrevivência está intimamente relacionada às vantagens competitivas que as mesmas venham conseguir no ambiente em que atuam.

O gerenciamento da informação torna-se crucial para o monitoramento do ambiente interno e externo a fim de que a instituição possa desenvolver suas estratégias. A informação orgânica em seu valor primário torna-se um elemento essencial, pois permite aos usuários do nível estratégico conhecer e monitorar o ambiente interno de atuação a fim de conduzir o andamento da organização.

Posteriormente, estes registros permitirão aos administradores a visualização do caminho percorrido pela organização ao longo do tempo, dotando-os de maior capacitação para a tomada de decisão mais segura.

Beal (2009, p. 75) afirma que:

A informação é um elemento essencial para a criação, implementação e avaliação de qualquer estratégia. Sem o acesso a informações adequadas a respeito das variáveis internas e do ambiente onde a informação se insere, os responsáveis pela elaboração da estratégia não têm como identificar os pontos fortes e fracos, as ameaças, as oportunidades, os valores corporativos e toda a variedade de fatores que devem ser considerados na identificação de alternativas e na tomada de decisões estratégicas. Com base nas informações coletadas sobre o ambiente interno e externo, a organização pode identificar alternativas e tomar decisões estratégicas para promover mudanças na estrutura e nos processos organizacionais, de modo a garantir a manutenção da sintonia com o ambiente externo e oferecer respostas adequadas para a sobrevivência e crescimento da organização.

A Informação Orgânica, que representa as variáveis do ambiente interno, é fundamental para que uma instituição possa identificar os problemas inseridos no ambiente institucional.

A informação orgânica exerce papel fundamental durante o planejamento estratégico. Ela representa um fator de apoio à decisão por reduzir a incerteza no processo decisório. Ela dota os usuários de maior capacitação para a execução dos procedimentos institucionais, além de permitir a integração das atividades exercidas em diferentes atividades. Atua, portanto, como fator de produção e sinergia. (BEAL, 2009, p. 21-22).

A autora apresenta alguns conceitos correlatos ao valor da informação no âmbito institucional. Apresenta sete leis que definem a relevância que a informação apresenta para os usuários do arquivo corrente, a saber:

- A informação é infinitamente compartilhável;
- O valor da informação aumenta com o uso;
- A informação é perecível;
- O valor da informação aumenta com a precisão;
- O valor da informação aumenta quando há combinação de informações;
- Mais informação não é necessariamente melhor; e
- A informação se multiplica. (BEAL, 2009, p. 22-28).

Antes da explanação de cada uma das leis, é necessário estabelecer algumas ponderações. O valor mencionado pela autora não está relacionado à valoração documental. Na valoração documental, a informação é analisada frente a sua relevância de valor de uso primário, pelos usuários da fase corrente e, posteriormente, pelo valor secundário, voltado aos usuários da fase permanente. Nesse contexto, entende-se que a informação orgânica está sendo abordada em relação ao comportamento informacional dos usuários da fase corrente, ou seja, ao produtor da documentação. Esses usuários geram e recuperam a informação arquivística para atender às demandas geradas por suas atribuições.

É necessário, portanto, elucidar que, ao mencionar o termo “valor”, a autora refere-se à relevância da informação frente aos processos administrativos que se apropriam dos registros informacionais. Este conceito de valor não está diretamente relacionado ao valor primário e secundário da informação orgânica.

Em relação à primeira lei, Beal afirma que a informação pode ser compartilhada e utilizada por infinitas vezes por um grande número de pessoas sem que seja consumida, ou seja, sem que sofra qualquer tipo de desgaste. No âmbito da informação orgânica, os usuários

geram informações arquivísticas em registros documentais e compartilham com usuários de atribuições semelhantes, a fim de executar suas atribuições. O compartilhamento da informação evita a criação de registros em duplicidade, que é custoso à instituição. A definição do fluxo informacional dos tipos documentais pode contribuir para que vários usuários possam usufruir do compartilhamento de informações.

Em relação à segunda, a autora enfatiza a necessidade que o usuário:

- tenha ciência da existência da informação que atenda o seu comportamento;
- saiba onde está armazenada;
- tenha acesso;
- saiba se apropriar e direcionar o uso de forma correta.

Consequentemente, o valor, ou seja, a relevância da informação frente às necessidades do usuário será ampliada com o uso. Quanto mais os usuários conhecerem e apropriarem-se deste, maior relevância a informação terá para os mesmos.

É importante frisar a responsabilidade dos profissionais de arquivo no desenvolvimento de mecanismos eficientes na recuperação da informação orgânica. Esse é um dos feitos de onde emana o reconhecimento dos usuários face à importância dos registros documentais, bem como dos benefícios das práticas arquivísticas nas atribuições dos usuários.

Em relação à terceira lei, Beal defende a ideia de que a informação é perecível, ou seja, perde o seu valor potencial à medida que o tempo passa. A autora exemplifica a importância da tempestividade da informação, ou seja, a disponibilização da informação necessária no exato momento em que for necessária. Com o passar do tempo, esse valor diminui.

Neste trecho é possível estabelecer uma relação com a prescrição do valor primário. A informação orgânica, registrada em documentos de arquivo, tende a perder a necessidade de uso por parte dos usuários da administração no decorrer do tempo. Na fase de guarda intermediária, os registros prescrevem o valor primário. É importante lembrar que muitos dos registros poderão adquirir valor secundário que é de caráter imprescritível. Esses documentos serão submetidos à guarda permanente. Esses registros não mais serão apropriados pelos usuários da fase corrente nas finalidades em que foram gerados, mas poderá ainda servir aos administradores para identificar os trâmites envolvidos na administração. (ROUSSEAU; COUTURE, 1998).

No que diz respeito à quarta lei, a autora comenta que grau de precisão determina a relevância da informação ao cumprimento das atribuições funcionais. Informações inexatas

atrapalham o andamento dos procedimentos administrativos podendo gerar prejuízos de diversas naturezas.

Ao explicar a quinta lei, Beal afirma que a informação quando devidamente contextualizada e integrada a outras informações passa a ser considerada dentro uma visão sistêmica de um processo. É possível observar que as informações registradas em cada tipologia documental se inter-relacionam, permitindo uma visão integrada de todo o acervo, ao mesmo tempo em que se resguarda a autenticidade dos documentos por meio da visualização do sincronismo das informações. Isso traz, por consequência, maior vinculação entre os procedimentos institucionais e ressalta a indivisibilidade do acervo como um todo, já que os documentos que registram a informação tornam este contexto de produção indivisível.

Em explanação à sexta lei, a autora afirma que o excesso de informação poderá também ser prejudicial aos usuários. Por esse motivo, as informações precisam ser filtradas em relação aos critérios de relevância para o usuário. O filtro informacional, portanto, deverá seguir as especificidades do comportamento informacional inerente às atribuições de cada um dos cargos funcionais exercidos no contexto institucional. Assim, os níveis informacionais operacional, tático e estratégico, já mencionados, podem prover parâmetros para a filtragem da informação.

Em relação à sétima lei, Beal comenta que, enquanto muitos recursos são esgotados com o uso, a informação se multiplica. Conforme já explanado, os usuários de arquivo multiplicam as informações em novos documentos após entrarem em contato com outros registros informacionais. Dessa forma, os novos conjuntos de informação orgânica dão embasamento contextual às informações registradas anteriormente, e vice-versa, solidificando a integridade do acervo.

A informação orgânica transita em um fluxo informacional da organização, atendendo aos procedimentos sistêmicos executados no órgão. Oliveira (2009, p. 31) comenta que “os fluxos de informação podem ser compreendidos como um sistema de canalização pelo qual fluem todas as informações de um sistema”.

5.3 O fluxo de informação na fase corrente e intermediária

Beal (2009, p. 30) comenta que na representação de um fluxo organizacional

[...] a atividade de identificação de necessidades e requisitos de informação age como elemento acionador do processo que pode estabelecer um ciclo contínuo de coleta, tratamento, distribuição/armazenamento e uso para alimentar os processos decisórios e/ou operacionais da organização, e leva também a oferta de informações para o ambiente externo.

Para prover o correto fluxo informacional é necessário integrar alguns elementos, dentre eles:

- Identificação de Necessidades e Requisitos;
- Obtenção;
- Tratamento;
- Distribuição;
- Uso;
- Armazenamento; e
- Descarte. (BEAL, 2009, p. 30-31).

A identificação de necessidades e pré-requisitos, conforme já trabalhado neste capítulo, diz respeito à compreensão do comportamento informacional do usuário. A partir da identificação das necessidades informacionais dos diferentes grupos de usuários institucionais é possível identificar as fontes documentais mais relevantes para cada um dos grupos. A necessidade informacional, como já mencionado, pode ser mapeada a partir do relacionamento que um determinado cargo exerça com o conjunto de funções institucionais.

A etapa de obtenção configura-se a partir da identificação de necessidades. É possível mapear as fontes documentais de informação que possam suprir o comportamento informacional e disponibilizá-las ao usuário.

Uma vez identificadas às fontes informacionais que atendam o comportamento dos usuários, é necessário submetê-las a um tratamento, que se configura em procedimentos de organização da informação, a fim de que sua relevância seja percebida.

Para tanto, são feitas análises, sínteses e apresentação. O tratamento engloba, ainda, procedimentos de classificação quanto à segurança relacionada à natureza do sigilo, bem como quanto à temática referente aos procedimentos que poderão ser referenciados ou representados.

A partir do mapeamento das necessidades informacionais e das fontes documentais relevantes, torna-se possível prover meios à satisfação do comportamento do usuário. A classificação funcional, provida na gestão documental, propicia maior clareza e precisão na inter-relação existente entre o corpo de funções e o comportamento informacional.

Para que o usuário possa apropriar-se da informação, é necessário evoluir para a etapa da distribuição. Esse processo permite o direcionamento dos recursos informacionais para aqueles que deles necessitam. Neste momento, são desenvolvidos instrumentos para a recuperação da informação, tais como instrumentos de pesquisa, tesauros, planos de

classificação e motores de busca. Essas ações permitem desenvolver meios que viabilizem tanto a divulgação de informações inerentes ao comportamento informacional do usuário, como o atendimento das demandas que os mesmos possam solicitar em mecanismos de busca. A Web Semântica, como será demonstrada em capítulos posteriores, parece prover excelentes recursos para que a distribuição de informação seja ampliada nos ambientes corporativos.

Uma vez que haja a ciência por parte do usuário de uma informação concernente ao seu comportamento informacional, o mesmo poderá dela apropriar-se e desenvolver suas atribuições. Embasando-se no uso dos recursos informacionais, o usuário produzirá outros documentos arquivísticos e os inserirá no sistema de informação orgânica.

Depois de produzidos novos registros documentais, os mesmos deverão ser armazenados para que possam ser recuperados no futuro. O armazenamento engloba o desenvolvimento de estratégias para que a documentação possa ser acondicionada corretamente e salvaguardada de riscos de deterioração, obsolescência tecnológica, bem como pela destruição causada por desastres naturais ou ainda por agentes químicos, físicos e biológicos.

Por fim, muitos dos registros informacionais armazenados sofrerão a prescrição de valor primário sem a adição do valor secundário. Nesse contexto, procedimentos de avaliação documental entram em cena para desenvolver tabelas de temporalidade documental que definam os prazos para a prescrição do valor primário e promovam a destinação dos documentos ao descarte ou a guarda permanente. Documentos que estejam desprovidos de ambos os valores serão vistos como irrelevantes aos interesses da instituição e serão descartados. Os demais serão recolhidos à guarda permanente, onde os usuários da fase corrente não exercem mais influência no contexto de produção documental.

Estes são, portanto, os elementos envoltos no fluxo informacional que percorre os sistemas de uma determinada instituição. É importante ressaltar o papel exercido pela informação orgânica na orientação e condução dos usuários no exercício de suas atribuições.

5.4 Considerações

A delineação do presente capítulo demonstrou que o comportamento do usuário consiste nas necessidades de informação que os mesmos apresentam frente a um determinado acervo documental. Neste âmbito, o cargo exercido pelos usuários apresentam vínculos significativos com determinadas funções, atividades e tipos documentais, que representam meios de mapeamento do comportamento informacional. A interação desmembra-se em níveis que definem atuação de nível operacional, tático ou estratégico.

Verifica-se que a informação orgânica deve ser apropriada de forma estratégica para que a instituição produtora possa se desenvolver em seu ambiente de atuação. Isso é alcançado por meio do desenvolvimento de um fluxo informacional que contemple o tratamento, distribuição, uso, armazenamento, eliminação ou guarda permanente dos recursos informacionais envolvidos em cada atribuição institucional.

Surge a necessidade da abordagem de programas de tratamento documental que formalizem os procedimentos mencionados neste capítulo, dotando-os de metodologia.

A elaboração de programas de gestão documental, a ser analisada no próximo capítulo, permite a aplicação do tratamento arquivístico junto ao fluxo de informações da instituição visando à salvaguarda dos registros de valor secundário para a guarda permanente, o descarte de documentos desprovidos de qualquer valoração no momento oportuno, e o desenvolvimento de mecanismos que permitam a satisfação do comportamento informacional do usuário trabalhado neste capítulo.

6 A GESTÃO DE DOCUMENTOS E DA INFORMAÇÃO ORGÂNICA

O capítulo anterior trouxe uma compreensão do comportamento informacional do usuário, da relevância da informação orgânica dentro do atual contexto organizacional e apresentou as atividades envolvidas no gerenciamento do fluxo informacional, tais como o uso, o armazenamento e a eliminação de documentos e informações.

O presente capítulo abordará a gestão de documentos e sua interação com o fluxo informacional de uma organização.

6.1 Gestão Documental: Definição e Elementos

Por gestão de documentos compreende-se “conjunto de procedimentos e operações técnicas referentes à produção, tramitação, uso, avaliação e arquivamento de documentos em fase corrente e intermediária, visando sua eliminação ou recolhimento”. (ARQUIVO NACIONAL, 2005, p. 100).

É possível observar nessa definição que o fluxo informacional e a gestão documental se integram para prover a satisfação das necessidades informacionais que os usuários apresentam na fase corrente do ciclo de vida documental.

Todos os procedimentos mencionados na definição da gestão documental coincidem com os elementos apresentados no fluxo informacional, descrito no capítulo anterior. É interessante acrescentar neste conjunto de procedimentos a classificação documental que é fundamental para que a informação possa ser contextualizada e recuperada.

A gestão documental é um processo eficiente na integração do conteúdo, contexto e estrutura documentais, já apresentados no primeiro capítulo, às necessidades dos usuários. O gerenciamento de documentos permite a consolidação das práticas arquivísticas em benefício dos administradores que necessitam da informação para a correta tomada de decisão.

Os elementos existentes na gestão documental serão apresentados nos parágrafos seguintes.

6.1.1 Produção documental

Santos, V. (2007, p. 178) afirma que a produção documental

Contempla os procedimentos relacionados à manutenção do maior rigor possível na produção dos documentos de arquivo, abrangendo definição de normas, conteúdo, modelos, formato e trâmite; [...] a execução adequada dessa função demanda um conhecimento profundo da instituição, seus objetivos e missão, as tecnologias disponíveis e os tipos adequados ao exercício do negócio da instituição.

Por produção documental compreende-se a criação do documento, ou seja, a gênese documental seguida da captura do registro em um sistema de gerenciamento documental. O CONARQ (2009, p. 32) afirma que a captura

[...] é a incorporação de um documento ao sistema de gestão arquivística, quando ele passa a seguir as rotinas de tramitação e arquivamento. Uma vez capturado, o documento pode ser incluído num fluxo de trabalho e, posteriormente, arquivado, ou ser, imediatamente, arquivado em uma pasta, no caso de documentos em papel, ou diretório, no caso de documentos digitais.

A produção antecede a tramitação e permite identificar os procedimentos que geram a documentação, servindo de base para a definição dos requisitos a serem cumpridos para o gerenciamento da documentação e dá às diretrizes a classificação.

6.1.2 Tramitação

Uma vez produzido e capturado, o documento será submetido a um fluxo informacional a fim de que possa cumprir com as finalidades para as quais foi criado e satisfazer o comportamento informacional dos usuários.

A tramitação está relacionada, portanto, ao percurso que o documento transitará nos procedimentos institucionais para que possam dar-lhes suporte e andamento. À medida que o documento tramita, informações vão sendo produzidas pelos usuários em novos registros, peculiares a cada uma das atividades correlacionadas. A geração de informações contextualizadas em outros documentos forma uma teia informacional orgânica que dará autenticidade aos atos representados no acervo documental.

6.1.3 Uso

É necessário prover os meios necessários para que os usuários se apropriem da informação e produzam documentos peculiares às suas atribuições para dar prosseguimento aos trâmites. Por esse motivo, deve-se trabalhar o uso que será feito destes registros identificando quem, onde, quando e porque é feita a apropriação da informação arquivística em um determinado tipo documental para prover o acesso aos usuários que deles necessitem para dar andamentos em suas atividades.

O relacionamento entre os procedimentos envoltos na tramitação permite identificar quais atividades produzem um determinado tipo documental e quais atividades são correlatas. A partir desse relacionamento é possível identificar os grupos de usuários potenciais à apropriação de determinados tipos documentais.

6.1.4 Classificação

Um dos principais pontos a serem trabalhados na gestão de documentos é a classificação documental. Essa é a base para a eficiência de todo o conjunto de elementos do programa de gestão documental, já que por meio dela será possível prover acesso, avaliação e a identificação dos procedimentos que produzem os tipos documentais.

Gonçalves (1998, p. 12) comenta que:

O objetivo da classificação é, basicamente, dar visibilidade às funções e às atividades do organismo produtor do arquivo, deixando claras as ligações entre os documentos. Podemos entender que a classificação é, antes de tudo, lógica: a partir da análise do organismo produtor de documentos de arquivo, são criadas categorias, classes genéricas, que dizem respeito às funções/atividades detectadas. [...] A classificação é geralmente traduzida em esquema no qual a hierarquia entre as classes e subclasses aparece representada espacialmente.

Schellenberg (2006, p. 83) afirma que uma das maiores preocupações das instituições que atuam com a documentação em sua fase corrente é que os documentos sejam organizados para que o usuário possa obter a recuperação imediata da informação quando isso se fizer necessário. Para que isso possa ocorrer, os documentos deverão estar bem classificados e armazenados. Afirma, então que:

A classificação é básica à eficiente administração de documentos correntes. Todos os outros aspectos de um programa que vise ao controle de documentos dependem da classificação. Se os documentos são adequadamente classificados, atenderão bem as necessidades das operações correntes. E, para tanto, devem ser arranjados em função do uso que têm em determinadas unidades administrativas de um órgão. Em todos os casos, desde os relativos a importantes assuntos de política até os de operações de rotina deverão agrupados em relação ao seu uso. Refletirão a função do órgão, no amplo sentido do termo, e, no sentido mais restrito, as operações específicas individuais que integram as atividades do mesmo órgão.

Devido à relevância apresentada para a gestão documental, o foco de abordagem do gerenciamento de documentos nesse capítulo estará voltado à classificação documental.

Afirma Schellenberg (2006, p. 84) que a classificação poderá ser vinculada às funções, à estrutura departamental, ou ainda por assuntos.

Recomenda o autor, no entanto, que a classificação por assuntos seja aplicada apenas em casos muito específicos, haja vista que esse modelo possa desfigurar toda a organicidade envolta nos documentos.

Desses três modelos, o mais recomendável é, sem sombra de dúvidas, a classificação funcional, já que ela permite identificar a relação que os documentos têm com os elementos funcionais e, a partir disso, interpretar o contexto de produção documental. Embora esse seja o modelo de classificação mais eficiente à contextualização do acervo, a classificação por funções é uma das mais difíceis de produzir e requer procedimentos especializados de análise

do processo, para que possa prover representação mais expressiva dos trâmites envolvidos no fluxo informacional.

Souza (2007, p. 138-139) apresenta as seguintes definições para função e atividade:

Função, portanto, corresponde a uma agregação de atividades análogas, interdependentes que se encadeiam num único campo especializado de trabalho, devendo, conseqüentemente, para uma maior eficácia, ter o desenvolvimento alocado à unidade organizacional específica. [...] Ela tem, geralmente, caráter duradouro e sem término previsto e provê o suporte para a viabilização da missão da organização. A função indica o que é feito sem detalhar como é feito [...] A atividade compreende um conjunto de tarefas/ações caracterizadas pelo consumo de recursos e orientadas para um objetivo definido, de caráter global. Em resumo, é um conjunto de procedimentos necessários para a execução de uma função.

A classificação por função, embora seja mais trabalhosa, traz a melhor base para o desenvolvimento de ferramentas que promovam a recuperação da informação e a satisfação do usuário, bem como a contextualização de toda a produção documental.

A classificação estrutural, por sua vez, consiste na organização das classes documentais pelos departamentos que formam a estrutura da organização. Esse tipo de classificação pode não ser muito eficiente na fase corrente de um acervo, haja vista que a estrutura departamental está sempre sofrendo alterações. No decorrer da existência da organização, departamentos são criados, extintos, divididos. As funções exercidas pelos mesmos podem ser adicionadas ou transferidas para outros departamentos.

A classificação estrutural pode ser recomendada para prover o arranjo, ou seja, a classificação de um acervo de guarda permanente em que a instituição e os usuários que dela participaram não mais existem, e cuja documentação não tenha recebido nenhum tipo de tratamento arquivístico na fase corrente. Contudo, torna-se muito difícil a identificação imediata dos trâmites envolvidos em cada uma das atividades. Identificar a nomenclatura e os procedimentos funcionais, executados por uma organização cuja existência vincula-se ao passado, torna-se um desafio praticamente intransponível.

A classificação estrutural desempenha um papel importante para que a documentação seja contextualizada a partir dos departamentos que a produziu, permitindo a organização e descrição do acervo. Como a instituição não tornará a existir, não haverá mudanças na estrutura do fundo arquivístico, configurando-o como fundo fechado. Nesse caso, a classificação estrutural não sofrerá constantes alterações.

Essas são, portanto, as formas de classificação encontradas em um acervo arquivístico.

6.1.5 Avaliação

Uma vez terminada a intensidade do uso corrente, a documentação será enviada para a fase intermediária a fim de que se estabeleçam os prazos de prescrição do valor primário e a verificação de aquisição do valor secundário. Caso fique constatada a inexistência do valor secundário, o documento será então eliminado, haja vista não haver mais a necessidade de suprir o comportamento informacional existente no valor primário, nem ser relevante aos usuários que consultam os acervos na vigência do valor secundário.

A definição dos prazos de guarda é vista pelo procedimento de avaliação, que identifica os prazos de prescrição e elabora a tabela de temporalidade documental.

6.1.6 Arquivamento

O arquivamento, por sua vez, está relacionado aos procedimentos de guarda documental, a fim de que o documento possa ser localizado sempre que solicitado. Procedimentos que garantam a conservação dos suportes documentais são executados para garantir a integridade dos documentos e evitar a perda da informação.

Estes elementos combinados permitem a compreensão dos procedimentos a serem trabalhados na gestão documental a fim de integrar a informação orgânica acessível aos usuários da fase corrente.

