

This paper studies the application of archival description in metadata for semantic links and interoperability of digital archival information, considered as information generated by institutions and funds files, with sections and documentary series. By applying archival approach for informational treatment in a technological network, it was objected to create dialogue between the areas of technology and information. The literature was reviewed including fundamental concepts of archival description and ontologies, also in the following the descriptive elements present in NOBRADE (standard based on ISAD (G) - International Standard Archival Description - General) were compared with Dublin Core metadata. This comparative study brings archival descriptive data structure that can be expressed in markup language belonging to the semantic web context. It can be noticed that the informational areas of NOBRADE have correspondents in Dublin Core scheme, so it was possible to conclude that these metadata can be used to represent archival descriptive data, showing that it was possible to create a kind of descriptive ontology to apply web 3.0, anchored in XML and RDF, which operate light ontologies easily applied in digital environments.

Estuda a aplicação da descrição arquivística em metadados de links semânticos e a interoperabilidade da informação arquivística digital, considerada como informação gerada por instituições em arquivos e fundos, com seções e séries documentais. Ao aplicar abordagem arquivística para tratamento informacional numa rede tecnológica, objetiva-se criar diálogos entre as áreas de tecnologia e de informação. Foi realizada revisão da literatura dos conceitos de descrição arquivística e ontologias, e a seguir foram comparados os elementos descritivos presentes na NOBRADE (norma baseada na ISAD(G) - *International Standard Archival Description – General*) com a estrutura do esquema de metadados *Dublin Core*. Este estudo mostra que a descrição arquivística pode ser expressa em linguagem de marcação pertencente ao contexto da *web* semântica. Pode-se perceber que as áreas informacionais da NOBRADE têm correspondentes no esquema *Dublin Core* concluindo-se que estes metadados podem ser utilizados para representar dados de descrição arquivística. Pode ser criado um tipo de ontologia descritiva para a *web* 3.0, ancorado em XML e em RDF, e de fácil aplicação em ambientes digitais.

The archival information on the semantic web: from description to ontology.

Charlley dos Santos Luz (Universidade de São Paulo, São Paulo, Brasil) – charlley@usp.br
Cibele Araújo Camargo Marques dos Santos (Universidade de São Paulo, São Paulo, Brasil) – cibileac@usp.br

Keywords: Archival Description, Semantic Web , Ontology, NOBRADE , Dublin Core .

Palavras-chave: Descrição arquivística, Web semântica, Ontologia, NOBRADE, Dublin Core.

Introdução

A *web* semântica, também chamada *web* 3.0, tem como base uma rede com informações ligadas através de significados. Por meio de recursos informacionais e tecnológicos as máquinas podem entender o conteúdo dos documentos e oferecer aos usuários respostas significantes e realizar inferências às questões e buscas realizadas, trazendo informação relevante e específica.

Estes recursos informacionais são resultantes do processo de organização semântica da informação, ou seja, são as informações rotuladas, estruturadas, combinadas e disponibilizadas em interfaces dinâmicas por meio de *links* abertos, utilizando o modelo conhecido como *Open Linked Data*. Tal modelo conecta dados relacionados que não estavam vinculados anteriormente entre si e reduz barreiras à ligação de dados estruturados outros métodos, como por exemplo, vocabulários controlados.

O processo de criação da rotulação das informações, ou seja, de dar sentido a determinado *link*, inclui a definição de padrões e o desenvolvimento de ontologias. Os padrões são recomendações de instituições como o W3C (*World Wide Web Consortium*) e as ontologias elaboradas por instituições de pesquisa ou conhecimento e até instituições privadas.

Este artigo apresenta a aplicação da NOBRADE (Norma Brasileira de Descrição Arquivística) como padrão para *links* semânticos, por meio de metadados, definindo uniformização para interoperabilidade da informação arquivística digital, aqui considerada como a informação gerada por instituições que descrevem e disponibilizam informações de seus arquivos e fundos, com suas respectivas seções