6.2 O programa de Gestão Documental

A gestão de documentos é realizada por meio de um Programa de Gestão Documental (PGD). O PGD desenvolve ferramentas regulamentadoras que dirigem os procedimentos envoltos no gerenciamento documental.

Por meio do PGD, ocorre o desenvolvimento de ferramentas de classificação, para a recuperação da informação; de avaliação, que é a tabela de temporalidade voltada à destinação dos registros documentais, bem como de políticas e procedimentos para o trâmite, uso, e o arquivamento da documentação.

A gestão documental provê, portanto, os recursos necessários à gestão da informação orgânica registrada nos documentos arquivísticos.

A norma ISO 15489-1 apresenta oito estágios integrados para o desenvolvimento de um programa de gestão documental, que serão apresentados a seguir:

- a) Levantamento preliminar;

- b) Análise das funções, das atividades desenvolvidas e dos documentos produzidos;
- c) Identificação das exigências a serem cumpridas para a produção de documentos;
- d) Avaliação dos sistemas existentes;
- e) Identificação das estratégias para satisfazer as exigências a serem cumpridas para a produção de documentos arquivísticos;
- f) Projeto do sistema de gestão arquivística de documentos;
- g) Implementação do sistema de gestão arquivística de documentos;
- h) Monitoramento e ajustes. (ARQUIVO NACIONAL, 2009; BANTIN, 2008; NATIONAL ARCHIVES OF AUSTRALIA, 2001)

Essas etapas conduzem a elaboração de um PGD desde o diagnóstico, que corresponde às etapas A –D, até as revisões a serem trabalhadas após a implementação. As etapas serão apresentadas a seguir:

6.2.1 Etapa A – Levantamento Preliminar

O levantamento preliminar visa à investigação do contexto jurídico e administrativo em que a instituição opera, buscando identificar o papel desempenhado pelo organismo, a regulamentação que norteia suas atividades, os pontos fortes e fracos do sistema de gerenciamento de documentos, bem como os principais *stakeholders* envolvidos na instituição. Por *stakeholder*, compreende-se serem os indivíduos ou as organizações que podem influenciar ou ser influenciados pela instituição. Os usuários do arquivo corrente são um exemplo de *stakeholder*. (NATIONAL ARCHIVES OF AUSTRALIA, 2001)

A finalização do levantamento preliminar é conduzida por um relatório em que se registram as informações investigadas o que inclui as principais fontes documentais encontradas na instituição, os riscos envolvidos nas atividades, pontos fortes e fracos da gestão documental, principais *stakeholders*, e a atuação do órgão.

6.2.2 Etapa B – Análise de Funções e Atividades

A análise das funções e atividades institucionais pode ser considerada uma das etapas mais importantes do PGD pois, a partir dela, será elaborada a classificação dos documentos de arquivo pelo método funcional.

O ARQUIVO NACIONAL (2009, p. 27) afirma que a análise:

Consiste em identificar, documentar e classificar cada função e atividade, bem como em identificar e documentar os fluxos de trabalho e os documentos produzidos. O objetivo deste passo é desenvolver um modelo conceitual sobre o que o órgão ou entidade faz e como faz, demonstrando como os documentos se relacionam com sua missão e suas atividades. O modelo subsidiará a definição dos procedimentos de produção, captura, controle, armazenamento, acesso e destinação dos documentos. Essa definição é particularmente importante em ambientes eletrônicos, onde os documentos adequados não são capturados e mantidos se o sistema não for projetado para isso.

A análise das funções e atividades existentes na instituição permite a obtenção de um mapeamento dos procedimentos executados pela entidade e os relacionamentos existentes entre os mesmos. É possível, portanto, relacionar cada função com suas respectivas atividades e cada uma das atividades com as tipologias documentais produzidas.

Neste ponto identifica-se a gênese documental, ou seja, as finalidades para as quais são produzidos cada um dos documentos e a identificação da atividade produtora de cada um dos tipos documentais. Esta fase define a classificação documental e traz a base necessária ao desenvolvimento de tabelas de temporalidade e instrumentos para indexação e recuperação da informação orgânica. Ao término dessa etapa é possível desenvolver o plano de classificação, elaborar tesouros e obter a estrutura da tabela de temporalidade documental.

6.2.3 Etapa C – Identificação das Exigências

A identificação das exigências está relacionada à compreensão dos requisitos necessários para que a documentação possa prover evidência aos eventos ocorridos em cada uma das atividades institucionais. Esses requisitos determinam quais documentos deverão ser produzidos, qual forma documental deverá conter e quem poderá produzir cada documento. Os requisitos são embasados na legislação vigente, nos documentos internos, na expectativa da sociedade, e no risco envolto em negligenciar a documentação. (ARQUIVO NACIONAL, 2009; NATIONAL ARCHIVES OF AUSTRALIA, 2001).

6.2.4 Etapa D – Avaliação dos Sistemas Existentes

Uma vez identificados os requisitos necessários ao gerenciamento documental, passa-se à avaliação dos sistemas de informação existentes, onde se verifica a ocorrência do cumprimento dos requisitos. Os resultados da investigação trarão os elementos a serem trabalhados no desenvolvimento de estratégias para sanar quaisquer problemas encontrados.

6.2.5 Etapa E – Identificação de estratégias

Uma vez analisado cada um dos sistemas existentes, passa-se ao desenvolvimento de estratégias que permitam à gestão documental o cumprimento dos requisitos encontrados na Etapa C e sanar quaisquer problemas encontrados na Etapa D. As estratégias consistem no desenvolvimento de políticas, padrões e procedimentos para a implantação de um sistema de gestão documental.

6.2.6 Etapa F – Projeto do Sistema de Gestão Arquivística

A partir das estratégias desenvolvidas, é elaborado o projeto do sistema de gestão documental que possa cumprir as exigências envolvidas em cada uma das funções e atividades identificadas nas etapas anteriores.

Neste momento, são planeadas as mudanças que serão implantadas nas praticas arquivísticas da instituição, bem como a integração dos sistemas e processos e existentes e a implementação das políticas de segurança institucionais.

Nessa fase também se elabora especificações detalhadas dos componentes de *hardware* e *software* envolvidos na aquisição de um sistema digital, bem como projeto de prédios para o armazenamento do acervo físico.

São também trabalhados modelos que, com os processos envolvidos no sistema, a previsão de treinamento de pessoal e um planeamento de implementação do sistema.

6.2.7 Etapa G – Implementação do Sistema

Uma vez projetado, o sistema será então implementado. Essa etapa consiste na integração do sistema de gestão arquivística com os outros sistemas informacionais existentes na organização, no treinamento de pessoal e na regulamentação das políticas desenvolvidas na Etapa E, por meio de normas e manuais.

6.2.8 Etapa H – Monitoramento e Ajustes

Uma vez implementado o sistema de arquivos, ocorre o monitoramento constante do sistema, a fim de possa ser avaliado o seu desempenho e se desenvolvam melhorias. O procedimento consiste no desenvolvimento de entrevistas com a administração, aplicação de questionários e na análise da documentação produzida na implementação do sistema. A partir dos dados levantados são produzidos os ajustes necessários a fim que o sistema possa acompanhar a evolução existente na instituição.

Essas etapas são, portanto, fundamentais à implementação de qualquer programa de gestão documental que vise satisfazer o comportamento informacional dos usuários de arquivo, já que regulamenta um sistema com procedimentos voltados ao fluxo informacional provido por cada uma das funções e atividades do organismo.

6.3 Ferramentas da classificação: Instrumentos para a gestão estratégica da informação.

Uma vez apresentados os procedimentos envoltos na elaboração de um sistema de gestão documental, apresentar-se-á uma rápida explanação das ferramentas de classificação. O *National Archives of Australia* (2003) apresenta duas ferramentas resultantes da classificação documental, a saber: Tesouro de Funções, e Plano de Classificação.

Ambos os instrumentos derivam da análise de funções e atividades institucionais que traz os vínculos existentes entre os procedimentos institucionais e os documentos de arquivo produzidos.

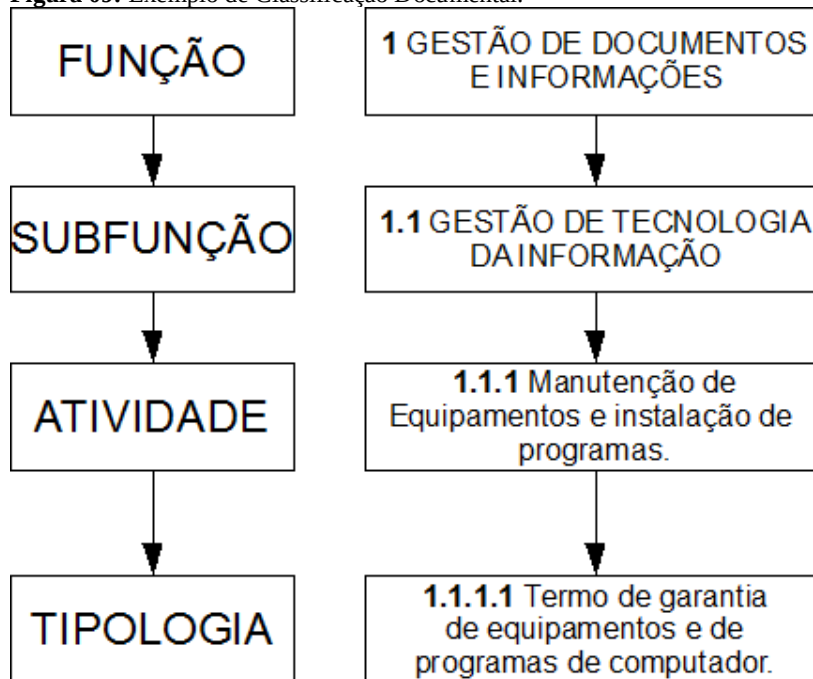
A estrutura provida pela classificação funcional representa, portanto, as relações orgânicas entre os documentos e os procedimentos produtores permitindo a completa contextualização do acervo. Essa vinculação permite a integração entre conteúdo, contexto e estrutura documentais em torno das necessidades informacionais dos usuários.

Tendo em consideração que os usuários da fase corrente recuperam a informação a partir das necessidades geradas por suas atribuições, é possível compreender que as ferramentas da classificação proveem os meios necessários para que os usuários possam acessar rápida e precisamente as informações de que necessitam.

O Plano de Classificação é um instrumento que apresenta as relações hierárquicas existentes entre os níveis de Função, Subfunção, Atividade, e Tipo Documental. Por meio dele é possível atribuir ao documento uma codificação que o identifica e contextualiza frente ao acervo. (NATIONAL ARCHIVES OF AUSTRALIA, 2003; SOUZA, 2007)

O plano de classificação consiste, portanto, em uma taxonomia que provê estruturas de classes, relacionadas hierarquicamente, essenciais à elaboração das tabelas de temporalidade documental e à recuperação da informação orgânica contida no acervo, conforme apresentado na figura a seguir:

Figura 09: Exemplo de Classificação Documental.



Fonte: Ministério Público do Estado de São Paulo (2010, p. 18).

Como é possível observar na figura cada um dos elementos representam uma classe. Função é a classe de maior hierarquia, podendo ser dividida em subclasses de funções, cada uma das subfunções são divididas em classes de atividades que, por sua vez, abrigarão a classe das tipologias.

O Arquivo Nacional (2005, p. 132) define plano de classificação como:

Esquema de distribuição de documentos em classes, de acordo com métodos de arquivamento específicos, elaborado a partir do estudo das estruturas e funções de uma instituição e da análise do arquivo por ela produzido.

O plano de classificação, portanto, permite uma ampla visão dos relacionamentos existentes entre os elementos funcionais com as tipologias documentais.

O plano de classificação é eficiente na representação das relações existentes entre atividades e dos documentos, sendo essenciais para qualquer PGD.

A recuperação da informação, por sua vez, traz a necessidade de se formalizar uma ampla estrutura de termos indexadores, a fim de que os usuários possam encontrar as informações registradas nos tipos documentais. O plano de classificação poderá não desenvolver grande precisão na recuperação da informação, uma vez que utiliza terminologia específica para prover a localização de uma determinada classe documental. É necessário, portanto, desenvolver ferramentas que possam aprimorar o processo de indexação documental a fim de melhor atender o comportamento informacional dos usuários.

Da classificação deriva, então, o tesouro de funções. Esta ferramenta apresenta uma lista de termos vinculados entre si por relações semânticas, hierárquicas, associativas ou de equivalência. Vocabulários controlados, de natureza técnica, podem ser associados a termos familiares aos usuários para uma recuperação mais precisa.

Neste instrumento, os termos são apresentados em ordem alfabética e descrevem relações existentes com outros termos, permitindo maior precisão na classificação, indexação e recuperação de informações e documentos orgânicos.

Santos, V. (2007, p. 195) apresenta o tesouro como uma ferramenta voltada à gestão da informação, visando trazer suporte aos procedimentos internos para assegurar a qualidade das operações executadas no órgão e para trazer eficiência e eficácia à administração.

O tesouro de funções será abordado mais detalhadamente no próximo capítulo, onde se desenvolverá a integração dos termos indexadores nele contidos com as linguagens da web semântica para instrumentalizar a automação da recuperação da informação orgânica em sistemas digitais de gerenciamento de documentos.

As ferramentas da classificação documental, portanto, são formas eficientes de prover o gerenciamento da informação orgânica em âmbito institucional.

Santos, V. (2007, p. 177) comenta que:

Com a valorização da informação como recurso para tomada de decisão e como ativo das instituições, o papel da unidade de arquivo pode passar a ser o de fonte de informações administrativas e técnicas e, em consequência, o arquivista que atua na gestão de documentos deve tornar-se um provedor de informações para a tomada de decisões.

A gestão documental é, portanto, indispensável ao gerenciamento estratégico da informação, pois provê métodos que permitem o tratamento do fluxo informacional envolvidos nos trâmites administrativos. Desenvolve, ainda, ferramentas essenciais ao mapeamento das informações e à satisfação comportamento informacional do usuário de arquivos correntes, provendo a recuperação precisa dos registros informacionais.

6.4 Considerações

A gestão documental promove, integra, e realça os relacionamentos existentes entre o contexto de produção, conteúdo e estrutura documental, envolvidos na tramitação da informação orgânica registrada nos documentos de arquivo. É um instrumento mantenedor do fluxo informacional indispensável ao gerenciamento estratégico da informação voltado à tomada de decisões.

Por meio das ferramentas de classificação e das políticas de gestão os usuários podem prover controle dos documentos e interagir com as informações ali existentes, visando ao cumprimento de suas atribuições para melhor condução da missão institucional. Integrar as ferramentas de classificação documental às linguagens de marcação, conforme será demonstrado no capítulo seguinte, poderá prover agilidade na recuperação da informação orgânica e melhor desempenho ao gerenciamento estratégico informacional.

A gestão de documentos permite, portanto, a integração das ações de administradores e profissionais da informação rumo à otimização do desempenho institucional.

7 AS CONTRIBUIÇÕES DAS LINGUAGENS DE MARCAÇÃO PARA A GESTÃO DOCUMENTAL: REPRESENTAÇÃO DE PLANOS DE CLASSIFICAÇÃO E TESAUSOS DE FUNÇÃO

Os capítulos anteriores versaram a respeito do comportamento informacional dos usuários de arquivo corrente e das demandas informacionais solicitadas ao sistema. Esclareceu-se que, para melhor atender ao comportamento informacional dos usuários, é necessário trabalhar a documentação dentro de um fluxo informacional que cuidará da tramitação da documentação no acervo e dos procedimentos arquivísticos a serem executados. Por consequência, a documentação será armazenada de modo a ser localizada quando necessário, assegurando-se a recuperação da informação orgânica ali registrada. Observou-se a importância que a gestão documental apresenta na consolidação do fluxo informacional, a fim de assegurar a construção de mecanismos que disponibilizem a informação orgânica registradas nos documentos de arquivo aos usuários que dela necessitar em um momento oportuno. Observou-se que o procedimento de classificação traz as diretrizes necessárias para o desenvolvimento de ferramentas de classificação, dentre elas o Tesauro de Funções.

O presente capítulo aprofundará na descrição desta ferramenta, visando integrá-la às linguagens de marcação utilizadas na web semântica, a fim de apresentar benefícios que as linguagens de marcação possam trazer à gestão da informação orgânica e dos documentos de arquivo durante a fase corrente do ciclo de vida documental.

7.1 A recuperação da informação na fase corrente da guarda documental

Para que a gestão de informação e de documentos de arquivo possa ser eficiente, é necessário prover a recuperação da informação orgânica pelos usuários do acervo documental.

Araújo Júnior (2007, p. 65) afirma que:

O processo de busca e recuperação da informação pode ser conceituado como o processo de localizar documentos e itens de informação que tenham sido objeto de armazenamento, com a finalidade de permitir o acesso dos usuários aos itens de informação, objetos de uma solicitação. A recuperação da informação se dá pela comparação do que se solicitou com o que está armazenado, bem como com o conjunto de procedimentos que este processo envolve.

Menezes, Cunha e Heeman (2004, p. 50) complementam a definição de recuperação da informação afirmando que esta consiste em:

Forma de obtenção seletiva de dados mediante a aplicação de argumentos de busca destinados a identificar os registros que cumprem uma determinada condição. A recuperação requer um meio de armazenamento, um dispositivo lógico destinado a comparar os dados que se deseja recuperar, com os dados armazenados. Requer

ainda métodos e procedimentos para representar o conhecimento e a formulação das estratégias de busca.

O processo de busca e recuperação da informação norteia todo o fluxo de informação, sendo de fundamental importância para a gestão de documentos. Os registros documentais deverão ser localizados no momento oportuno para que uma determinada instituição possa cumprir com suas atribuições e prover a prova necessária da execução dos procedimentos.

É necessário, portanto, que os usuários disponham de ferramentas que lhes permitam a recuperação da informação sempre que necessário. A esse respeito, Cervantes (2006, p. 12) comenta que:

Cabe ao sistema de informação possibilitar a transferências da informação, por meio de procedimentos seletivos que regulem a distribuição e uso [...] esta transferência acontece mediante um processo de comunicação entre os estoques de conhecimentos e os usuários desses conhecimentos.

É possível observar, portanto, que o sistema de informação arquivística, provedor da gestão documental, recebe a incumbência de transmitir aos usuários as informações registradas nos arquivos munidas de toda a contextualização envolta na produção, a fim de que possa satisfazer o comportamento informacional do usuário.

7.2 O controle de vocabulário na gestão documental

O controle de vocabulário pode ser compreendido como um processo documentário voltado ao desenvolvimento de instrumentos terminológicos, a saber, o vocabulário controlado.

Smit e Kobashi (2003, p. 11) comentam que o processo de ordenação física de documentos não é o único elemento necessário à recuperação da informação. Isso ocorre porque o volume de documentos cresce rapidamente na fase corrente e intermediária e a ordenação em si mesma, por mais elaborada que possa ser, não traz os elementos necessários ao mapeamento preciso da informação.

Em paralelo, a produção de documentos no meio digital torna o efeito da organização física praticamente nula, o que implica em maior dificuldade na recuperação da informação.

A aplicação do controle de vocabulário na gestão documental amplia a consistência e confiabilidade no processo de recuperação da informação, devendo o seu uso ser empregado nas nomenclaturas utilizadas nas funções e atividades institucionais, nas tipologias documentais, bem como na elaboração do plano de classificação, da tabela de temporalidade e nas listagens de assunto. (SIMIT; KOBASHI, 2003).

O estabelecimento do controle de vocabulário, no ambiente arquivístico, se dá com a interação entre o profissional da CI e o produtor do acervo documental. Deverá ser produzido com embasamento nas políticas institucionais, entendendo-se sobre a teia terminológica existente nos conjuntos de funções e atividades. Por conseguinte, seus benefícios se refletirão na utilização de um determinado público alvo que interagirá com procedimentos de indexação e recuperação da informação orgânica registrada no acervo.

Os instrumentos do vocabulário controlado deverão, portanto, relacionar e conceituar elementos terminológicos, hierarquizando as respectivas funções e atividades, além de estabelecer controle do vocabulário utilizado, particularmente no que diz respeito à sinonímia.

Tais instrumentos, além de prover a precisa recuperação da informação nas fases corrente e intermediária, serão os meios que facilitarão a compreensão dos feitos institucionais registrados na documentação, quando o acervo pertencer à fase da guarda permanente.

É por meio da terminologia utilizada na fase corrente que o usuário do grupo secundário poderá compreender todo o contexto de produção documental representado no acervo. O vocabulário controlado transforma-se em um importante elemento de representação da informação orgânica que conduz à interação investigativa que os arquivistas e os usuários pesquisadores desenvolverão na vigência do valor secundário.

Uma vez explanado acerca do controle de vocabulário na gestão de documentos, delinear-se-á a respeito do seu reflexo na linguagem documentária, no plano de classificação e no tesauro de função.

7.3 A linguagem documentária enquanto elemento de interação informacional

A interação entre o conteúdo informacional do acervo com o comportamento do usuário é feita pelo sistema de informações por meio do uso de linguagens. A linguagem pode ser compreendida como um conjunto de elementos dotados de significados voltados à comunicação entre sistemas e indivíduos.

É importante ressaltar que o foco de linguagem neste trecho do capítulo é direcionado à interação que o usuário faz com o conteúdo informacional existente no acervo arquivístico. É necessário fazer essa explanação, pois os conceitos de linguagens de marcação, trabalhados nesta dissertação, poderão confundir o leitor. As linguagens de marcação permitem prover uma estrutura para a automação da recuperação da informação e da interação da comunicação entre o usuário e o sistema de arquivo. O conceito de linguagem, que está

sendo trabalhado neste momento, diz respeito ao modo utilizado pelo usuário para prover a recuperação da informação.

Uma vez feito esse esclarecimento, é necessário apresentar a distinção da linguagem em dois tipos, a saber: linguagem natural e linguagem documentária.

Para Araújo Júnior (2007, p. 33), a linguagem natural constitui em um sistema de signos de espectro amplo, em que todos os outros sistemas de linguagem poderão repassar de língua. Em outras palavras, a linguagem natural pode ser compreendida como um amplo conjunto de signos utilizados pelo usuário para que ele possa comunicar-se no meio social. É a linguagem utilizada no cotidiano, com expressividade das características dos indivíduos que a utiliza. A linguagem natural compreende, desta forma, em uma estrutura ampla de comunicação em que cada grupo ou indivíduo desenvolve características peculiares para a troca de informação. Surge, então, a construção de gírias e expressões vocabulares vinculadas à existência do indivíduo, sendo difícil prover uma padronização e um controle de vocabulário, neste meio linguístico.

A linguagem documentária, por sua vez, poderá ser compreendida como um sistema de signos que tem por finalidade a representação abreviada da informação contida em uma peça ou acervo documental com o propósito de promover a recuperação da informação contida nos registros. O resultado dessa representação deverá ser, portanto, similar ao texto original, com equivalência de conteúdo informacional (ARAÚJO JÚNIOR, 2007; CERVANTES, 2006; MENEZES, CUNHA, HEEMAN, 2004).

Cervantes (2006, p. 12) afirma ainda que “com isso, permite-se que o sentido de um texto original, elaborado em Linguagem Natural, seja representado por meio da Linguagem Documentária.” A linguagem documentária permite a construção de uma estrutura padronizada para a definição de pontos de acesso que permita a recuperação de registros informacionais.

O ponto de acesso pode ser compreendido como elementos de conteúdo informacional por onde o usuário efetua a recuperação da informação. Um dos pontos de acesso utilizados na recuperação da informação orgânica são os termos descritores, providos pela indexação.

Segundo Araújo Júnior (2007, p. 20) a indexação é, portanto, um procedimento que traduz o teor documental em termos-chave que possibilitem a recuperação da informação ali registrada. No contexto da gestão documental e da informação orgânica, é necessário desenvolver instrumentos que relacionem a terminologia existente no corpo funcional da instituição com cada um dos tipos documentais existentes. Para possibilitar este feito, os

profissionais da informação que atuam nos arquivos deverão executar procedimentos de indexação que possam, por meio de um relacionamento de terminologias e tipos documentais, prover ao usuário a informação orgânica requerida.

Cervantes (2007, p. 19) comenta que um dos critérios de representação das Linguagens Documentárias é a ordenação dos conceitos, que diz respeito se estes serão pré ou pós-coordenados no momento da busca.

Segundo a autora, nas linguagens pré-coordenadas o indexador determina os assuntos a serem abordados no teor documental e os reúne embasado na dedução das estratégias de busca a serem empregadas pelo usuário na recuperação da informação. Desta maneira, o usuário deverá conhecer os procedimentos deduzidos pelo indexador, a fim de que possa obter a recuperação da informação. O desconhecimento por parte do usuário implicará em sérias dificuldades em recuperar a informação orgânica necessária.

Já as linguagens pós-coordenadas permitem a combinação dos assuntos no momento da busca informacional. Isso permite o desenvolvimento de um ambiente dinâmico de interação entre o usuário e os elementos informacionais de um acervo, dando-lhe maior autonomia para satisfazer seu comportamento informacional. Neste contexto, ainda segundo a autora, o indexador representa cada conceito separadamente para que possam ser combinados posteriormente.

O dinamismo existente no comportamento informacional dos usuários do arquivo corrente e intermediário impossibilita e traz a necessidade de desenvolver ferramentas de recuperação da informação trabalhado em ambos os elementos de ordenação de conceitos. O Plano de Classificação Funcional figura como elemento de linguagem pré-coordenada, enquanto o Tesauro de Função configura-se como um termo de linguagem pós-coordenada.

7.4 O plano de classificação documental enquanto uma ferramenta de vocabulário controlado a serviço da recuperação da informação orgânica.

Muito se explanou a respeito da composição do Plano de Classificação Funcional no capítulo anterior, e de toda a metodologia empregada em sua elaboração. A abordagem, neste capítulo, estará voltada ao forte vínculo dessa ferramenta com o controle de vocabulário e seus reflexos, no controle da informação.

O Plano de Classificação consiste em uma classificação rigorosa, tendo por base o princípio da hierarquia, em que funções e atividades são alocadas em classes documentais

auto excludentes. Tal hierarquização determina os critérios de ordenação física a serem empregados nos documentos, a fim de que possam ser encontrados.

Essa ferramenta possibilita que o controle de vocabulário seja empregado nas nomenclaturas utilizadas nos grupos funcionais e nas tipologias documentais, o que traz consistência na produção e recuperação de registros documentais.

O Plano de Classificação Funcional pode ser considerado um exemplo de instrumento de recuperação da informação orgânica que se utiliza nas linguagens pré-coordenadas, uma vez que esse instrumento traz uma taxonomia que apresenta relacionamento hierárquico existente entre função, atividade e tipologia documental. Para que o usuário possa encontrar um documento deverá conhecer a estrutura de relacionamento existente entre os elementos do plano, a saber: funções, atividades, e tipologias documentais, para que possa, então, efetuar a consulta e a recuperação do documento solicitado.

Embora seja essencial a um programa de gestão documental, o plano de classificação poderá impor ao usuário algumas restrições no momento da busca informacional. Se o mesmo desconhecer os termos utilizados para as atividades que geram os documentos de que necessita, poderá obter certa dificuldade para a recuperação da informação orgânica registrada no acervo institucional.

Segundo Smit e Kobashi (2003, p. 39), isso ocorre porque “uma categorização hierárquica de termos representa – ou atualiza – sempre um ponto de vista”, que está vinculado à cultura institucional. Se usuário buscar a recuperação de um ponto de vista ligeiramente diferente, poderá ser custosa a recuperação.

É necessário complementar as necessidades informacionais com instrumentos de recuperação da informação que trabalhem com linguagens pós-coordenadas, em que os assuntos relativos ao acervo documental, ou seja, os teores informacionais existentes no relacionamento entre tipologias e atividades, possam ser elencados e combinados no momento da busca.

Em outras palavras, torna-se necessário fazer menção a termos que não estão incluídos em determinada função, atividade ou tipologia documental, mas que possam ser considerados relacionados. (SMIT; KOBASHI, 2003, p. 40)

Observa-se que os tesauros são ferramentas de linguagem pós-coordenada que traz significativas contribuições para o atendimento dessas demandas. Smit e Kobashi (2003 p.40) comentam, ainda, que essas ferramentas oferecem duas formas distintas e não opostas de representação terminológica.

De fato, o plano de classificação e o tesauro se complementam enquanto ferramentas de vocabulário controlado, trazendo maior expressividade a todo o processo de representação e recuperação da informação orgânica nas fases corrente e intermediária, além de trazer grande reflexo à guarda permanente. Os tesauros e suas contribuições para a gestão da informação orgânica serão apresentados nos itens a seguir.

7.5 Os tesauros e suas contribuições para a recuperação da informação orgânica

O tesauro é um conjunto de termos relacionados semanticamente permitindo interação entre Linguagem Natural familiar ao usuário e a terminologia controlada da Linguagem Documentária. Consiste, portanto, em uma linguagem controlada, especializada, normalizada e pós-coordenada, voltada às finalidades documentárias. Essa linguagem permite a representação do conteúdo dos documentos com o uso de termos sinônimos, conceitos genéricos, específicos e relativos. (MENEZES, CUNHA, e HEEMAN; 2004).

Araújo Júnior (2009, p. 37) define o tesauro como uma:

lista estruturada de conceitos destinados a representar, de maneira unívoca, o conteúdo dos documentos e das buscas em um dado sistema e a apoiar o usuário na indexação dos documentos. Os conceitos são extraídos de uma lista finita, estabelecida *a priori*. Somente os termos que figuram nesta lista podem ser empregados para indexar os documentos e as buscas. Possui uma estrutura semântica com relações entre os termos de equivalência, hierarquia e associação.

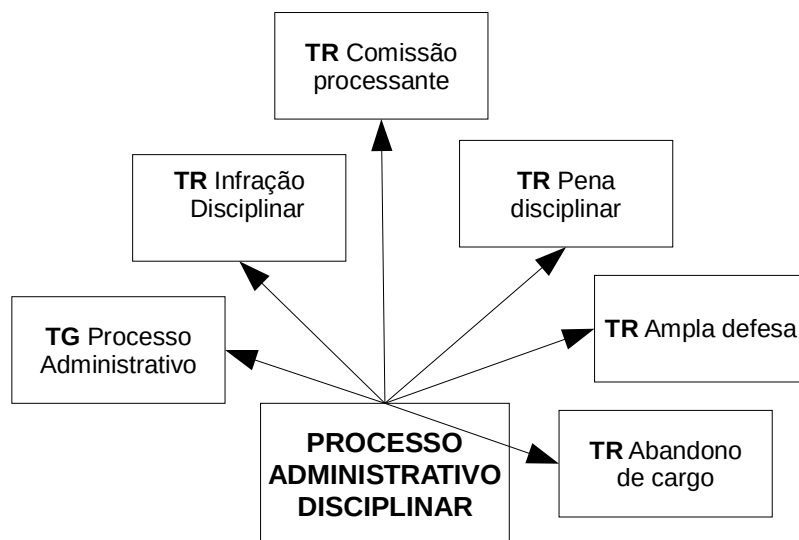
O relacionamento existente dentro do tesauro enriquece os sistemas informacionais, ampliando as possibilidades de busca que apresentar-se-ão em cada um dos tipos de relacionamentos existentes nos tesauros.

O relacionamento hierárquico permite estabelecer hierarquia terminológica entre Termos Genéricos (TG) e Termos Específicos (TE). A hierarquia possibilita a concepção de categorizações entre os termos, seguindo a de classificação. No contexto da produção documental, os termos que representam cada uma das funções são termos genéricos, e cada uma das atividades existentes em uma função são termos específicos.

O relacionamento associativo permite o desenvolvimento de associações de termos relacionados em um mesmo nível. Os Termos Relacionados (TR) não são sinônimos, nem tão pouco um termo genérico ou específico. O relacionamento associativo traz à recuperação da informação ampla a possibilidade para prover a recuperação da informação orgânica, pois amplia o vínculo de relacionamento os termos de uma determinada informação.

A figura a seguir traz um exemplo do uso de termos genéricos e termos relacionados.

Figura 10: Exemplificação do relacionamento hierárquico e associativo.



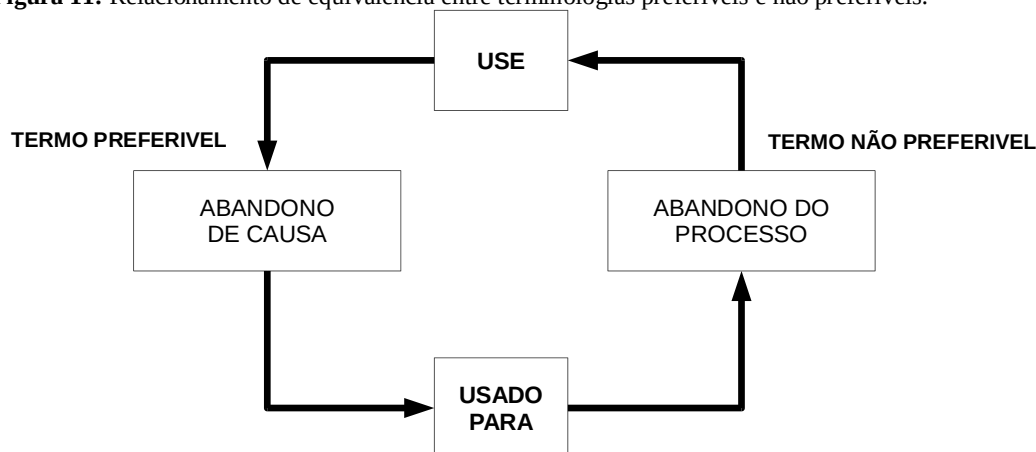
Fonte: Superior Tribunal de Justiça.

A figura apresentada permite observar os relacionamentos terminológicos envolvidos no conceito de Processo Administrativo Disciplinar, desenvolvido na apuração de uma infração cometida por um servidor público. Esse termo é hierarquicamente específico ao conceito de processo administrativo, que figura como termo genérico. Neste contexto, os termos referentes à ampla defesa, abandono de cargo, pena disciplinar e comissão processante associam-se ao conceito de processo disciplinar, figurando-se, portanto, como termos relacionados.

O relacionamento de equivalência, por sua vez, permite a designação de termos adotados e termos não preferíveis. Enquanto na Linguagem Natural vários termos são designados para representar um mesmo conceito, na Linguagem Documentária uma denominação é selecionada para ser o termo descritor e os demais, considerados não descritores, e fazem remissivas àquele.

O relacionamento de equivalência permite que termos sinônimos ou ainda inapropriados para a representação da informação sejam vinculados aos descritores preferíveis, constantes na classificação. A remissiva é indicada com a terminologia “Use” e “Usado Para”. A figura a seguir apresenta um exemplo de relacionamento de equivalência:

Figura 11: Relacionamento de equivalência entre terminologias preferíveis e não preferíveis.



Fonte: O autor.

Verifica-se, na imagem, que o termo “Abandono do Processo” não é autorizado; por este motivo, é estabelecida uma relação de equivalência “USE” que ratifica o posicionamento do termo adotado com o termo não autorizado: em vez de buscar “Abandono do Processo”, “USE” “Abandono de Causa”. O termo adotado, por sua vez, é usado para substituir o termo não autorizado, e para essa finalidade, utiliza-se a relação “USADO PARA”.

Essas relações enriquecem os procedimentos de recuperação da informação orgânica, pois mesmo desconhecendo o termo adotado, o usuário consegue encontrar as remissivas que o conduzirão à informação demandada. (CERVANTES, 2006; NATIONAL ARCHIVES OF AUSTRALIA, 2003).

O Arquivo Nacional da Austrália publicou, no ano de 2003, um manual que desenvolve, a partir da classificação funcional, um tesauro de funções e atividades visando à obtenção do controle de vocabulário para o gerenciamento de informações orgânicas na fase corrente e intermediária do ciclo de vida documental.

A partir do relacionamento entre funções e atividades identificados na implantação de um programa de gestão documental, torna-se possível o desenvolvimento de um tesauro de funções em que apresenta a terminologia identificada na análise função e atividades em cada um dos tipos de relacionamentos identificados.

Cervantes (2006, p. 30) afirma que a construção de um tesauro apresenta as seguintes etapas:

- Definição do domínio;
- Coleta e seleção de termos;
- Definição dos termos;
- Categorização;

- Organização das relações;
- Notas de escopo e de aplicação; e
- Apresentação do tesauro.

O domínio está relacionado ao âmbito a ser coberto pelo tesauro, ou seja, a área em que o instrumento atuará. No contexto da gestão documental, o domínio a ser observado é o de atuação da instituição para a qual o instrumento está sendo desenvolvido, englobando as atividades ali executadas.

A coleta e seleção de termos dá-se a partir da classificação funcional trabalhada em uma instituição. Nela é possível obter a visualização dos termos adotados representativos de cada uma das funções e das atividades.

A definição dos termos é elaborada a partir dos conceitos e noções das áreas a que dizem respeito. Este procedimento, segundo a autora, permite a delimitação das relações a serem privilegiadas neste instrumento.

A categorização consiste na identificação dos termos principais que são reagrupados em categorias. Neste procedimento são trabalhados os métodos dedutivos, que inicia na generalização e termina no específico, e o método indutivo que vai do particular, ao geral. A autora recomenda a adoção dos dois métodos.

A organização das relações é executada depois da definição das categorias e consiste em definir a hierarquia, associação e equivalência existente nos termos coletados.

A normalização, por sua vez, é direcionada à padronização do tesauro, visando assegurar o controle de vocabulário, bem como da circulação de informações. Estende-se em âmbito gramatical, e semântico. A autora (2006, p. 30) afirma que:

A normalização gramatical tem como objetivo uniformizar a apresentação das unidades informacionais do tesauro quanto aos aspectos relativos a gênero, número, grau; já a normalização semântica dos termos é garantida pela definição e pelas terminologias dos domínios e subdomínios da especialidade.

A normalização tem por incumbência trazer ao instrumento a devida conformidade com os conceitos vigentes na área de representação, e também dos padrões gramaticais.

As notas de escopo apresentam instruções a respeito da aplicação de alguns dos descritores, bem como dos usos a eles atribuídos.

A apresentação do tesauro, por fim, utiliza-se da forma sistemática, ou da forma alfabética. A forma sistemática refere-se à apresentação embasada na categorização estabelecida pelo tesauro, sendo as classes apresentadas alfabeticamente. Já na apresentação por ordem alfabética, cada um dos descritores aparecem vinculados àqueles que estão

relacionados na ordem de equivalências, relações genéricas, e relações associativas. (CERVANTES, 2006, p. 30-36).

No caso específico do tesauro de função, o Arquivo Nacional da Austrália apresenta oito procedimentos a serem seguidos para a construção do instrumento, apresentados na quadro a seguir:

Quadro 11: Procedimentos para a criação de um Tesauro de Funções.

PROCEDIMENTO	INSTRUÇÕES
1	- Adicionar os termos principais obtidos no plano de classificação relacionados às funções, e assegurar-se de que estejam rotulados como termo adotado. - Agrupar termos funcionais adicionais elaborados para atender aos requerimentos dos usuários. Rotular como termo adotado. - Inserir notas relacionadas aos termos da função. - Assegurar que estejam em ordem alfabética.
2	- Adicionar os termos principais obtidos no plano de classificação relacionados às atividades, e assegurar-se de que estejam rotulados como termos adotados. - Agrupar termos funcionais adicionais elaborados para atender aos requerimentos dos usuários. Rotule como termos adotados. - Inserir notas relacionadas aos termos da função. - Assegurar que estejam em ordem alfabética.
3	- Inserir os termos correlatos ao terceiro nível obtido na classificação, correlatos às transações desenvolvidas em uma atividade. Esses termos poderão ser correlatos a tipologias documentais ou “assuntos”. - Certificar-se de que os termos estejam rotulados como adotados. - Se for necessária uma nota para a explanação do termo, inseri-la vinculada ao recurso terminológico. - Reconhecer as fontes do termo. - Assegurar que estejam em ordem alfabética
4	- Estabelecer os relacionamentos hierárquicos genéricos e específicos existentes entre os recursos terminológicos. - Assegurar que os termos de uma função obtenham grupos de termos de atividades específicas relacionados. - Assegurar que cada uma dos termos de atividades tenham uma função como termo genérico para relacionamento hierárquico. - Assegurar que cada uma das atividades tenham relacionamento hierárquico superior aos termos do terceiro nível. - Assegurar que cada um dos termos de terceiro nível estejam relacionados a uma atividade.

5	- Estabelecer relações associativas entre termos de um mesmo nível hierárquico. Neste contexto, se definem termos relacionados a cada um dos termos referentes à função, atividades, e transações.
6	- Adicionar os termos não preferíveis a cada um dos termos existentes no tesauro. Os termos não preferíveis, embora inseridos no instrumento, não são indicados para o uso e, portanto, devem ser configurados para dar remissão aos termos adotados. - Assegurar a relação de equivalência vinculando o termo não preferível com o termo adotado por meio da etiqueta “USE”. - Assegurar que estejam ordenados alfabeticamente.
7	- Assegurar que cada um dos termos não preferíveis estejam direcionados ao termo adotado específico. - Indicar a relação de equivalência “Usado para” em que um termo adotado e utilizado em substituição da um termo não autorizado.
8	Revisar a lista terminológica para assegurar que nenhum relacionamento ilógico foi declarado.

Fonte: *Natrional Archives of Australia* (2003, p. 17-18, tradução nossa).

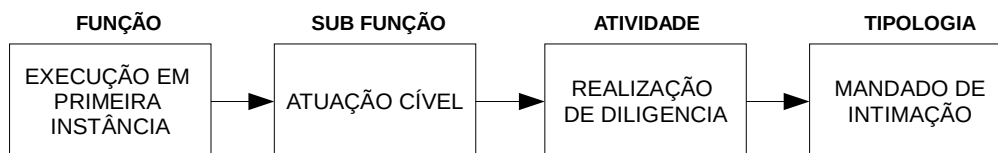
É possível observar que os tesouros de funções consistem em uma linguagem documentária construída a partir do relacionamento de elementos terminológicos referentes a funções e atividades. A cada um dos termos adotados são atribuídos relacionamentos de equivalência com termos não preferíveis, mas que poderão ser objeto de busca.

Os recursos providos pela estrutura de um tesauro permitem que um determinado tipo documental, item referenciado como elemento de transação, seja relacionado hierarquicamente à atividade e, conseqüentemente, à função que o produziu provendo a contextualização da produção documental. Permite, ainda, a agregação de conceitos relacionados que poderiam não ser identificados de imediato em uma função ou atividade.

Para fins de exemplificação, apresentar-se-ão alguns relacionamentos terminológicos elaborados a partir de um plano de classificação.

A figura apresentada a seguir traz um conjunto de classes documentais referentes à execução de funções e atividades de uma instituição jurídica.

Figura 12: Segmento de Funções e Atividades de um Ministério Público.

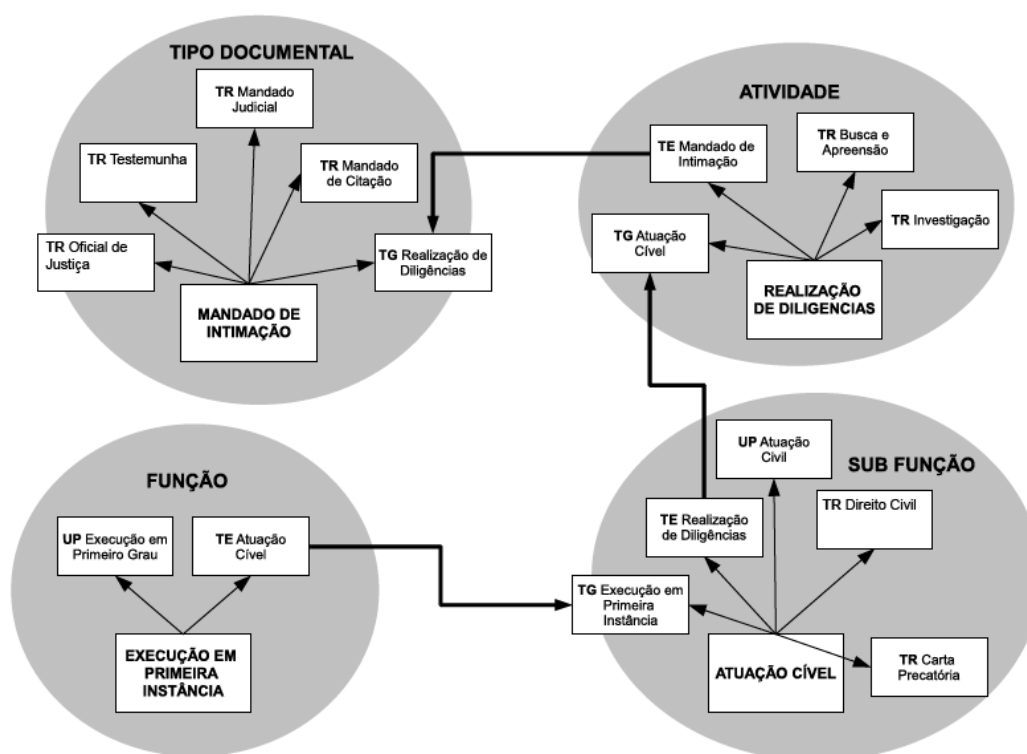


Fonte: Ministério Público do Estado de São Paulo.

Cada um destes elementos são termos adotados que representam um segmento específico de atuação de um Ministério Público. A função Execução em Primeira Instância diz respeito à atuação específicas das promotorias em processos judiciais que tramitam em instituições jurídicas de primeiro grau. A atuação cível traz uma delimitação da atuação voltada à área cível, configurando-se em uma subfunção. Dentro da atuação cível, uma das atividades da promotoria será a realização de investigações para apurar uma determinada denúncia. Uma das tipologias documentais produzidas neste procedimento é o mandado de intimação.

Esses elementos terminológicos apresentam relacionamento hierárquico entre si. A figura apresentada a seguir traz sugestões de possíveis relacionamentos terminológicos a serem desenvolvidos para cada um dos níveis hierárquicos apresentados:

Figura 13: Relacionamentos terminológicos em um Tesauro de Função.



Fonte: O autor.

Como é possível observar na figura, cada um dos termos apresentados na classificação recebe um conjunto de termos que os contextualizam e ampliam as possibilidades da busca informacional. Os vínculos são iniciados com abreviaturas em que UP indica relação de equivalência, TE e TG indicam relacionamento hierárquico e TR indica relacionamento associativo.

O conceito de Mandado de Intimação, por exemplo, pode ser vinculado a terminologias mandado de citação e mandado judicial. A realização de diligências pode se relacionar com a busca e apreensão e investigação. A atuação cível apresenta a forma autorizada de escrita frente à terminologia não preferível “atuação civil”, sendo que o mesmo ocorre no termo empregado na função.

Uma vez exemplificada a construção dos relacionamentos terminológicos dos tesouros de função, torna-se necessário definir a nomenclatura dos relacionamentos em outros idiomas. O quadro a seguir apresenta as nomenclaturas existentes no idioma português e inglês:

Quadro 12: Elementos do Tesouro.

PORTUGUÊS		INGLÊS	
TG	Termo Genérico	BT	Broader Term
TE	Termo Específico	NT	Narrower Term
TR	Termo Relacionado	RT	Related Term
NE	Nota Explicativa	SN	Scope Note
TGM	Termo Genérico Maior	TT	Top Term
UP	Usado Para	UF	Used For
USE	Use	USE	Use

Fonte: Cervantes (2006, p. 33-34).

Estes elementos relacionais trazem ao usuário um guia pormenorizado que os auxiliam na satisfação do comportamento informacional. O tesouro de função traz maior precisão à recuperação da informação, à medida que ajusta terminologias de conhecimento do usuário aos termos documentários que regem toda a classificação documental.

Uma vez apresentada a metodologia de construção de tesouro de função, surge a necessidade de desenvolver algumas explanações referentes aos benefícios obtidos na gestão da informação orgânica e seus reflexos no desenvolvimento de metodologias para a arquivística.

Com relação à interação do plano de classificação e do tesouro funcional, é possível observar que a construção do tesouro se dá a partir da estrutura definida no plano, permitindo a reordenação dos termos ali existentes. É importante salientar que o tesouro de função não substitui o plano, mas apresenta diferentes visualizações para os conjuntos terminológicos existentes na classificação funcional.

Smit e Kobashi (2003, p. 44) comentam ainda que “o tesauro não supõe que todas as ramificações das atividades sejam previstas *a priori*, mas que estas possam ser livremente acopladas às atividades, quando a documentação assim o exigir”.

Este instrumento de vocabulário controlado possibilita que uma gama de ramificações terminológicas sejam consolidadas posteriormente, inclusive no momento de uma busca por recursos informacionais. Isso ocorre porque o usuário poderá realizar combinações terminológicas no momento da solicitação, o que traz maior flexibilização. Outro benefício é o relacionamento de termos não autorizados, que poderão estar mais familiarizados com a linguagem do usuário, o que garantirá maior precisão na recuperação da informação.

A respeito das funções desempenhadas pelo tesauro, Smit e Kobashi (2003, p. 45) afirmam que:

As funções desempenhadas pelo tesauro funcional podem ser resumidas como segue: opera como índice de termos do plano de classificação; opera como índice para encontrar séries documentais a serem eliminadas; provê pontos de acesso ao usuário, permitindo navegar de sinônimos ou outros termos não adotados para a terminologia adotada pelo arquivo; pode ser usado como ferramenta para ajudar o funcionário a encontrar documentos necessários para as tarefas do dia a dia.

O maior benefício que o plano de classificação e o tesauro funcional apresentam para a gestão da informação orgânica é o controle de vocabulário e o acesso à informação orgânica em conformidade com o comportamento informacional do usuário.

Em relação à arquivologia, é importante salientar que a abordagem sobre a elaboração de tesouros funcionais confere à área crescimento em torno de sua identidade, voltada ao desenvolvimento de metodologias que promovam tratamento documental que atenda aos interesses de seus usuários, em detrimento de um olhar focado num saber meramente pragmático.

A discussão e sobre os instrumentos de vocabulário controlado no meio aquívistico provoca maior profundidade nos estudos relacionados à representação da informação relacionadas ao comportamento dos diferentes usuários que interagem com a informação orgânica durante o ciclo de vida documental.

Tais discussões reafirmam a nobreza dos princípios arquivísticos consolidados nos últimos séculos, renovando-lhes o esplendor. Torna evidente, então, que a aplicabilidade dos mesmos é indispensável à satisfação dos usuários do grupo primário e secundário, principalmente nestes tempos em que a tecnologia da informação está sendo empregada na produção de registros informacionais orgânicos.

Após ser delineado o controle de vocabulário e seus reflexos na constituição dos instrumentos de representação da informação orgânica no processo da gestão documental, explanar-se-á sobre a interação desses instrumentos com as linguagens de marcação, em benefício de toda a gestão documental.

7.6 A interação das linguagens de marcação com planos de classificação e os tesauros de função: ferramentas de apoio à gestão da informação arquivística

Depois de apresentados os processos metodológicos envolvidos na gestão da informação orgânica e na constituição da web semântica, surge a necessidade de apresentar a interação existente entre as linguagens de marcação e o vocabulário controlado da gestão documental.

7.6.1 A construção de planos de classificação em linguagens de marcação

A explanação a respeito da web semântica, no capítulo 3 (três), tornou evidente que a linguagem RDF pode prover os recursos necessários à representação dos planos de classificação no meio digital. Estas linguagens, trabalhando no ambiente *web* das instituições, poderão expandir a classificação empregada para diversas unidades institucionais, padronizando e trazendo maior conformidade a todo processo classificatório.

No capítulo 6 (seis), foi apresentado um fragmento do plano de classificação que será retomado na figura a seguir:

Figura 14: Fragmento de Plano de Classificação.

ELEMENTO FUNCIONAL	ÍTEM DE CLASSIFICAÇÃO
Função	Gestão de Documentos e Informações
Subfunção	Gestão de Tecnologia da Informação
Atividade	Manutenção de Equipamentos Instalação de Programas
Tipologia	Termo de Garantia de Equipamentos e de Programas de Computador

Fonte: O autor.

A partir deste quadro de itens funcionais, apresenta-se uma proposta de representação dos itens classificação na linguagem de RDF. Tais itens poderiam ser representados da seguinte maneira:


```

<?xml version="1.0"?>

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="http://www.arquivos.com/classificacao#">

  <rdf:Description rdf:ID="Gestao_de_Documentos_e_Informacoes">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:comment>
      Essa é a classe das Funções institucionais que são executadas para que
      a organização alcance sua missão institucional.
    </rdfs:comment>
  </rdf:Description>

  <rdf:Description rdf:ID="Gestao_de_Tecnologia_da_Informacao">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:subClassOf rdf:resource="#Gestao_de_Documentos_e_Informacoes"/>
    <rdfs:comment>
      Essa classe é referente à subfunção institucional. Essa subfunção é
      subordinada à função Gestão de Documentos e Informações.
    </rdfs:comment>
  </rdf:Description>

  <rdf:Description
    rdf:ID="Manutencao_de_equipamentos_e_instalacao_de_programas_de_computador">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:subClassOf rdf:resource="#Gestao_de_Tecnologia_da_Informacao"/>
    <rdfs:comment>
      Essa é a classe das atividades institucionais. Essa atividade é um
      subgrupo da Subfunção Gestão de Tecnologias da Informação.
    </rdfs:comment>
  </rdf:Description>

  <rdf:Description
    rdf:ID="Termo_de_garantia_de_equipamentos_e_de_programas_de_computador">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:subClassOf
      rdf:resource="#Manutencao_de_equipamentos_e_instalacao_de_programas_de_computador"/>
    <rdfs:comment>
      Essa é a classe referente a tipologia documental, que é um subgrupo da
      atividade de manutenção de equipamentos e instalação de programas
    </rdfs:comment>
  </rdf:Description>

</rdf:RDF>

```

O exemplo apresentado a seguir é uma singela demonstração do potencial proporcionado pela modelagem RDF e RDFS à representação de planos de classificação.

Nele é possível observar a edificação de declarações que delimitam os níveis de relacionamento hierárquico entre as classes terminológicas. Tais declarações permitem aos agentes de *software* a compreensão dos elementos terminológicos em forma de taxonomias e, desta forma, o sistema conseguirá contextualizar os itens relacionados em cada nível.

Isso possibilita que sistemas digitais de gestão documental possam reconhecer a rigidez hierárquica envolta nos processo classificatório e aplicá-la na organização de um acervo documental, com documentos analógicos e digitais, preservando, dessa maneira, informações referentes ao contexto de produção.

Como o RDF e RDFS possibilitam a construção dos mais variados tipos de relacionamento entre declarações, é possível desenvolver a representação detalhada de planos de classificação, e até mesmo dos prazos de guarda existentes em uma tabela de temporalidade, o que facilita a gestão dos itens documentais. Essas linguagens trazem, portanto, uma estrutura eficiente para a construção de sistemas que promova o gerenciamento de documentos digitais.

Elas apresentam recursos amplos para a representação de tesouros, conforme será demonstrado a seguir

7.6.2 A construção de tesouros de função em linguagens de marcação

Os tesouros de função sempre proporcionaram ao usuário grande precisão na recuperação da informação, mesmo quando eram manuseados em catálogos impressos. Por um longo período de tempo, esse foi o método mais empregado para a disseminação destes instrumentos de indexação. A implementação da tecnologia da informação nos arquivos e bibliotecas trouxe aos profissionais da informação métodos inovadores para a construção, disseminação e utilização desses instrumentos.

As relações existentes entre os termos podem ser representadas no meio digital, estabelecendo-se vínculos entre a estrutura terminológica e o conteúdo informacional dos registros documentais correlatos.

Quando o usuário gerar uma demanda informacional sobre um sistema de busca digital, a estrutura de um tesouro poderá instruir as ferramentas de *software* que atuam nos motores de busca para que os elementos informacionais encontrados estejam de acordo com o comportamento informacional do usuário.

Lambe (2007, p. 213) comenta ainda que, se o motor de busca consultar um tesouro sempre que uma palavra chave for introduzida, poderá ser construída uma taxonomia rica, extensa e avançada de termos-chave não autorizados.

Em relação aos usuários do arquivo corrente e do arquivo intermediário, os tesouros de função poderão prover meios para a antecipação de registros arquivísticos que satisfaçam o comportamento informacional. Isso é possível quando se combina o cargo exercido pelo usuário ao conjunto de funções e atividades de atuação do indivíduo.

Neste contexto, as linguagens de marcação provêm uma forte estrutura para a construção de tesauros no ambiente digital, dentre elas destacam-se a linguagem XML e as modelagens provenientes da web semântica, a saber: a RDF, OWL e SKOS, já abordadas nesta dissertação.

Em capítulos anteriores foi demonstrado que a linguagem de modelagem de dados RDF e RDF Schema (RDFS) trazem uma estrutura para o desenvolvimento de declarações semânticas entre termos, constituindo-se em triplas que inter-relacionam o sujeito, o objeto e o predicado. Por meio da RDFS, é possível a definição de declarações entre classes e subclasses, e definir algum controle de vocabulário.

A linguagem OWL traz recursos ontológicos à modelagem de dados permitindo maior interação entre as classes. Por meio dessa linguagem, é possível obter representações complexas de um determinado domínio do conhecimento com a construção de ontologias digitais.

A modelagem SKOS trabalha em conjunto com cada uma dessas linguagens e permite a definição de taxonomias e tesauros. O presente trabalho tem o foco direcionado à construção de tesouro de funções e sua representação digital por meio das linguagens de marcação. O quadro a seguir apresenta as principais relações existentes na terminologia de um tesouro com as definições da linguagem SKOS:

Quadro 13: Tesauros e SKOS.

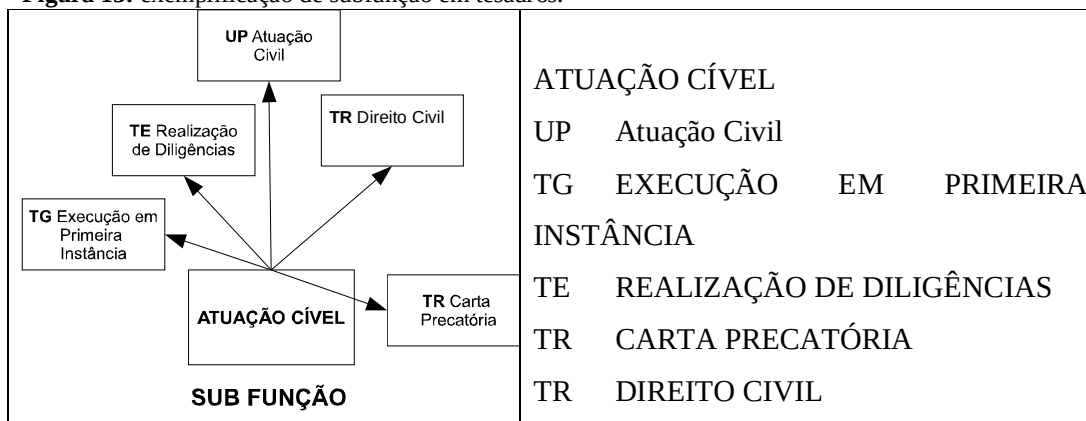
PORTUGUÊS		INGLÊS		SKOS
TG	Termo Genérico	BT	Broader Term	skos:broader
TE	Termo Específico	NT	Narrower Term	skos:narrower
TR	Termo Relacionado	RT	Related Term	skos:related
NE	Nota Explicativa	SN	Scope Note	skos:scopeNote
TGM	Termo Genérico Maior	TT	Top Term	
UP	Usado Para	UF	Used For	skos:prefLabel
USE	Use	USE	Use	skos:altLabel

Fonte: O autor.

Constata-se que a modelagem SKOS provê elementos para a descrição de cada um dos recursos relacionais existentes em um tesouro. Tais recursos são combinados na construção de um tesouro digital. Neste contexto, os termos controlados da classificação são relacionados, a fim de se estabelecer os tipos de relacionamentos existentes em um tesouro.

Para a construção de uma exemplificação, toma-se a subfunção “ATUAÇÃO CÍVEL” apresentada na figura a seguir:

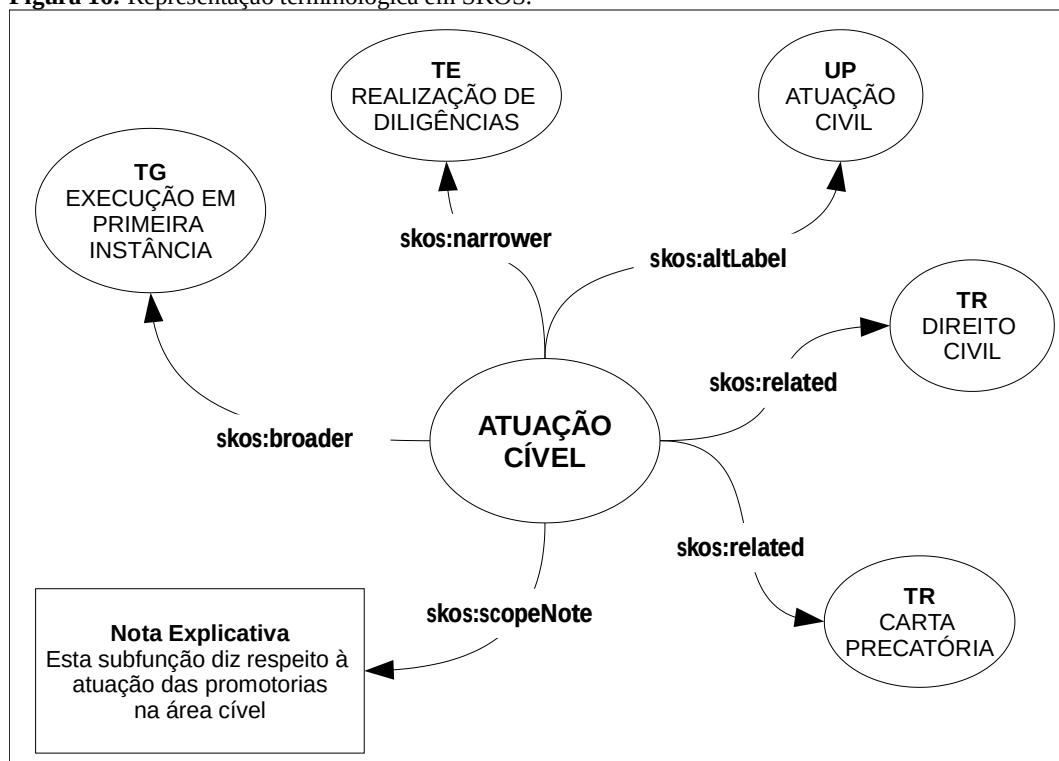
Figura 15: exemplificação de subfunção em tesouros.



Fonte: O autor.

A figura apresenta a construção do relacionamento terminológico da subfunção atuação cível descrita nos relacionamentos de um tesouro, em que os termos são posicionados alfabeticamente em ordem sistemática de relação equivalência (UP), hierárquica (TG e TE) e associativa (TR).

Conforme já mencionado, essas relações podem ser representadas na web semântica por meio da linguagem SKOS. Esta modelagem trabalha em conjunto com a RDF e proporciona a construção de declarações de relacionamentos entre termos de um tesouro. A imagem a seguir apresenta um esquema de codificação em SKOS para a terminologia da subfunção atuação cível:

Figura 16: Representação terminológica em SKOS.

Fonte: O autor.

(Foi inserida uma nota explicativa para ampliar a dimensão do exemplo) Observa-se a possibilidade de recursos disponibilizados pela linguagem SKOS para descrever os relacionamentos terminológicos de um tesouro. Com esses recursos, a proposição de tesouros de funções feita pelo Arquivo Nacional da Austrália fica completa no meio digital. Por meio da integração de tesouros de funções com as linguagens de marcação, os sistemas computacionais adquirem habilidades para trabalhar todo o fluxo de informação orgânica. A figura a seguir apresenta a transcrição do exemplo anterior em um documento XML, com as modelagens RDF e SKOS:

Figura 17: Representação de Fragmento de Tesauro de Função em XML, RDF e SKOS.

```

<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:skos="http://www.w3.org/2004/02/skos/core"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dct="http://purl.org/dc/terms/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:vs="http://www.w3.org/2003/06/sw-vocab-status/ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <skos:Concept
    rdf:about="http://www.exemplo.org/tesauro_funcoes#atuacao_civel">
    <skos:prefLabel xml:lang="pt">Atuação Cível</skos:prefLabel>
    <skos:scopeNote xml:lang="pt"> Esta subfunção diz respeito à
    atuação das promotorias na Área Cível.
    </skos:scopeNote>
    <skos:altLabel xml:lang="pt">Atuação Civil</skos:altLabel>
    <skos:broader
      rdf:resource="http://www.exemplo.org/tesauro_funcoes#execucao_em_primeira_instancia"/>
    <skos:narrower
      rdf:resource="http://www.exemplo.org/tesauro_funcoes#realizacao_de_diligencias"/>
    <skos:related
      rdf:resource="http://www.exemplo.org/tesauro_funcoes#direito_civil"/>
    <skos:related
      rdf:resource="http://www.exemplo.org/tesauro_funcoes#carta_precatoria"/>
    </skos:Concept>
    <skos:Concept
      rdf:about="http://www.exemplo.org/tesauro_funcoes#execucao_em_primeira_instancia">
      <skos:prefLabel xml:lang="pt"> Execução em Primeira Instância
      </skos:prefLabel>
    </skos:Concept>
    <skos:Concept
      rdf:about="http://www.exemplo.org/tesauro_funcoes#realizacao_de_diligencias">
      <skos:prefLabel xml:lang="pt"> Realização de
      Diligências</skos:prefLabel>
    </skos:Concept>

  </rdf:RDF>

```

Fonte: O autor.

É possível observar que a linguagem OWL não precisou ser utilizada na consolidação do exemplo, tamanhos são os recursos providos pela RDF e RDFS. O uso da linguagem OWL poderá trazer maiores recursos na descrição do conteúdo informacional, trazendo declarações de relacionamentos de nível ontológico, que vão além dos recursos dispostos em um tesauro. Para a exemplificação utilizada neste trabalho, no entanto, é suficiente a RDFS.

O fragmento de tesauro apresentado anteriormente, referente à Atuação Cível, foi transcrito em RDF e SKOS. É possível observar a existência de declarações que descrevem os relacionamentos terminológicos entre os elementos.

Observa-se, na figura, a existência do elemento *skos:concept* que não fora apresentada anteriormente. Esse elemento delimita a área conceitual do recurso terminológico que deseja trabalhar, neste caso, a subfunção “Atuação Cível”.

É utilizado, então, o elemento *rdf:about* para indicar a localização do conceito que deseja descrever, o qual está indicado em uma URL fictícia “http://www.exemplo.org”. A partir desta declaração, inicia-se a declaração terminológica envolta no conceito. É utilizado o elemento *skos:prefLabel* em conjunto com *xml:lang* para a definição do termo autorizado, a saber, “Atuação Cível”. Caso haja terminologia correlata em outros idiomas, é possível declarar cada um dos referidos termos especificando a sigla correta no elemento *xml:lang*.

Este benefício permite que tesauros de funções sejam construídos em múltiplas linguagens para padronizar e controlar elementos terminológicos de classificação em instituições multinacionais. A partir de então, declara-se a nota explicativa, que explana a finalidade da subfunção, podendo esta também ser descrita em múltiplos idiomas.

São declarados, na sequência, os termos não preferíveis, ou seja, a “Atuação Civil”. Para isso é utilizado o elemento *skos:altLabel*, que constitui as relações de equivalência. Define-se, então, o termo genérico, com *skos:broader*, e o termo específico, com *skos:narrower*. Estes elementos permitem a definição da hierarquia em que se situam os conceitos. Declara-se, por fim, os termos relacionados com o elemento *skos:related*, encerrando, posteriormente, a construção do conceito.

A título de ilustração, é apresentada a abertura e encerramento dos termos hierárquicos relacionados ao conceito principal, ou seja, “Execução em Primeira Instância” e “Realização de Diligências” para situar o leitor da construção de declaração de termos genéricos e específicos.

A partir da união de várias declarações como esta, é possível concluir a representação de um tesouro de função com as linguagens de marcação, que permitirão o estabelecimento do controle de vocabulário e da recuperação da informação em sistemas digitais de gestão de documentos.

Por meio deste exemplo, é possível obter uma singela dimensão dos benefícios que as linguagens de marcação podem trazer à gestão da informação e de documentos arquivísticos. Quando tesauros de funções são representados nas linguagens RDF e SKOS, poderá haver um direcionamento significativo dos sistemas digitais de recuperação da informação orgânica, voltado à demanda construída pelos usuários. Ao mesmo passo em que supre a demanda, o tesouro de função ressalta a organicidade do acervo e realça o vínculo do

registro documental com o seu produtor, consolidando a gênese documental. Neste contexto, a demanda dos usuários e gênese documental são supridas por um único instrumento.

A linguagem SKOS pode ser também implementada conjuntamente com RDF e RDFS na linguagem OWL, servindo de elemento para a construção de ontologias que representem um domínio do conhecimento. Não é, entretanto, o foco deste trabalho a construção de ontologias a partir dos tesouros de funções; embora seja necessário estabelecer algumas ponderações.

A respeito das ontologias, Ferneda (2003, p. 116) comenta que:

Na maioria das vezes, uma ontologia toma a forma de uma árvore hierárquica de classes, de maneira que cada classe herda as características de uma ou mais classes superiores. Cada classe representa um conceito do domínio que está sendo modelado, e seu significado é expresso pelas suas propriedades, similaridades e diferenças em relação a outros conceitos. No contexto da Ciência da Informação, este recurso é utilizado em larga medida, denominado “Plano de Classificação” ou “Tesouro”.

É possível observar que a ontologia pode ser construída a partir de classes advindas de um tesouro. A esse respeito, Torres, Lora e López (2008, p. 09) comentam que a utilização de um tesouro é um ponto de partida relevante para a construção de uma ontologia, uma vez que a obtenção de recursos informacionais para a representação de um determinado domínio do conhecimento é custosa, e complexa.

Uso dos tesouros poderá reduzir os custos de composição da ontologia, além de trazer pistas para a composição terminológica do domínio do conhecimento a ser representado. O tesouro poderá ser compreendido como uma ontologia simples. (TORRES; LORA; LÓPEZ, 2008)

Os tesouros de função trabalham em conjunto com os usuários de arquivo, visando satisfazer o comportamento informacional dos mesmos. Esta ferramenta, portanto, poderá prover termos que poderão ser utilizados para direcionar uma determinada área do conhecimento ao comportamento informacional obtido no exercício de funções e atividades institucionais.

No ambiente digital, os tesouros de função podem trazer inúmeros benefícios ao gerenciamento estratégico da informação orgânica e ao programa de gestão documental, a saber:

- Padronização da terminologia empregada na classificação da informação orgânica em organizações de grande porte, inclusive em múltiplos idiomas;
- recuperação da informação orgânica a partir do contexto de atuação do usuário;

- suporte para processos digitais de automação da gestão documental como workflows e gestão de conteúdo;
- usuário munido de informações necessárias ao cumprimento de suas atribuições; e
- procedimento de avaliação otimizado, que permite a identificação de cargos que indicam stakeholders e assuntos para avaliação documental.

Os tesauros podem ser, ainda, uma forma eficiente para a pré-disponibilização da informação aos usuários, apontando-lhes antecipadamente os recursos informacionais existentes em um acervo arquivístico. Isso é possível por meio da relação terminológica existente entre funções e atividades institucionais. O tesouro de funções é, ainda, um recurso que permite o controle terminológico sem prejuízo da recuperação devido ao uso de termos adotados e não preferíveis para um mesmo conceito.

Este instrumento poderá, portanto, alimentar sistemas computacionais que trabalhem com o gerenciamento da informação para a disponibilização dos recursos informacionais aos usuários.

7.7 Considerações

O plano de classificação e o tesouro de função, conforme explanado neste capítulo, trazem inúmeros benefícios à gestão documental e informacional, uma vez que estabelecem controle de vocabulário sobre as ações dos usuários de arquivo corrente.

São, portanto, ferramentas muito visadas em todo o procedimento relacionado à gestão de documentos de uma instituição. Por meio destes instrumentos, os usuários localizam as informações necessárias ao exercício de suas atribuições podendo, inclusive, identificar usuários potenciais para a apropriação dos recursos informacionais.

Os instrumentos tecnológicos providos pelas linguagens de marcação provêm a estrutura necessária para a automação da recuperação informacional por meio da incorporação de tesauros e planos de classificação junto aos agentes de *software*. Quando aliadas a motores de busca, estas representações poderão ampliar a precisão e confiabilidade dos resultados apresentados aos usuários.

Ante ao exposto, resta afirmar que as linguagens de marcação trazem benefícios incontestes à informação orgânica, no que tange a representação de planos de classificação e de tesauros para automação do fluxo informacional em arquivos correntes e intermediários, por meio de sistemas digitais de gestão de documentos.

Uma vez apresentadas as contribuições das linguagens de marcação para a Gestão da Informação Arquivística, resta realizar uma última abordagem concernente à preservação digital, que será o foco do próximo capítulo.

8 LINGUAGENS DE MARCAÇÃO: UMA ALTERNATIVA PARA A PRESERVAÇÃO DE DOCUMENTOS DIGITAIS

O presente capítulo tem por objetivo apresentar os benefícios oferecidos pelas linguagens de marcação aos procedimentos voltados à preservação digital.

O advento das tecnologias da informação possibilitou uma série de benefícios aos produtores de arquivo na fase de uso corrente dos registros arquivísticos. A praticidade envolta no processo de transmissão do fluxo informacional faz com que a migração dos registros em meio analógico para documentos em meio digital ocorra exponencialmente.

A natureza dos documentos digitais, no entanto, faz com que os registros fiquem obsoletos e se deteriore rapidamente, já no início da fase de guarda intermediária. Quando for necessária a recuperação para os fins de guarda permanente, as mídias e os recursos de *software* poderão estar indisponíveis, o que causará a perda completa da informação.

Um dos principais problemas reside na obsolescência de formatos de arquivos digitais proprietários, que dificulta o processo de leitura da informação, ainda que as mídias e os equipamentos de leitura do mecanismo de *hardware* estejam em boas condições.

Neste contexto, as linguagens de marcação emergem com significativas contribuições que serão apresentadas a fim de contornar os desafios interpostos pela obsolescência tecnológica.

Antes de adentrar na abordagem propriamente dita é necessário prestar um esclarecimento. Neste capítulo, o maior foco será direcionado à representação dos formatos de arquivo digitais nas linguagens de marcação para a preservação digital, em detrimento da representação informação, embora essa última possa ser mencionada eventualmente. Por formatos digitais compreende-se ser a forma de composição e interpretação do conjunto de bits que foram um objeto digital.

8.1 O advento das tecnologias da informação e suas implicações na produção documental

Até meados do século XX, a produção documental arquivística embasava-se, unicamente, nos suportes de documentos analógicos. O desenvolvimento dos princípios arquivísticos e grande parte da metodologia para tratamento documental desenvolveu-se embasado nesse tipo de suporte.

A partir de meados do século XX, surgiram as Tecnologias da Informação (TI), inicialmente restritas às entidades militares, sendo popularizadas e disseminadas

posteriormente em larga escala. Observa-se, a partir de década de 1970, o surgimento de computadores pessoais que trouxeram grande popularização dos recursos digitais de registro de informação.

No Brasil, as tecnologias da informação, especificamente os recursos computacionais, se popularizaram a partir da década de 1990, momento em que computadores passaram a ser adquiridos e utilizados em larga escala por indivíduos e instituições. Na década de 2000, o avanço computacional se intensificou de modo que indivíduos e instituições passaram a utilizar e depender dos recursos tecnológicos para melhor execução de suas atribuições.

Os registros digitais estão substituindo paulatinamente documentos específicos do meio analógico. É possível observar a existência de diversos tipos de gêneros arquivísticos registrados em um único ambiente digital.

A TI revolucionou os tipos de suportes documentais utilizados para a produção informacional. Documentos que antes eram bem definidos para o registro de um determinado tipo de informação passaram a ser registrados no meio digital em forma de bits.

São muitos os benefícios oferecidos pela documentação digital:

- São relativamente fáceis de serem transmitidos de um local para outro;
- não ocupam espaço físico nas mesmas proporções dos documentos tradicionais;
- a busca e recuperação da informação digital ocorrem de forma ágil e precisa;
- traz a possibilidade para se controlar o fluxo informacional e a atuação dos usuários de arquivos correntes; e
- etc.

Esses benefícios fazem com que a produção documental cresça exponencialmente. A ampliação da capacidade de armazenamento de registros digitais traz sustentação para o crescimento de informações registradas unicamente no ambiente digital, bem como para a migração dos suportes documentais convencionais.

Sobre esta questão, Arellano (2004) comenta que

A natureza dos documentos digitais está permitindo ampla produção e disseminação de informação no mundo atual. É fato que na era da informação digital se está dando muita ênfase à geração e/ou aquisição de material digital, em vez de manter a preservação e o acesso a longo prazo aos acervos eletrônicos existentes.

As tecnologias da informação trouxeram procedimentos eficientes para gerenciar a informação na fase de uso corrente e, por este motivo, tem incorporado registros documentais que antes eram registrados unicamente nos suportes digitais.

A incorporação no meio digital, quando não planejada, poderá suprimir as questões relacionadas à preservação por longo prazo em prol da modernização do acervo. Por conseguinte, quando estes registros estiverem na fase de guarda intermediária e permanente, poderão surgir imprevistos que poderão inviabilizar por completo o acesso à informação.

Isso se dá porque a composição dos documentos digitais diverge da produção de documentos em suporte convencional. As peculiaridades do suporte convencional serão apresentadas no item a seguir.

8.1.1 A composição dos documentos digitais: características frente aos documentos tradicionais

No primeiro capítulo, abordou-se algumas diferenças existentes entre o suporte convencional e os suportes digitais. A abordagem trouxe a percepção de que o suporte tradicional traz registros analógicos que podem ser lidos diretamente, ou por meio de equipamentos mecânicos, pois eles fazem uma analogia ao objeto que representam.

Os registros em papel, por exemplo, podem ser lidos por um olho humano saudável sem que haja nenhuma necessidade de conversão. Os registros em microfilme necessitarão de equipamentos mecânicos para a ampliação dos registros, podendo ser lidos sem grande dificuldade.

O suporte digital, por sua vez, traz um procedimento mais complexo para a leitura, uma vez que as informações não estão disponíveis de modo a possibilitar visualização imediata. Uma série de procedimentos de interação entre *hardware* e *software* será necessária para que o documento digital seja apresentado em um dispositivo de saída. (NATIONAL ARCHIVES OF AUSTRALIA, 2002)

Ferreira (2006, p. 24) apresenta uma ilustração que sistematiza os passos de interpretação dos objetos digitais, conforme segue:

Figura 18: Cadeia de interpretação do nível físico ao conceitual.



Fonte: Ferreira (2006, p. 24).

Observa-se que, entre a mídia digital e a interpretação do documento por um periférico de saída, existe um longo percurso a ser corretamente processado para que a apresentação original do registro arquivístico não se perca.

É possível verificar, ainda, que um registro digital é composto de um conjunto de bits que, desprovido de sua contextualização, poderá significar qualquer coisa.

Os diversos gêneros documentais, tão distintamente abordados nos suportes tradicionais, são registrados da mesma maneira. Os recursos de *software*, que processam e apresentam o documento em sua forma original, são também registrados em bits 0 e 1.

Santos, V. (2007, p. 26) afirma que a estrutura de um documento digital é formada por um equipamento físico, constituindo o *hardware*, um equipamento lógico, envolvendo o *software*, e a informação que compreende a soma do suporte e dos bits.

Comenta ainda que os documentos digitais são formados de três maneiras: diretamente, com o uso de *software* específico; por procedimentos de digitalização e por sistemas gerenciadores de bancos de dados.

Observa-se, portanto, que a implementação de sistemas informatizados para o gerenciamento documental trouxe muitos avanços nos meios de produção documental.

Dentre os avanços, destaca-se o uso de ferramenta de banco de dados como suporte ao registro da informação orgânica, que será abordado a seguir.

8.1.2 Banco de dados: uma nova concepção na produção de registros documentais arquivísticos

A tecnologia da informação possibilitou a adoção de métodos inovadores na produção da informação orgânica, em que o documento deixa de ser, essencialmente, um suporte físico com o registro informacional. Os novos sistemas computacionais trouxeram a construção das bases de dados que segmentam a informação orgânica em diversificados campos.

Desta forma, um banco de dados é constituído com diversas tabelas em que cada linha significa um registro de informação, sendo estes segmentados em diversas colunas que constituem os campos de registro de dados. As tabelas poderão conter, ainda, campos para chave primária e estrangeira que permitirão o relacionamento de informações registradas entre as tabelas.

Neste contexto, uma base de dados conterá inúmeros registros informacionais orgânicos que representarão o trâmite informacional das atividades no organismo produtor. Devido à segmentação em nível de campos de dados, o documento arquivístico perde a sua estrutura tradicional, tendo o seu conteúdo reduzido em nível de dados.

Os sistemas de gerenciamento de bancos de dados permitem a composição de relatórios que poderão reconstituir a estrutura de conteúdo existente no documento de acordo como a ordem dos elementos diplomáticos, de modo que seja constituído o documento arquivístico em seu formato tradicional. A estrutura da base de dados, no entanto, não permitirá a imediata compreensão da informação orgânica como ocorre com o documento de arquivo tradicional.

Durante a fase corrente, os bancos de dados possibilitam maior dinamismo e eficiência aos sistemas de informação uma vez que um mesmo segmento de dados poderá ser combinado e utilizado em inúmeras situações, proporcionado economia de espaço, maior agilidade no registro e na recuperação da informação orgânica ali vigente; o que resultará na economia de recursos humanos, materiais e financeiros.

Observa-se que na guarda permanente os registros informacionais registrados em banco de dados deverão apresentar a composição da base e o contexto de produção informacional devidamente registrados e estruturados para que a informação orgânica possa ser compreensível e representável aos usuários do arquivo permanente.

Embora o presente trabalho não tenha o seu foco direcionado ao tratamento informacional a ser dado a uma base dados e nele não possa se aprofundar, é importante frisar que essa nova modalidade de documento arquivístico precisa ser devidamente

contextualizada, a fim de que possa receber a devida descrição e recuperação da informação na guarda permanente.

Tão importante quanto a representação da informação é a provisão de procedimentos adequados para a preservação da informação registrada em bases de dados de valor de guarda permanente para que a mesma possa ser mediada aos usuários que dela necessitar. As próximas seções abordarão os desafios apresentados pela preservação de registros arquivísticos digitais e as contribuições das linhagens de marcação para a produção documental.

8.2 A preservação de registros digitais: desafios, princípios e estratégias abordadas pela arquivística

Depois de apresentadas as características inerentes aos documentos digitais, surge a necessidade de prover uma abordagem sobre os conceitos envolvidos na preservação de registros arquivísticos.

O Arquivo Nacional (2005, p. 135) define preservação digital como procedimentos adotados que visam, sobretudo, a prevenção contra a deterioração e danos físicos ou químicos que possam ser causados na documentação.

Silva, S. (2008, p. 75) traz uma definição mais abrangente ao processo de preservação, e o define como “um conjunto de decisões estratégicas, que envolve opções políticas e tecnológicas como, por exemplo, a reformatação do suporte por meio da microfilmagem e documentação”.

Essa ação, segundo o autor, abarca um número abrangente de políticas e práticas operacionais que compreendem os tratamentos de aquisição, conservação, e operacionalização de recursos financeiros e humanos, que visem retardar o início do processo de deterioração do acervo documental.

O autor defende ainda que a essência do gerenciamento da preservação digital deverá ser focada na alocação de recursos eficientes em assegurar que seja dada a proteção adequada aos elementos informacionais. Isso configura o aspecto essencial na preservação dos arquivos institucionais, que é o estabelecimento de um gerenciamento capaz de prover o acesso à informação por longo prazo.

Para que os arquivos permanentes possam prover eficácia e satisfazer a necessidade informacional dos usuários, é imprescindível que sejam fornecidos meios para a preservação e as garantias de cumprimento dos requisitos inerentes à autenticidade e integridade do acervo em conjunto com a viabilização do acesso por longo prazo.

A complexidade existente na leitura dos registros digitais faz com que surjam desafios significativos no planejamento de ações que visem à preservação digital por longo prazo dos complexos procedimentos aplicados na interpretação dos formatos de arquivos digitais, alvos de rápida obsolescência tecnológica. Tais desafios serão abordados no tópico a seguir.

8.2.1 Desafios trazidos pela preservação de arquivos digitais

O *National Archives of Australia* (2002, p. 09) afirma que o documento arquivístico registrado em suporte tradicional possibilita interação direta entre usuário e o registro da informação documental. O acesso à informação ocorre, portanto, sem a intervenção de nenhuma tecnologia de interpretação dos dados ali registrados.

Por conseguinte, o foco da preservação está voltado ao objeto documental. Estando o suporte informacional preservado, a informação também estará preservada.

Os documentos de arquivo digitais, por sua vez, são constituídos apenas de uma fonte de registros binários que deverão ser interpretados por procedimentos de *hardware* e *software* para que o documento possa ser apresentado em sua forma original.

Surge, portanto, a necessidade de preservar o processo tecnológico que produziu o registro arquivístico, para que ele possa ser interpretado e apresentado ao usuário. A recomposição dos procedimentos de *hardware* e *software*, somada à correta configuração dos dispositivos de saída, é de fundamental importância para que a informação seja acessível ao usuário. Sem a preservação dos processos envolvidos na interpretação de documentos, os cuidados envolvidos na preservação das mídias tornam-se nulos. (FERREIRA, 2006; NATIONAL ARCHIVES OF AUSTRALIA, 2002)

Um fator agravante é a obsolescência dos equipamentos tecnológicos utilizados na produção dos registros digitais. Com o passar do tempo, os elementos de *hardware* são renovados em desempenhos de processamento e armazenamento. A concorrência acirrada existente no mercado de tecnologias, unificada ao forte incentivo do consumo de ferramentas tecnológicas, faz que os sistemas computacionais posteriores tornem-se completamente incompatíveis em curtíssimo prazo de tempo, fazendo com que a leitura dos registros seja incompatibilizada.

O *National Archives of Australia* (2006) comenta que, estando um determinado ambiente de *hardware* e *software* obsoleto, os registros digitais tornam-se órfãos e de difícil leitura. Este problema é agravado no decorrer do tempo fazendo que, em alguns casos, a perda da informação seja completa.

A obsolescência poderá englobar as mídias digitais e os formatos de arquivos digitais utilizados na leitura da informação.

8.2.1.1 Obsolescência e deterioração das mídias digitais

As mídias digitais, que configuram o suporte dos registros digitais, poderão se deteriorar no decorrer do tempo. Diferentemente do que ocorre nos suportes convencionais, a deterioração de parte da mídia resultará na perda completa da informação registrada, pois o rompimento de uma sequência de bits poderá impossibilitar a interpretação dos dados remanescentes. (SANTOS, V., 2005, 2007).

Outra questão a ser abordada é a obsolescência das mídias digitais no decorrer do tempo. O avanço no desenvolvimento de do *hardware* faz com que novos modelos de mídias e de leitores sejam incorporados rapidamente no mercado. A popularização de novas mídias digitais possibilita que o uso dos suportes anteriores torne-se inviável e, por conseguinte, a fabricação dos leitores para modelos obsoletos é descontinuada.

Pode ser mencionado para exemplificação o ocorrido com os discos flexíveis. Muito utilizados nas décadas de 1980 e 1990, os disquetes foram se desenvolvendo em capacidade de armazenamento e em complexidade tecnológica para a sua leitura. Eram indispensáveis em quaisquer ambientes computacionais e registravam os backups de informações orgânicas digitais de muitas instituições.

Atualmente, os leitores para este tipo de mídia encontram-se em avançado estágio de obsolescência, embora menos de trinta anos tenham transcorrido. Em termos de temporalidade arquivística, considera-se que os documentos registrados nessas mídias digitais estariam completando o ciclo de guarda vigente na fase intermediária, o que é um período de tempo arquivístico muito curto para que haja tamanha obsolescência de *hardware*.

8.2.1.2 Obsolescência de formatos de *software*

Mais rápida que o processo de *hardware* é a obsolescência dos formatos de arquivos digitais proprietários. Observa-se que as ferramentas de *software* apresentam lançamento de diferentes versões em curtíssimo prazo. Por conseguinte, as ferramentas lançadas posteriormente não possibilitam a leitura precisa de formatos de arquivos digitais produzidos por versões anteriores. Estima-se que em um período de tempo superior a cinco anos, muitos dos arquivos produzidos em formato digital estejam em um estágio avançado de

obsolescência. Neste curto prazo de tempo, os registros estarão por adentrar na fase de guarda intermediária, podendo se perder por completo.

Neste contexto, os direitos de propriedade industrial e intelectual em formatos de arquivos proprietários terminam por agravar a problemática da preservação dos procedimentos de leitura digital. Estando protegidos por direitos autorais, os formatos de arquivos proprietários não poderão ser manuseados por desenvolvedores de *software* sem que haja expressa autorização do produtor da ferramenta. (NATIONAL ARCHIVES OF AUSTRALIA, 2002 , 2006).

O produtor da ferramenta, embora possua todos os direitos relacionados ao seu uso, não é obrigado a produzir meios que mantenham os formatos digitais acessíveis ao longo do tempo. Os profissionais da informação que necessitem preservar o acesso aos documentos digitais por prazo indeterminado poderão encontrar-se em um grande impasse técnico e jurídico, a fim de que possam sanar a problemática envolta na preservação digital.

Com o avanço e a substituição do uso das tecnologias da informação dos procedimentos analógicos de produção documental, a informação orgânica registrada em documentos digitais está submetida ao risco de desaparecer sem deixar quaisquer vestígios de sua existência.

Preocupados com a questão envolta na preservação de registros digitais, profissionais e instituições da CI e TI têm se unido no desenvolvimento de métodos que possam assegurar a preservação de documentos digitais por longo prazo.

As discussões encaminharam-se no desenvolvimento de alguns princípios que devem ser observados nas práticas relacionadas à preservação dos registros digitais. Alguns dos princípios serão abordados a seguir.

8.2.2 Princípios a serem observados na preservação de acervos digitais

A fim de se minimizar os riscos relacionados à preservação de registros digitais, deverão ser observados alguns princípios que nortearão todos os procedimentos.

Um programa de preservação digital documental deverá embasar-se em políticas contínuas de ações que gerencie as estruturas e as cadeias relativas à preservação dos registros digitais. Essas políticas deverão apresentar:

- definição clara de escopo e objetivos;
- prover aquisição de recursos informacionais;

- ter o foco direcionado aos documentos digitais. (INTERPARES PROJECT; SANTOS,V., 2007).

A preservação digital deverá ser eficiente para manter os documentos arquivísticos sob custódia arquivística independentemente de aplicações ou sistemas específicos que os produziu e dos formatos de dados em que foram estruturados. Em outras palavras, não deverá haver qualquer dependência de ferramentas de *hardware*, *software* ou sistema de gestão específico. Os sistemas de preservação deverão, ainda, embasar-se em tecnologias não proprietárias para que o acervo não seja refém de um produtor de ferramenta de *software* (NATIONAL ARCHIVES OF AUSTRALIA, 2002; SANTOS, V., 2007).

Não se pode confiar cegamente em um suporte de armazenamento digital. Surge, então, a necessidade de constituir um programa de preservação que migre de suporte periodicamente. Também deverá determinar e preservar a essência dos documentos digitais sob a custódia arquivística, recreando a apresentação original ao longo do tempo. (NATIONAL ARCHIVES OF AUSTRALIA, 2002; SANTOS, V., 2007)

Para minimizar os riscos envoltos na integridade dos registros, o programa de preservação deverá reduzir o número de tratamentos de preservação, a saber, migração aplicados em cada um dos documentos digitais. Deve-se, ainda, manter o formato lógico mais antigo para que possam ser providos procedimentos de recuperação em caso de emergência. (NATIONAL ARCHIVES OF AUSTRALIA, 2002; SANTOS, 2007)

Uma vez abordados os principais princípios concernentes à preservação, apresentar-se-ão algumas estratégias para a preservação de registros arquivísticos digitais. Tais estratégias são medidas desenvolvidas na tentativa de se contornar os desafios impostos pela obsolescência tecnológica.

8.2.3 Estratégias para a preservação digital

As estratégias implementadas para prover a recuperação da informação são de fundamental importância em um processo de preservação digital. Sobre esta questão, Arellano (2004) comenta que “a aplicação de estratégias de preservação para documentos digitais é uma prioridade, pois sem elas não existiria nenhuma garantia de acesso, confiabilidade e integridade dos documentos a longo prazo.”

São quatro as estratégias de preservação a serem abordadas neste trabalho: migração, emulação, encapsulamento e modelo referencial OAIS.

8.2.3.1 Migração

A migração pode ser compreendida como a passagem de um documento de um suporte ou formato obsoleto para um formato lógico atualizado. A migração parece ser uma boa alternativa para a preservação de registro arquivísticos digitais. As críticas a ela referenciadas dizem respeito à continuidade de cópias sucessíveis que perderão informações relevantes na passagem de um formato para outro. (FERRREIRA, 2006; NATIONAL ARCHIVES OF AUSTRALIA, 2002).

8.2.3.2 Emulação

A emulação digital permite a reconstituição dos recursos de *hardware* e *software* em que os registros informacionais foram produzidos, permitindo a interação com o ambiente de *software* original para a produção documental. A emulação permite, por um lado, o acesso e manuseio do registro digital em seu ambiente original de operação. Exige-se do usuário, no entanto, que conheça o sistema original de produção documental para que possa haver interação. Os emuladores, no entanto, tendem a ficar obsoletos com o tempo, pois estarão diretamente vinculados a um sistema específico de *hardware* e *software*. (FERRREIRA, 2006; NATIONAL ARCHIVES OF AUSTRALIA, 2002).

8.2.3.3 Encapsulamento

O encapsulamento, por sua vez, permite a criação de pacotes de dados onde se encontram o objeto digital e a descrição de todos os recursos informacionais necessários a sua interpretação. Desta forma, os documentos digitais poderão ser compreendidos e interpretados quando for necessário. O encapsulamento trabalha em conjunto com migradores ou emuladores. (FERREIRA, 2006).

8.2.3.4 Modelo Open Archival Information System (OAIS)

O presente trabalho não poderia prosseguir sem uma breve explanação sobre o modelo conceitual do *Open Archival Information System* – Sistemas de informação de arquivo aberto (OAIS).

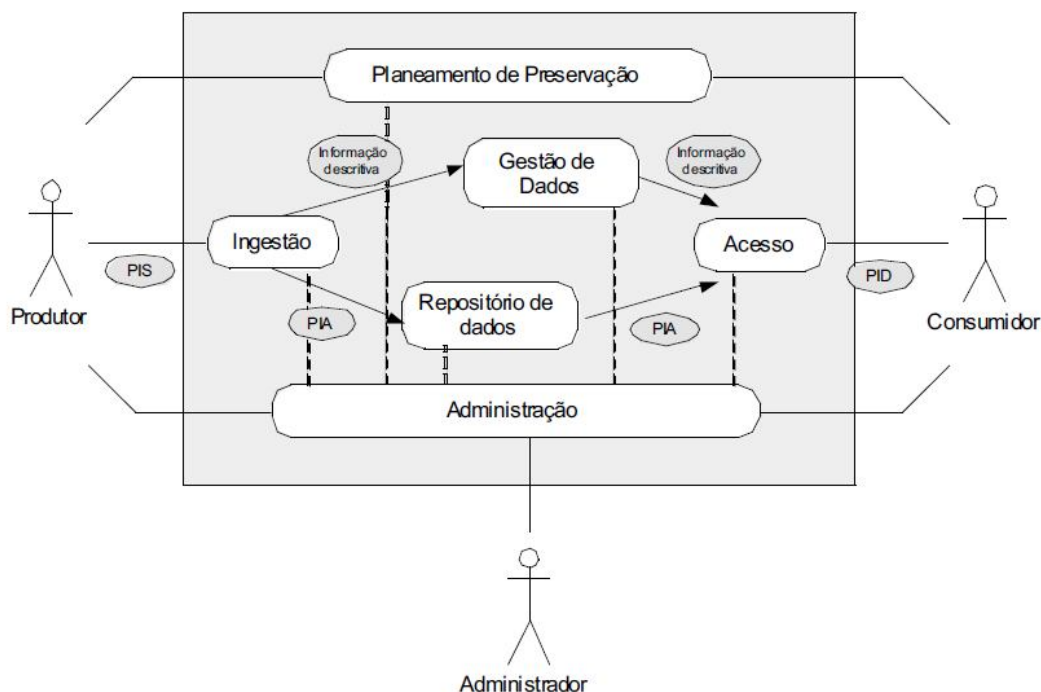
Sobre esse modelo, Santos, V. (2007, p. 37) comenta que:

O Open Archival Information System tem como propósito a definição de um modelo referencial (International Organization for Standardization) para o desenvolvimento de Sistemas Abertos de Informações em Arquivos. A aplicação do OAIS em arquivos consiste na organização de pessoas e sistemas, tendo como responsabilidade a preservação e o acesso da informação à comunidade interessada e

como foco principal a informação digital, as formas primárias de armazenamento e suporte da informação para os materiais de arquivos digitais e físicos.

Esse modelo constituiu uma norma ISO para a preservação de registros digitais e traz a seguinte estrutura:

Figura 19: Modelo Conceitual OAIS.



Fonte: Ferreira (2006, p. 28).

Observa-se que o modelo é composto dos seguintes componentes:

- Ingestão, que gerencia a entrada de registros digitais no acervo arquivístico e o encapsula em pacotes de metadados de submissão, que serão encaminhados ao repositório de dados;
- O repositório de dados, que armazena os pacotes arquivísticos, devendo ser ressaltadas todas as questões pertinentes à obsolescência;
- A gestão de dados; que gerencia as informações descritivas dos objetos existentes no repositório de dados; e
- Acesso, que promove a mediação da informação aos usuários do arquivo permanente.

Esses são os principais elementos existentes dentro do modelo conceitual OAIS.

Uma vez apresentadas algumas das estratégias relativas à preservação digital de documentos arquivísticos, apresentar-se-ão as contribuições que as linguagens de marcação apresentam para preservação da informação arquivística permanente digital.

8.3 As contribuições das linguagens de marcação para a preservação de documentos arquivísticos digitais

Neste contexto em que existe uma grande incerteza em relação ao futuro dos registros digitais, as linguagens de marcação apresentam recursos que poderão trazer impacto significativo nas questões inerentes à obsolescência de formatos de *software* digitais.

A linguagem XML, por exemplo, têm características que são fundamentais aos procedimentos de preservação. Conforme já mencionado, ela é estruturada em formato de texto, em código aberto, o que facilita questões relacionadas aos obstáculos interpostos ao acesso de um conteúdo binário em um formato de arquivo digital. A extensibilidade desta linguagem permite, ainda, a obtenção de estrutura para a representação de estruturas de formatos de arquivos digitais.

O uso das linguagens de marcação construídas a partir da XML poderá ser uma alternativa para contornar e solucionar muitas das questões concernentes à preservação de registros arquivísticos digitais.

O presente capítulo abordará o uso da linguagem XML na representação de formatos de arquivos digitais em três linguagens, a saber:

- XENA: *XML Electronic Normalization of Archives* – Normalização eletrônica de Arquivos em XML; e
- SIARD: *Software Independent Archival of Relational Database* – Software Arquivístico Independente para Bases de Dados Relacionais;
- DBML: *Database Markup Language* – Linguagem de Marcação para Banco de Dados.

Essas linguagens podem apresentar significativas contribuições à preservação de registros digitais durante a fase da guarda permanente.

8.3.1 A transcrição de formatos de arquivos digitais para XML com a Ferramenta XENA

O Arquivo Nacional da Austrália, desenvolvedor do XENA, trabalha com o armazenamento de documentos digitais que remontam à década de 1960. Inicialmente, as mídias digitais eram gerenciadas da mesma maneira que os suportes tradicionais. Percebeu-se, no início da década de 1990, que mesmo estando devidamente catalogadas, havia grande dificuldade em recuperar o conteúdo das fitas computacionais devido à obsolescência de *hardware* e *software*.

Preocupado com essa questão, a instituição promoveu uma série de discussões a respeito do gerenciamento de documentos eletrônicos na década de 1990. (NATIONAL ARCHIVES OF AUSTRALIA, 2004).

No ano de 2002, a instituição consolidou um modelo de preservação que utilizaria XML como base para a normalização de formatos digitais proprietários, sendo em 2005 o lançamento da primeira versão da ferramenta XENA, que vem sendo aprimorada no decorrer do tempo.

A ferramenta XENA, portanto, permite a migração de documentos arquivísticos eletrônicos de formatos proprietários para o formato aberto XML, possibilitando o controle dos recursos tecnológicos empregados no processamento de apresentação do conteúdo informacional.

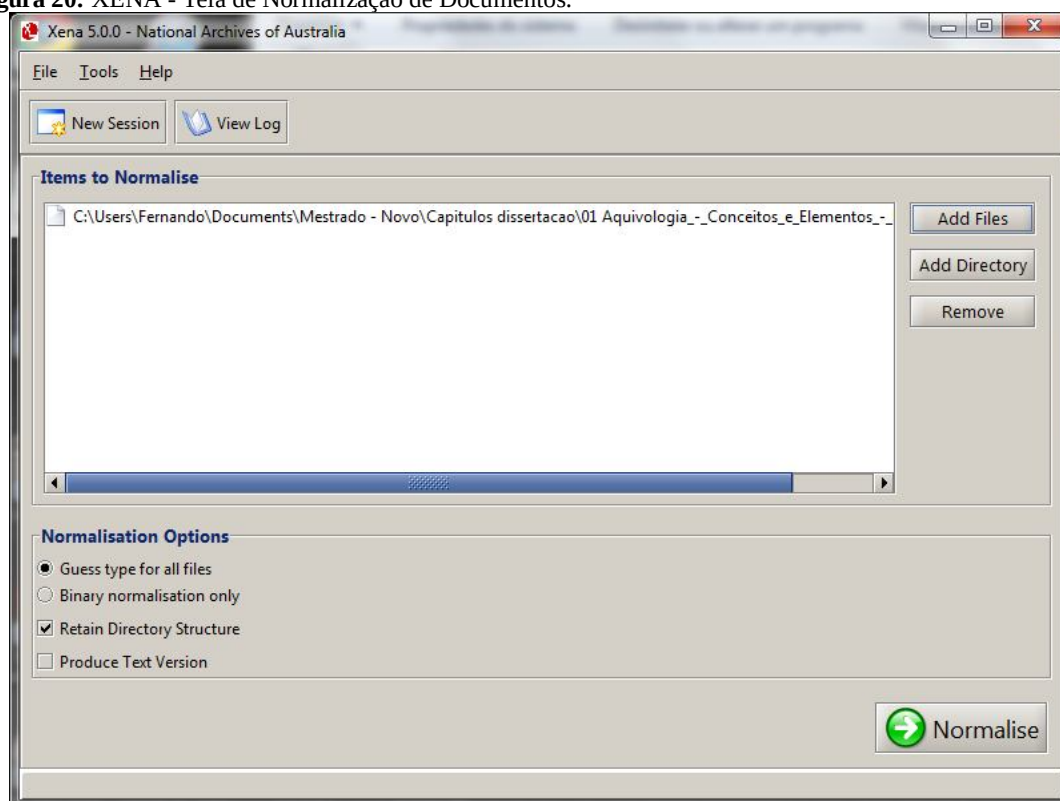
A versão atual, que é a 5.0, permite a conversão de imagens, documentos textuais e sonoros que foram compostos em formatos do Microsoft Office, Musicas (MP3, WAV, etc.), imagens (JPG, PNG, TIFF, etc.).

A ferramenta XENA apresenta também a possibilidade de conversão binária. Neste caso específico, a cadeia binária será representada em XML, mas necessitará da ferramenta que produziu o registro para que possa ser lido. A conversão binária em XML é útil para a retenção completa dos dados existentes no formato de arquivo original, que poderão se perder no processo de normalização e para resguardar uma cópia em XML de formatos que poderão ser normalizados futuramente. Apresentado o contexto de produção da ferramenta XENA, far-se-á uma explanação a respeito do funcionamento da ferramenta.

8.3.1.1 O funcionamento da ferramenta XENA

A primeira tela apresentada é a tela de normalização, que permite a inicialização e execução do processo de conversão de formatos.

Figura 20: XENA - Tela de Normalização de Documentos.



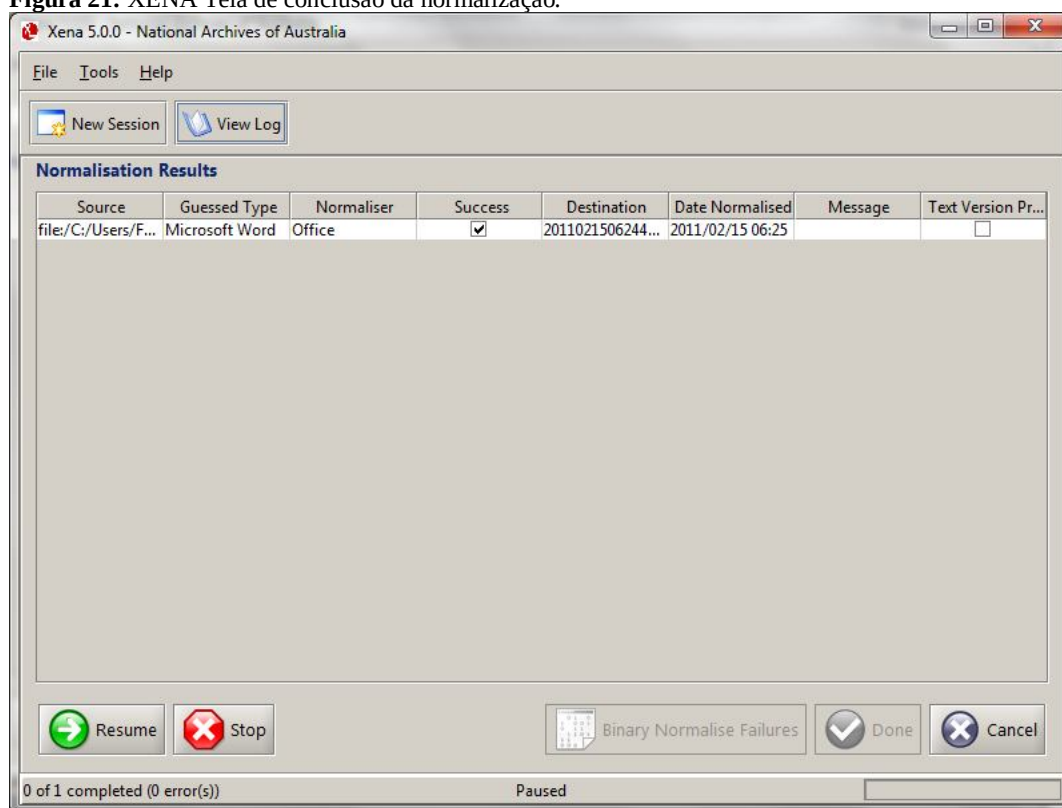
Fonte: O autor.

O Botão *New Session* permite abrir uma nova sessão de normalização. O Botão *Add Files* permite adicionar documentos individuais, enquanto o botão *Add Directory* permite adicionar pastas de documentos.

O tipo de formato a ser convertido, binário ou normalizado, poderá ser escolhido no campo *Normalization Options*. A partir da seleção da opção *Binary normalisation only* o documento será convertido para o modo binário. Já a opção *Guess type for all files* permite a conversão no formato normalizado.

O botão *Normalise* dará início ao processo de conversão e, depois de finalizado, avisará que a conversão foi concluída, conforme apresentado na figura a seguir:

Figura 21: XENA Tela de conclusão da normalização.



Fonte: O autor.

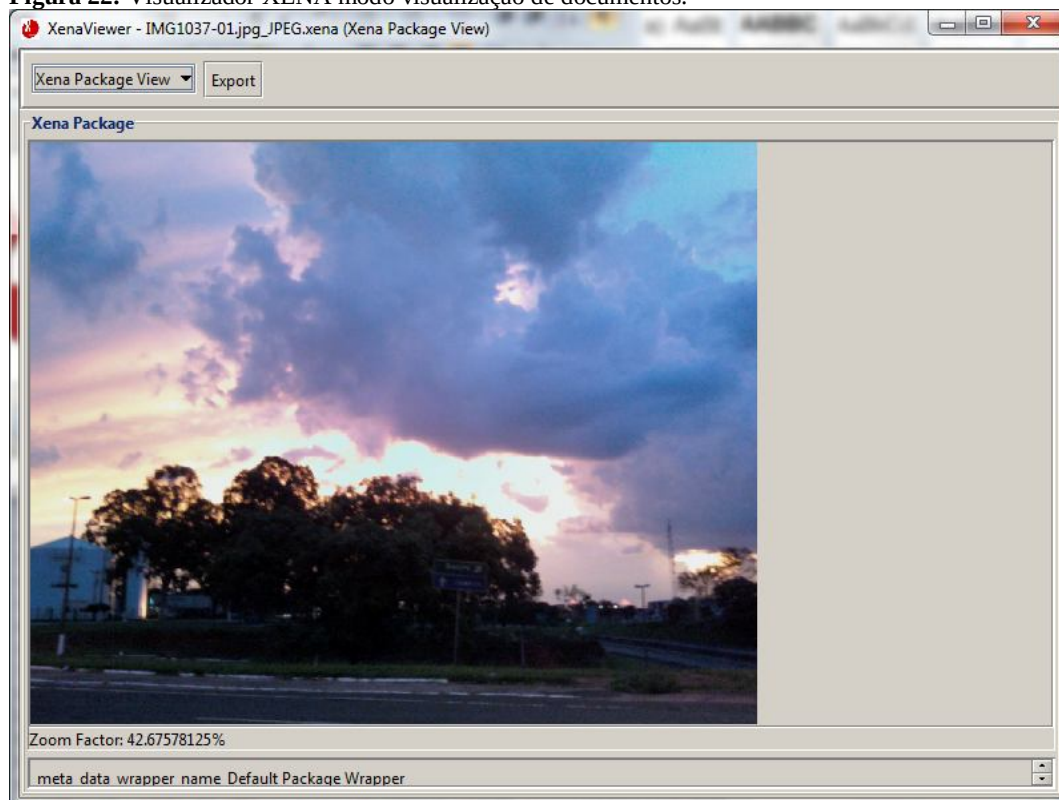
Essa tela informa se a conversão foi bem sucedida, o tipo de normalizador e possibilita ao usuário a visualização do conteúdo existente no registro documental; caso o formato de conversão esteja normalizado.

O Formato XML Normalizado permite três representações no visualizador, a saber:

- Apresentação original do documento;
- visualização da árvore de dados XML; e
- visualização de XML bruto.

A figura a seguir traz a apresentação de um registro fotográfico, em formato JPG convertido para XML

Figura 22: Visualizador XENA modo visualização de documentos.

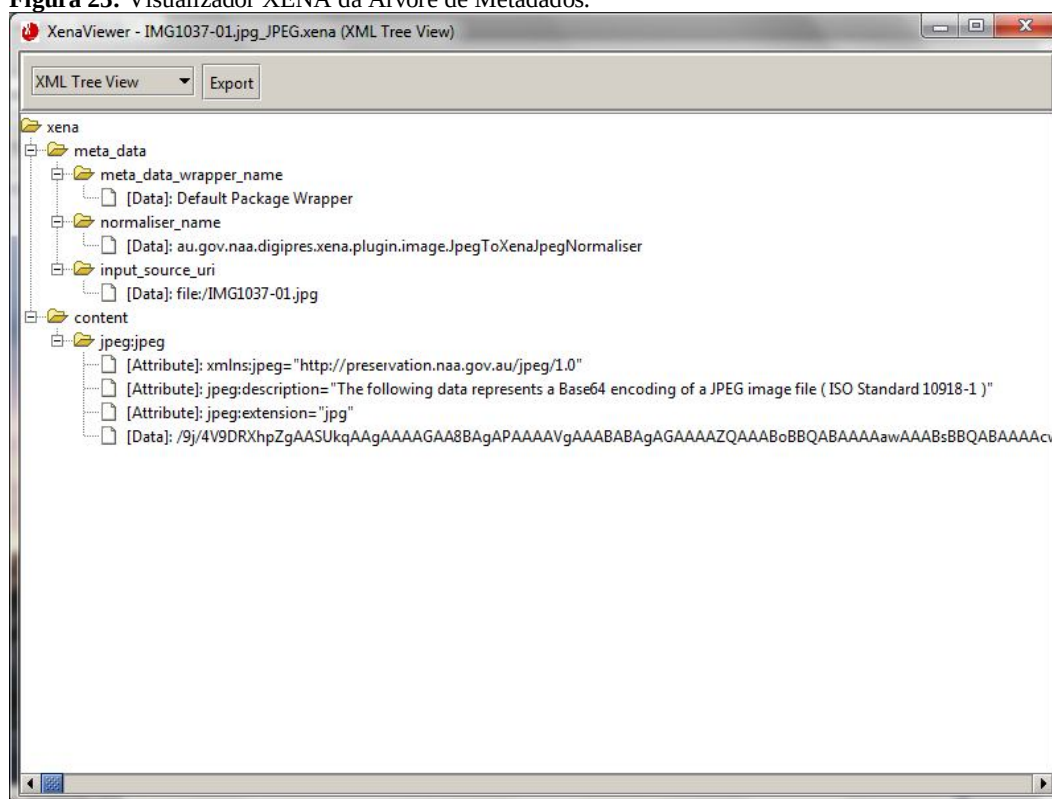


Fonte: O autor.

Observa-se que o documento é visualizado de forma semelhante ao aplicativo original. A apresentação original do registro documental é mantida, mas agora migrada para um formato de arquivo aberto que traz uma série de metadados que identificam o ambiente de *software* que produziu o documento original.

Tais metadados são apresentados na opção *XML Visualização da Árvore de Metadados*, disposto na figura a seguir:

Figura 23: Visualizador XENA da Árvore de Metadados.



Fonte: O autor.

Essa tela permite a identificação de cada um dos elementos de metadados existentes no formato normalizado. Por conseguinte, é facilitada a compreensão da estrutura do formato de *software*, bem como dos padrões utilizados para a composição do mesmo.

Esses metadados são de grande relevância para a preservação dos registros arquivísticos digitais, uma vez que denominam o registro documental e os recursos utilizados para a composição da estrutura do formato.

A visualização do XML bruto, por fim, apresenta o documento XENA em sua forma primária, a saber, o formato texto, conforme o disposto na figura a seguir:

Figura 25: Arquivo XENA Normalizado.

```

<?xml version="1.0" encoding="UTF-8"?>
<xena>
<meta_data>
<meta_data_wrapper_name>Default Package
Wrapper</meta_data_wrapper_name><normaliser_name>au.gov.naa.digipre
s.xena.plugin.image.JpegToXenaJpegNormaliser</normaliser_name>
<input_source_uri>file:/IMG1037-01.jpg</input_source_uri>
</meta_data>

<content>
<jpeg:jpeg xmlns:jpeg="http://preservation.naa.gov.au/jpeg/1.0"
jpeg:description="The following data represents a Base64 encoding
of a JPEG image file ( ISO Standard 10918-1 )"
jpeg:extension="jpg">
/9j/4V9DRXhpZgAASUkqAAgAAAAGAA8BAGAPAAAVgAAABABAgAGAAAZQAAABoBBQA
BAAAAawAAABsBBQABAAACwAAACgBAwABAAAAAgAAAGmHBAABAAAmAAAAKoAAABMR1
9FbGVjdHJvbmljcwBLRjMwMABIAAAAAQAAAEgAAAAAABAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAABAACQBwAEAAAAMDiyMAAAAAACAAECBAABAAAAyAAAAICBAAB
AAAC14AAAAAAD/2P/bAEMAAgEBAQEBAgEBAQICAgICBAMCAgICBQQEAWQGBQYGBGU
GBgYHCQgGBwkHBgYICwgJCgoKCgoGCAsMCwoMCQoKCv/bAEM
[...] conteúdo suprimido para exemplificação
ek/clf8Ag9a19nJ04ouV9S87TXcMqtp6tNGo/eKf4aJ5VtbXY9mz7l+QlqqpqzwW/km
fLhsuM4JFLd69ZahblU2DEu1mbop/nws71HyYkXFRGyPK0qmcEM2BG7cKR/jU895ELv
yw7M8KjzFX/wBCqiWeXUV89nLbPubvk+oqeKbUJZftdhsG/wD1rvWd5wQlCPMxl/Fcz
WqQWd5LDIZV2NweM90auWrvgFmVzFj01ZDye/FUZKZoH3MjF0qE9/WoNS1Bd1utrMd0
Uiq28dV/i/GsYwkm6typ2gjYglnSdLhxNvk6IfWor42y3P2gwEs5x/eqlPqkt+0lqsw
8rJwF+/j1p0Vw5idolTeR8jS1cas1F9Sebu//2Q==
</jpeg:jpeg>
</content>
</xena>

```

Fonte: O autor.

É possível observar que o formato XENA possui duas grandes estruturas de metadados, a saber:

- <meta_data> Metadados, que identifica o normalizador utilizado na conversão e traz os dados do registro documental, ex. o nome do arquivo.
- <content> Conteúdo, que abriga a conversão propriamente dita e identifica a extensão de formato de arquivo convertida.

Estes campos poderão apresentar elementos que descrevam a composição do formato, com remissivas às normas de sua estrutura. Neste caso específico, o elemento conteúdo faz uma remissiva à norma ISO 10918-1, que normatiza a construção de uma codificação em arquivo JPG, sendo de grande relevância para a compreensão da normalização XENA.

O elemento metadados identificou o normalizador utilizado e registrou o nome do arquivo original, embora também possa registrar outros metadados como a data de criação, o nome do produtor e etc.

8.3.1.3 Considerações sobre a iniciativa australiana

Essa iniciativa do Arquivo Australiano traz contribuições significativas à Ciência da Informação, Arquivologia e Tecnologias da Informação, pois visa ao desenvolvimento de padrões sólidos que possam suportar a demanda de preservação digital por longo prazo.

Cumprе ressaltar que a normalização de formatos de arquivos digitais em XML não é a solução definitiva para a problemática concernente à preservação, pois os sistemas conversores e visualizadores parecem ser alvo da obsolescência. Um fator positivo, no entanto, é a documentação produzida em torno da ferramenta aberta, que poderá guiar desenvolvedores de *software* na composição de novas ferramentas de leitura compatível com os sistemas operacionais de vigência futura.

Quando essas iniciativas se consolidarem em padrões robustos para a preservação digital, ter-se-á desenvolvido uma estrutura metodológica eficiente para sanar por completo, ou significativamente, as questões concernentes à obsolescência de *software* em arquivos permanentes digitais.

Isso representará um grandioso avanço para a Ciência da Informação e a Arquivologia, que têm a grande necessidade de unir-se com os profissionais e pesquisadores da Tecnologia da Informação no desenvolvimento de métodos eficientes para o tratamento da informação orgânica digital.

Depois de apresentado o formato XENA, segue-se para a explanação sobre o formato SIARD e a linguagem de marcação DBML.

8.3.2 O Formato SIARD e a Linguagem DBML enquanto alternativas para a transcrição de bases de dados em XML

Uma vez apresentadas as contribuições do XENA para a preservação de documentos digitais, seguir-se-á uma explanação a respeito das linguagens de marcação para a preservação de bancos de dados.

Para atender aos objetivos desta pesquisa, será dado foco às linguagens de marcação apresentados a seguir:

- *Software Independent Archival of Relational Database* (SIARD) – *Software* Arquivístico Independente de Banco de Dados relacionais; e
- *Database Markup Language* (DBML) - Linguagem de marcação de banco de dados.

Após serem indicadas as linguagens utilizadas para a preservação de banco de dados, far-se-á uma breve explanação a respeito de cada uma delas.

8.3.2.1 A preservação de banco de dados com SIARD

O formato SIARD foi desenvolvido pela *Swiss Federal Archives* (SFA) e pode ser utilizado na conversão de bases de dados relacionais — tais como Oracle, SQL Server e MS Access para a linguagem XML. Esse formato embasa-se ainda na linguagem de banco de dados SQL de 1999.

O SIARD apresenta também uma ferramenta de conversão, que está em conformidade com o modelo OAIS, permitindo a geração de pacotes de submissão (SIP), arquivamento (AIP) e disseminação (DIP).

A composição de um arquivo SIARD ocorre no formato *zip*, em que há um arquivo XML *Schema* de metadados e um diretório de arquivos que abrigam os conteúdos declarados no registro de metadados.

O arquivo de metadados registra, portanto, a estrutura do banco de dados, que inclui:

- *Schema Level Metadata* - Esquema da base de dados, com informações globais sobre a base;
- *Table Level Metadata* - Metadados de tabela, registra a estrutura de uma determinada tabela;
- *Column Level Metadata* - Metadados de tabela, registra a estrutura dos campos de uma tabela;
- *Primary Key Metadata* – Metadados da chave primária, registra o campo da chave primária;
- *Foreign Key Metadata* – Metadados da chave estrangeira, registra o campo da chave estrangeira;
- *Trigger Level Metadata* – Metadados de ações, registra as ações que são executadas no banco de dados;
- *User Level Metadata* – Metadados de Usuários, registra informações sobre os usuários do banco de dados; e
- Etc. (SWISS FEDERAL ARCHIVES, 2008).

O arquivo de metadados produzido pelo SIARD traz ampla descrição da base de dados, dos procedimentos nela declarados e também nos dados sobre cada um dos sistemas. Cada elemento recebe campos em que são declarados subelementos que possibilitam a

identificação e descrição dos recursos, de modo que seja possível restabelecer, sempre que necessário, os procedimentos para a recuperação da informação orgânica permanente registrada na base.

Este eficiente formato foi adotado, recentemente, pelo projeto *Planets*, que promove a preservação digital das bases de dados de instituições europeias. (SWISS FEDERAL ARCHIVES, 2008).

8.3.2.2 A transcrição de banco de dados com DBML

A linguagem de modelagem de dados DBML permite a representação do conteúdo e estrutura existente em uma base de dados.

Seguindo o formato SIARD, a linguagem de marcação DBML permite a divisão dos elementos existentes em um banco de dados em estrutura e conteúdo (FREITAS, RAMALHO, 2009; RAMALHO ET. AL, 2007).

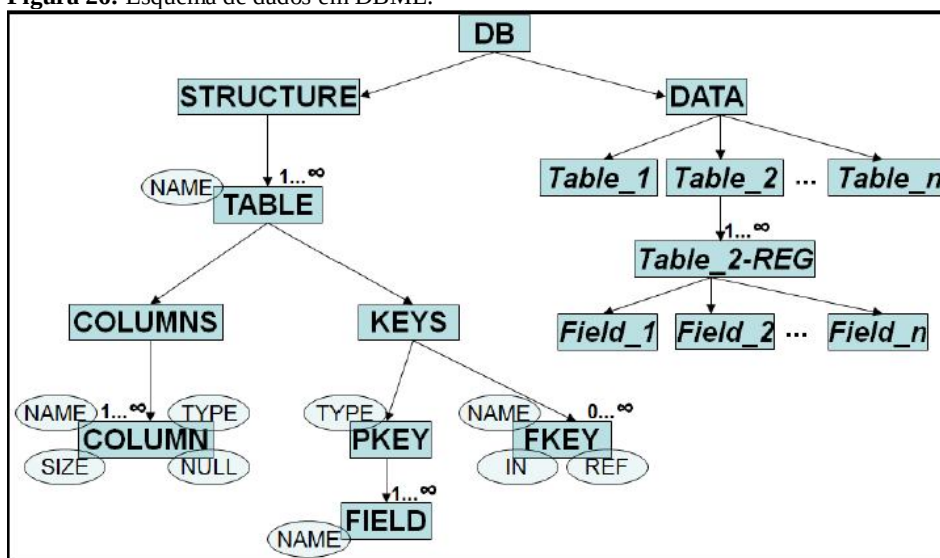
A estrutura registra informações a respeito dos seguintes elementos:

- *Tables* – Tabelas – traz o nome da tabela;
- *Columns* – Colunas – traz informações de cada um dos campos (*field*) existentes na tabela tais como tamanho, nome e tipo de dados armazenados;
- *Keys* – Chaves – traz informações a respeito das chaves existentes na tabela podendo ser primária ou estrangeira.

O conteúdo, por sua vez registrará os dados concernentes aos elementos declarados na estrutura.

Freitas e Ramalho (2008, p. 03) afirmam que essa divisão permite prover a contextualização dos dados unificada a descrição de cada um dos componentes existentes na estrutura do banco, e apresentam o esquema de estrutura e de dados disposto na figura a seguir:

Figura 26: Esquema de dados em DBML.

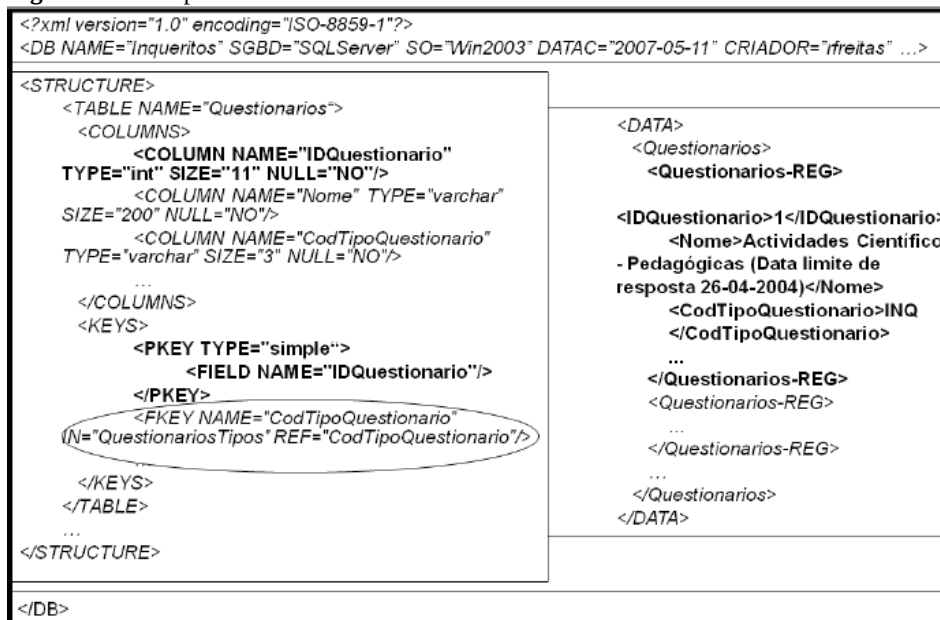


Fonte: Ramalho (2008, p. 04).

É possível observar nessa figura a composição da linguagem DBML no que tange à composição dos elementos concernentes à estrutura (*structure*) e aos dados (*data*). O vínculo entre ambos, ao invés de se desintegrar, é consolidado por meio da definição estrutural, e também é possibilitada a integração com esquemas de metadados EAD, para a representação da informação orgânica.

Os autores também apresentam uma exemplificação em DBML, que será apresentada na figura a seguir:

Figura 27: Exemplo em DBML.



Fonte: Ramalho (2008, p. 4).

É possível observar na figura a transcrição do esquema no texto em XML. Verifica-se que a estrutura e conteúdo do banco de dados são transcritos no arquivo XML, possibilitando a leitura por pessoas e máquinas.

O registro de banco de dados com valor de guarda permanente possibilita que o conteúdo informacional ali existente seja conservado por longo período de tempo livre de do vínculo direto com as tecnologias proprietárias que os gerou.

Por conseguinte, os sistemas de arquivo que necessitem efetuar a guarda permanente de registros orgânicos nascidos e gerenciados em bancos de dados digitais poderão utilizar destas linguagens, enquanto recursos para a ampliação da vida útil dos registros arquivísticos digitais de guarda permanente presentes no banco de dados.

As linguagens de marcação para a representação de bancos de dados solidificam alternativas para a preservação de bases de dados digitais. Assim como o XENA, elas poderão apresentar significativas contribuições para que seja provida a guarda de bases de dados por longo prazo. Certamente, apresentam-se como uma alternativa de solução frente à problemática da perda de informação devido à obsolescência em formatos proprietários de base de dados.

8.4 Considerações

O presente capítulo abordou de modo sucinto parte dos problemas concernentes à preservação digital. A obsolescência de *software* tem feito que muitas das informações registrada nos sistemas arquivísticos digitais corram forte risco de destruição permanente.

As linguagens de marcação figuram-se, neste contexto, como uma alternativa à preservação de documentos de arquivos digitais permanentes.

Isso se dá porque a flexibilidade da XML permite a construção de formatos digitais abertos que possibilitam a representação de formatos de arquivos digitais.

Observa-se na literatura que essa não é uma questão definida na arquivologia. É necessário um trabalho conjunto de pesquisadores e profissionais da CI e TI, a fim de aprimorar e desenvolver metodologias e ferramentas sólidas para a preservação digital. Nesse contexto a XML apresenta uma estrutura sólida ao desenvolvimento de linguagens para a representação de formatos de arquivo digitais de caráter proprietário, possibilitando o delineamento de alternativas plausíveis à preservação digital.

CONSIDERAÇÕES FINAIS

O presente trabalho buscou estabelecer um diálogo entre a Arquivologia, a Ciência da Informação e a Tecnologia de Informação, inter-relacionando as linguagens de marcação com a gestão de documentos e a preservação digital, visando discutir meios de benefícios aos usuários e aos acervos arquivísticos.

O delineamento deste trabalho buscou responder alguns questionamentos apresentados na introdução, que indagavam acerca das contribuições que as linguagens de marcação poderiam trazer à arquivística, dando-se enfoque aos relacionamentos entre a web semântica e a gestão de documentos, bem como da linguagem XML em suporte à preservação digital.

As indagações realizadas a partir do levantamento bibliográfico permitiram a compreensão do papel da arquivologia dentro da Ciência da Informação, que está voltado ao tratamento da informação orgânica que, por sua vez, está registrada nos documentos de arquivo. Verificou-se que os registros documentais podem produzir uma gama de metadados que representam a informação registrada e cumprem com os princípios norteadores da área. Constatou-se que as linguagens de marcação são um meio eficiente para o registro de metadados, pois promovem uma série de recursos para a representação da informação.

Realizou-se uma análise do comportamento informacional do usuário de arquivos correntes em que tornou-se evidenciado que a necessidade de informação se dá em torno das atribuições inerentes ao cargo funcional dos indivíduos, e sua inter-relação com o conjunto de funções e atividades institucionais. Delineou-se ainda o fluxo de informações a que são submetidos os documentos arquivísticos, visando ao atendimento das demandas.

Depois de apresentada a necessidade informacional do usuário do arquivo corrente, tornou-se possível a explanação do programa de gestão documental e das ferramentas de classificação documental em que o controle de vocabulário poderá ser empregado, a saber, o plano de classificação e os tesouros de função.

A partir da colocação destas indagações, foi possível apresentar as contribuições que as linguagem de marcação XML e a modelagem das linguagens da web semântica podem apresentar ao gerenciamento de documentos e da informação arquivística. Tais contribuições consistem na representação dos instrumentos de classificação documental no ambiente digital, possibilitando que agentes de *software* automatizem todo o processo de recuperação da informação orgânica registrada no acervo documental com maior precisão e confiabilidade.

O desenvolvimento de tesouros de função e planos de classificação nas linguagens de marcação da web semântica além de aprofundar as relações existentes entre arquivologia e TI, trazem maiores oportunidades para o aprimoramento da área de arquivos no que diz respeito ao estudo de usuário e as demandas informacionais impostas ao acervo. Isso permite amplo desenvolvimento e amadurecimento da área.

Por fim, abordou-se a problemática envolta na preservação de formatos de arquivos digitais proprietários, em que a linguagem de marcação apresentou-se como uma opção recomendável à representação destes códigos binários em formato de arquivo aberto. Isso traz profundos reflexos nas questões inerentes à migração de formatos de softwares em programas de preservação digital, pois poderia estancar os constantes processos migratórios preservando-se com maior fidelidade a apresentação original dos documentos digitais.

É importante salientar que o presente trabalho está muito longe de esgotar a complexa temática inerente às contribuições das linguagens de marcação para os documentos de arquivo, apesar de evidenciar o potencial das linguagens de marcação frente ao tratamento da informação orgânica.

Observa-se que nos últimos anos o uso das tecnologias da informação se consolidou na produção e gerenciamento do acervo documental arquivístico. A era dos documentos digitais ampliou, por conseguinte, a necessidade de observância dos princípios da arquivologia, bem como dos métodos empregados no tratamento da informação orgânica em todos os estágios do ciclo de vida documental. Isso é necessário para se assegurar a autenticidade e a integridade do acervo, bem como a continuidade de acesso por longo prazo.

O uso das linguagens de marcação pode vir a se consolidar como um dos principais requisitos para a provisão do tratamento documental arquivístico no ambiente digital. É necessário, portanto, que a arquivologia desenvolva, em parceria com os pesquisadores de TI, profundos estudos e reflexões voltados à aplicação das diversas formas de utilização dos recursos inerentes às linguagens de marcação em prol do tratamento arquivístico de documentos digitais.

REFERÊNCIAS

ALMEIDA, M. B. Uma introdução ao XML, sua utilização na internet e alguns conceitos complementares. **Ciência da Informação**, Brasília, v. 31, n. 2, p. 5-13, maio/ago 2002.

ALMEIDA JÚNIOR, O. F. Mediação da Informação e Múltiplas Linguagens. **Tendências da Pesquisa Brasileira em Ciência da Informação**, João pessoa, v.2, n. 1, 2009. Disponível em: <http://inseer.ibict.br/ancib/index.php/tpbci/article/view/17/39>. Acesso em 20 dez. 2009.

ANDONEGI, A. R. Nuevas herramientas para el estudio de la diplomática: los lenguajes de marcado TEI y EAD. Documenta et Instrumenta, Madrid, v.6, 2008. Disponível em: http://europa.sim.ucm.es/compludoc/GetSumario?zfr=0&r=/S/10810/16974328_6.htm. Acesso em: 01 fev 2009.

ANDRADE, R. S. Aspectos introdutórios da representação de informação arquivística: a Norma Brasileira de Descrição Arquivística (NOBRADE), a Descrição Arquivística Codificada (EAD-DTD) e o projeto Archives Hub. **Ponto de Acesso**, Salvador, v. 1, n. 2, 2007. Disponível em: < <http://www.portalseer.ufba.br/index.php/revistaici/article/view/1589> > Acesso em: 01 fev. 2009.

_____. **Uma nova geração de instrumentos arquivísticos de referência na web: novas possibilidades para as instituições arquivísticas públicas brasileiras** [manuscrito] / Ricardo Sodré Andrade. – 2010.87f. Disponível em < <http://ricardo.arquivista.net/wp-content/uploads/2010/06/DissertacaoRicardoSodreAndrade-2010.pdf> > Acesso em dez. 2010

ANTONIOU, G.; HARMELEN, F. **A semantic web primer**. Massachusetts:MIT, 2004.

ARAUJO JUNIOR, R. H. **Precisão no processo de busca e recuperação da informação**. Brasília:Thesaurus, 2007

ARCHIVES PORTAL EUROPE. **EAG as used within the Archives Portal Europe for describing archival institutions** – Description and best practice guide. 2011. Disponível em <http://www.apenet.eu/images/docs/apenet_eag_archival_institution.pdf>. Acesso em dez. 2010.

ARELLANO, M. A. Preservação de documentos digitais. **Ciência da Informação**, Brasília, 2004, vol.33, n.2, pp. 15-27.

ARELLANO, M. A. M.; ANDRADE, R. S. Preservação digital e os profissionais da informação. **Datagramazero**, Rio de Janeiro, v. 7, n. 5, out. 2006. Disponível em: <http://www.dgz.org.br/out06/Art_05.htm>. Acesso em 01 fev 2009.

BANTIN, P. C. **Understanding data and information systems for recordkeeping**. New York:Neal-Schhman Publishers, 2008.

BARBEDO, F. Arquivos Digitais: da origem à maternidade. **Cadernos BAD**, Lisboa, v.2, p. 07-17, 2002.

BARRETO, A. A. Mitos e lendas da informação: o texto, o hipertexto e o conhecimento. **Datagramazero**, Rio de Janeiro, v. 8, n. 1, fev. 2007. Disponível em < http://www.dgz.org.br/fev07/Art_02.htm >. Acesso em: 01 fev. 2009

BEAL, A. **Gestão estratégica da Informação**: como transformar a informação e a tecnologia da informação em fatores de crescimento e de alto desempenho nas organizações. São Paulo: Atlas 2009.

BELLOTO, H. L. **Arquivos permanentes**: tratamento documental, 4. ed. Rio de Janeiro: FGV, 2006.

_____. **Arquivística**: Objetos, Principios e Resumos. São Paulo: Associação dos Arquivistas, 2002.

_____. **Diplomática e tipologia documental em arquivos**. Brasília: Briquet de Lemos, 2008.

BONAL ZAZO, J. L. **La descripción archivística normalizada**: origen, fundamentos, principios y técnicas. Gijón, Asturias : Ediciones Trea, 2001

BORBINHA J. L. et al. Manifesto para a preservação digital. **Cadernos BAD**, Lisboa, v.2, p 54-67, 2002.

BORKO, H. Information Science: What is it? **American Documentation** [s:l], 1968

BRASIL. ARQUIVO NACIONAL. **Dicionário Brasileiro de Terminologia Arquivística**, Rio de Janeiro: Arquivo Nacional, 2005. Disponível em: <<http://www.portal.arquivonacional.gov.br/Media/Dicion%20Term%20Arquiv.pdf> > Acesso em 01 fev. 2009.

_____. Resolução n. 20 de 16 de julho de 2004. Dispõe sobre a inserção de documentos digitais em programas de gestão arquivística de documentos dos órgãos e entidades do Sistema Nacional de Arquivos. Diário Oficial. Brasília, 19 de julho de 2004.

_____. Resolução n. 24 de 24 de agosto de 2006. Estabelece diretrizes para a transferência e recolhimento de documentos arquivísticos digitais para instituições arquivísticas públicas. Diário Oficial. Brasília, 07 de agosto de 2006.

_____. Resolução n. 25 de 27 de abril de 2007. Dispõe sobre a adoção do Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos – e-ARQ Brasil pelos órgãos e entidades integrantes do Sistema Nacional de Arquivos - SINAR. Diário Oficial da União. Ed. 81, 27 de abril de 2007 – Seção 1

_____. CÂMARA TÉCNICA DE DOCUMENTOS ELETRÔNICOS. **Carta para a preservação do patrimônio arquivístico digital**. Disponível em: <<http://www.conarq.arquivonacional.gov.br/Media/publicacoes/cartapreservpatrimarqdigitalconarq2004.pdf>> . Acesso em: 01 fev. 2009.

_____. CONSELHO NACIONAL DE ARQUIVOS. 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 01 fev. 2010.

_____. CONSELHO NACIONAL DE ARQUIVOS. Câmara Técnica de Documentos Eletrônicos. **e-ARQ Brasil: Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos**, Rio de Janeiro: Arquivo Nacional, 2009. Disponível em: <<http://www.conarq.arquivonacional.gov.br/media/publicacoes/earqmet/earqbrasilv1.1.pdf>> Acesso em: 01 fev. 2010.

BREITMAN, K. K. **Web Semântica: a internet do futuro**, Rio de Janeiro: LTC 2005.

BRIET, S. What is documentation? [s.l]: Scarecrow, 2006

BULLOCK, A. **Preservation of digital information: issues and current status**. Disponível em: <<http://epe.lac-bac.gc.ca/100/202/301/netnotes/netnotes-h/notes60.htm>>. Acesso em: 01 fev. 2009.

CARVALHO, R. A.; ZANAGA, M. P. Perspectivas da *Web* semântica para a Ciência da Informação. In ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. **Anais...** São Paulo: ANCIB, 2008.

CERVANTES, B. M. N. **Terminologia do Processo de Inteligência Competitiva: estudo teórico e metodológico**. Londrina:Eduel, 2007.

CHO, A; GIUSTINI, D. *Web 3.0 and health librarians: An introduction*. **Journal of the Canadian Health Libraries Association**, 29, 2008, 13-18. Disponível em: <<http://pubs.nrc-cnrc.gc.ca/jchla/jchla29/c07-035.pdf>> Acesso em 01 fev. 2009.

CONSELHO INTERNACIONAL DE ARQUIVOS. **ISAD (G): Norma Internacional de Descrição Arquivística**, 2. ed. Rio de Janeiro: Arquivo Nacional, 2003. Disponível em: <<http://www.portalan.arquivonacional.gov.br/Media/ISADG.pdf>> Acesso em 01 jan. 2010.

_____. **ISAAR (CPF): norma internacional de registro de autoridade arquivística para entidades coletivas, pessoas e famílias**, 2. ed. Rio de Janeiro: Arquivo Nacional, 2004. Disponível em <<http://www.portalan.arquivonacional.gov.br/Media/ISAAR%20Brasil%20final.pdf>> Acesso em 01 jan. 2010.

_____. **ISDIAH: Norma Internacional para descrição de instituições com acervo arquivístico**. Rio de Janeiro: Arquivo Nacional, 2009. Disponível em: <<http://www.portalan.arquivonacional.gov.br/media/isdiah%20final09.pdf>> Acesso em 01 jan. 2010.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS). **Reference model for an Open Archival Information System (OAIS)**, Washington:CCSDS, 2002. 139p. Disponível em: <<http://public.ccsds.org/publications/archive/650x0b1.pdf>> Acesso em: 01 fev. 2009.

CURTY, R. G; GAMA F. A. Conjugando diplomática e XML: aproximação possível no contexto da proveniência de documentos jurídicos digitais. **Arquivistica.net** v. 3, n. 2, 2007 Disponível em: <<http://www.arquivistica.net/ojs/viewarticle.php?id=149>> acesso em dez. 2010

DACONTA, M. C.; OBRST, L. J.; SMITH, K. T. **The Semantic Web: A guide to the future of XML, Web Services, and Knowledge Management**, Indianapolis: Wiley Publishing, 2003.

DAUM, B. **Arquitetura de sistemas com XML**: conteúdo, processo, e apresentação. Rio de Janeiro: Campus, 2002

DIAS, M. P.; CARVALO, J. O. F. A visualização da informação e sua contribuição para a Ciência da Informação. **Datagramazero**, Rio de Janeiro, v. 8, n. 5, out. 2007. Disponível em < http://www.dgz.org.br/out07/Art_02.htm >. Acesso em: 01 fev. 2009

DZIEKANIAK, G. V.; KIRINUS, B. J. *Web Semântica*. **Encontros Bibli**, Florianópolis, v. 9, n. 18, 2004. Disponível em < <http://www.periodicos.ufsc.br/index.php/eb/article/view/155> >. Acesso em: 01 fev. 2009

DURANTI, L. **Diplomática**: usos nuevos para una antigua ciencia. Carmona: S&C Ediciones, 1996. (Colección Biblioteca Archivística, 5).

DUCHEIN, M. O respeito aos fundos em Arquivística: princípios teóricos e problemas práticos. **Arquivo & Administração**, Rio de Janeiro, v. 10-14, n. 1, abr. 1986.

FARIA, R. A. **Treinamento avançado em XML**. São Paulo: Digerati Books, 2005

FEITOSA, A. **Organização da informação na web**: das tags à web semântica. Brasília: Thesaurus, 2006.

FERNANDES, R. M. XML y registros electrónicos: principales estándares en la descripción archivística. **Ciência da Informação**, Brasília, v. 35, n.3, p.45-53, set/dez 2006.

FERNEDA, E. **Recuperação de informação: análise sobre a contribuição da ciência da computação para a ciência da informação** 2003. 388f. Tese (Doutorado em Ciências da Comunicação) - Programa de Pós-Graduação em Ciências da Comunicação, Universidade de São Paulo, São Paulo, 2003.

FERREIRA, M. **Introdução à preservação digital**: conceitos, estratégias e actuais consensos . Guimarães: Escola de Engenharia da Universidade do Minho, 2006. Disponível em: <<http://hdl.handle.net/1822/5820>> Acesso em: 01 fev. 2009.

FONSECA, M. O. K. **Arquivologia e Ciência da Informação**, Rio de Janeiro: FGV, 2005.

FREITAS, R. A. P.; RAMALHO, J. C. Relational databases digital preservation. In: INFORUM : SIMPÓSIO DE INFORMÁTICA, Lisboa, 2009 **atas...** [Lisboa : Universidade de Lisboa, 2009]. Disponível em <<http://repositorium.sdum.uminho.pt/handle/1822/9740>> Acesso em dez. 2010.

GARCEZ, A. N. R. Fontes Complementares na Pesquisa Historiográfica. IN CASIMIRO, A. P. B. S; LOMBARDI, J. C.; MAGALHÃES, D. R org. **A pesquisa e a preservação de arquivos e fontes para a educação, cultura e memória**. Campinas: Alínea, 2009.

GIL, A. C. **Como elaborar projetos de pesquisa**, 4. ed. São Paulo: Atlas, 2008

GREENBERG, J. Understanding Metadata and Metadata Schemes . In SMIRAGLIA, R. **MetaData**: A Cataloger's Primer has been co-published simultaneously as Cataloging & Classification Quarterly, °Mine 40, Numbers 3/4 2005.

HALPIN, H; THOMPSON, H. S. **One document to Bind Them**: Combining XML, Web Services, and the Semantic Web, Edinburgh: University of Edinburgh, 2006. Disponível em <

<http://www.ibiblio.org/hhalpin/homepage/publications/www2006/halpin-www2006.html> >. Acesso em: 01 fev 2009.

INNARELLI, H. C. Preservação Digital e seus dez mandamentos. In SANTOS, Vanderlei Batista dos (org). **Arquivística: temas contemporâneos...**, Distrito Federal: SENAC, 2007.

INTERPARES PROJECT. **The Long-term Preservation of Authentic Electronic Records:** findings of the InterPARES project, Disponível em <<http://www.interpares.org/book/index.cfm>> Acesso em 01 fev. 2009

_____. **Creator Guidelines:** making and maintaining digital materials: guidelines for individuals. Disponível em: <[http://www.interpares.org/ip2/display_file.cfm?doc=ip2\(pub\)creator_guidelines_booklet.pdf](http://www.interpares.org/ip2/display_file.cfm?doc=ip2(pub)creator_guidelines_booklet.pdf)> Acesso em: 01 fev. 2009.

_____. **Preserver Guidelines:** preserving digital records guidelines for organizations. Disponível em: <[http://www.interpares.org/ip2/display_file.cfm?doc=ip2\(pub\)preserver_guidelines_booklet.pdf](http://www.interpares.org/ip2/display_file.cfm?doc=ip2(pub)preserver_guidelines_booklet.pdf)> Acesso em: 01 fev. 2009.

LAKATOS, E. M.; MARCONI, M. A. **Fundamento de Metodologia Científica**, 3. ed. São Paulo: ATLAS, 1991.

LAMBE, P. **Organising knowledge:** taxonomies, knowlwdge and organizational effectiveness. Oxford: 2007.

LE COADIC, Y. **A Ciência da Informação**, 2. ed. Brasília: Briquet de Lemos, 2004.

LEUF, Bo. **The Semantic Web:** crafting infrastructure for agency. West Sussex: John Wiley, 2006.

LIBRARY OF CONGRESS. **Development of the Encoded Archival Description DTD**, 2002. Disponível em: < <http://www.loc.gov/ead/eaddev.html> > Acesso em 01 fev. 2009.

LOPEZ, A. P. A. **Tipologia documental de partidos e associações políticas brasileiras**. São Paulo: Loyola, 1999.

LUNARDELLI, R. S. A.; CALDERON, W. R. Fundo arquivístico: múltiplas leituras a respeito do termo. In: BERTALO, Linete; MORENO, Nádina Aparecida. **Gestão em arquivologia:** abordagens múltiplas. Londrina: EDUEL, 2008.

MENEZES, E. M.; CUNHA, V. C.; HEEMAN, V. M. **Glossário de Análise Documentária**. Londrina: ABECIN 2004.

MINISTÉRIO PÚBLICO DO ESTADO DE SÃO PAULO. **Tabela de Temporalidade de Documentos**. Disponível em : <http://www.mp.sp.gov.br/portal/page/portal/diretoria_geral/Documental/Tabela_Temporalidade> Acesso em 10 nov. 2010

NATIONAL ARCHIVES OF AUSTRALIA. **An Approach to the Preservation of Digital Records** Canberra: National Archives of Australia, 2002 Disponível em : <http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm2-888.pdf> Acesso em: 01 fev. 2009.

_____. **Developing a Functions Thesaurus:** Guidelines for Commonwealth Agencies, Canberra: National Archives of Australia, 2003 Disponível em: <<http://www.naa.gov.au/records-management/publications/Functions-thesaurus.aspx>> Acesso em: 01 fev. 2009.

_____. **Looking Back to the Future: 30 Years of Electronic Records in the National Archives of Australia.** 2004. Disponível em : <http://ourhistory.naa.gov.au/library/looking_back_to_future.html> Acesso em: 17 ago. 2010.

_____. **Digital Preservation:** Illuminating the past, guiding the future. Canberra: National Archives of Australia, 2006. Disponível em: <http://www.naa.gov.au/Images/XENA_brochure%5B1%5D_tcm2-918.pdf> Acesso em: 17 abr. 2007.

_____. **Digital Recordkeeping:** Guidelines for Creating, Managing and Preserving Digital Records. Canberra: National Archives of Australia, 2004 Disponível em : <http://www.naa.gov.au/Images/Digital-recordkeeping-guidelines_tcm2-920.pdf> Acesso em: 01 fev. 2009.

_____. **DIRKS:** A strategic Approach to Managing Business Information, Canberra: National Archives of Australia, 2001. Disponível em: <<http://www.naa.gov.au/records-management/publications/DIRKS-manual.aspx>> Acesso em: 01 fev. 2009.

_____. **How we preserve digital records of Australian Government agencies** [2009]. Disponível em: <<http://www.naa.gov.au/records-management/secure-and-store/e-preservation/at-NAA/index.aspx>> Acesso em: 01 fev. 2009.

_____. **Looking Back to the Future:** 30 Years of Electronic Records in the National Archives of Australia, Canberra: National Archives of Australia, 2004. Disponível em: <http://ourhistory.naa.gov.au/library/looking_back_to_future.html> Acesso em: 01 fev. 2009.

_____. **Overview of Classification Tools for Records Management:** 30 Years of Electronic Records in the National Archives of Australia, Canberra: National Archives of Australia, 2003. Disponível em: <<http://www.naa.gov.au/records-management/publications/classification-tools.aspx>> Acesso em: 01 fev. 2009.

OLIVEIRA, D. P. R. **Sistemas, Organização e Métodos:** uma abordagem gerencial. São Paulo:Atlas 2009.

PAES, M. L. **Arquivos:** Teoria e Prática, 3. ed. Rio de Janeiro: FGV, 1997

RAMALHO, J. C. et AL. XML e Preservação Digital. In XATA - XML: Aplicações e Tecnologias Associadas **Paper...**, Lisboa:FCUL, 2007. Disponível em: <<http://repositorium.sdum.uminho.pt/handle/1822/6224>> Acesso em: 01 fev. 2009.

RAMALHO et. al. Relational database preservation through XML modelling. In: INTERNATIONAL WORKSHOP ON MARKUP OF OVERLAPPING STRUCTURES (EXTREME MARKUP 2007), Montréal, Canadá, 2007 – “Extreme markup languages 2007 : **proceedings**”. [S.l. s.n., 2007]. Disponível em <<http://repositorium.sdum.uminho.pt/handle/1822/7120>>. Acesso em dez. 2010

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**, Rio de Janeiro, v. 8, n. 6, dez. 2007. Disponível em < http://www.dgz.org.br/dez07/Art_04.htm >. Acesso em: 01 fev. 2009

RODRIGUES, A. M. L. A teoria dos arquivos e a gestão de documentos. **Perspectivas em ciência da informação**, Belo Horizonte, v.11 n.1, p. 102-117, 2006.

RONDINELLI, R. C. **Gerenciamento arquivístico de documentos eletrônicos**: uma abordagem teórica da arquivística contemporânea. 2ed Rio de Janeiro:Editora FGV 2002.

ROUSSEAU, Jean-Yves; COUTURE, C. **Os Fundamentos da disciplina Arquivística** 1.ed. Lisboa: Dom Quixote, 1998.

SAMPIERI, H. et al. Metodologia de la investigación, México: Mc Gray Hill, 1994.

SANTOS, A. R. **Metodologia Científica**: A construção do conhecimento. Rio de Janeiro: De Paulo Editora, 1999

SANTOS, V. B. **Gestão de documentos eletrônicos**: uma visão arquivística, 2. ed. Brasília: ABARQ 2005

_____. A prática arquivística em tempos de gestão do conhecimento. In SANTOS, Vanderlei Batista dos (org). **Arquivística**: temas contemporâneos..., Distrito Federal: SENAC, 2007.

SARAMAGO M. L. Preservação digital a longo prazo: boas práticas e estratégias. **Cadernos BAD**, Lisboa, v.2, p 69-81, 2002.

SAYÃO L. F. **Preservação digital**: uma brevíssima introdução. Disponível em : <<http://libdigi.unicamp.br/document/?view=8398>>. Acesso em: 01 fev. 2009.

SCHULTZ, M. **Filosofía y producciones digitales**, Buenos Aires: Alfagrama, 2006.

SILVA, M. F.; LIMA, G. A. B. O. SEMANTICWEB 1.0: Uma abordagem para web semântica a partir de ontologias e tecnologias atuais. In SEMINÁRIO DE PESQUISA EM ONTOLOGIA NO BRASIL, 2008, Niterói. **Comunicação...** Niterói: UFF, 2008

SILVA, et al. **Arquivística**: Teoria e Prática de uma Ciência da Informação. 2 ed. Porto: Edições Afrontamento, 2002

SILVA, S. C. A. **A preservação da informação arquivística governamental nas políticas públicas do Brasil**. Rio de Janeiro: AAB:FAPERJ, 2008

SMIRAGLIA, R. P. Introducing Metadata. In SMIRAGLIA, R. **MetaData**: A Cataloger's Primer has been co-published simultaneously as Cataloging & Classification Quarterly, Mine 40, Numbers 3/4 2005.

SMIT, J. W; KOBASHI, N. Y. **Como elaborar vocabulário controlado para aplicação em arquivos**. São Paulo: Arquivo do Estado:Imprensa Oficial, 2003.

SOCIETY OF AMERICAN ARCHIVISTS. **Encoded Archival Description Tag Library**. Chicago:SAA, 2002. Disponível em < <http://archivists.org/publications/epubs/EAD2002-TL-05-03.pdf> > Acesso em dez. 2010

SOCIETY OF AMERICAN ARCHIVISTS; STAATSBIBLIOTHEK ZU BERLIN. Encoded Archival Context—Corporate Bodies, Persons, and Families (EAC-CPF), 2010. Disponível em < <http://www3.iath.virginia.edu/eac/cpf/tagLibrary/cpfTagLibrary.html#d1e8025> > acesso em dez. 2010

SOUZA, R. T. B. A classificação como função matricial do que-fazer arquivístico. In SANTOS, Vanderlei Batista dos (org). **Arquivística: temas contemporâneos...**, Distrito Federal: SENAC, 2007.

_____. **Classificação em Arquivística: trajetória e apropriação de um conceito**. 2004. São Paulo. Tese (Doutorado em História Social) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.

SOUZA, R. R. A Web Semântica e suas contribuições para a Ciência da Informação. **Ciência da Informação**, Brasília, v. 33, n. 1., p. 132-141, jan/abr 2004.

SHELLENBERG, T. R. **Arquivos Modernos: Princípios e Técnicas**, 6. ed. Rio de Janeiro: FGV, 2006.

STEWART, D. L. **Building Enterprise Taxonomies**. [s.l]:Morikota, 2008.

SUPERIOR TRIBUNAL DE JUSTIÇA. Vocabulário Jurídico. Disponível em <<http://www.stj.jus.br/SCON/thesaurus/>> acesso em jan. 2010.

SWISS FEDERAL ARCHIVES. **SIARD Format Description**. Berne:SFA, 2008. Disponível em <<http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html?lang=en&download=M3wBPgDB/8ull6Du36WenojQ1NTTjaXZnqWfVp3Uhmfnapmmc7Zi6rZnqCkklN0hHx+bKbXrZ6lhuDZz8mMps2gpKfo>> Acesso em jan. 2011

TEI CONSORTIUM. **TEI P5: Guidelines for Electronic Text Encoding and Interchange**. 2010. Disponível em < <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf> > Acesso em jan. 2011

TESTBED DIGITALE BEWARING. **XML and Digital Preservation**, 2002 Disponível em: <http://www.digitaleduurzaamheid.nl/bibliotheek/docs/white-paper_xml-en.pdf > Acesso em 01 fev. 2009.

THOMAZ, K. P. A preservação digital e o modelo de referência Open Archival Information System (OAIS). **Datagramazero**: Revista de Ciência da Informação, v.5, n1, fev/04 Disponível em: <http://www.dgz.org.br/fev04/F_I_art.htm>. Acesso em: 01 fev. 2009.

_____. **A preservação de documentos eletrônicos de caráter arquivístico: novos desafios, velhos problemas**. 2004. 388f. Tese (Doutorado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.

THURMAN, A. C. Metadata Standards for Archival Control: An Introduction to EAD and EAC. In SMIRAGLIA, R. **MetaData: A Cataloger's Primer** has been co-published simultaneously as Cataloging & Classification Quarterly, °Mine 40, Numbers 3/4 2005.

TORRES, A. G.; LORA, A. P.; LÓPEZ D. P. Reutilización de tesauros: el documentalista frente al reto de la web semántica. **El Profesional de la Información** . v. 17, n. 1, 2008. Disponível em < <http://elprofesionaldelainformacion.metapress.com/app/home/contribution.asp?referrer=parent&backto=searcharticlesresults,1,2;> > Acesso em dez. 2010.

VALENTIM, M. L. P. Análise de Conteúdo. In VALENTIM, Marta Lúcia Pomim (org). **Métodos qualitativos de pesquisa em Ciência da Informação**, São Paulo: Polis, 2005.

W3C. **HTML 4.0 Specification**, [s:l], 1998. Disponível em: <<http://www.w3.org/TR/1998/REC-html40-19980424/>> Acesso em: 01 fev. 2009

_____. **Enabling Standards and Technologies: Layer Cake**, 2002. Disponível em < <http://www.w3.org/2002/Talks/04-sweb/slide12-0.html> > Acesso em dez. 2010.

_____. **Extensible Markup Language (XML)**, [s:l], [2003?]. Disponível em: <<http://www.w3.org/XML/>> Acesso em: 01 fev. 2009

_____. **W3C Semantic Web Activity**, [s:l], [2001?]. Disponível em: <<http://www.w3.org/2001/sw/>> Acesso em: 01 fev. 2009

_____. **SKOS Simple Knowledge Organization System Reference**, 2009. Disponível em < <http://www.w3.org/TR/2009/REC-skos-reference-20090818/> > acesso em dez. 2010.

_____. **Semantic web**. [s:l], 2010. Disponível em: <<http://www.w3.org/2001/sw/>> Acesso em: 01 fev. 2010

_____. **SKOS Current Status** [s:l], 2010b. Disponível em < http://www.w3.org/standards/techs/skos#w3c_all > Acesso em dez. 2010.

WEINBERGER, D. **Everything is Miscellaneous: The Power of the New Digital Disorder**, New York: Times Book, 2007

WILSON, T. D. Human information behaviour. **Informing Science**, v.3, n.2, p.49-55, 2000. Disponível em: <<http://informationr.net/tdw/publ/papers/2000HIB.pdf>> Acesso em: 22 jun. 2009.

YOTT, P. Introduction to XML. In SMIRAGLIA, R. **MetaData: A Cataloger's Primer** has been co-published simultaneously as Cataloging & Classification Quarterly, °Mine 40, Numbers 3/4 2005.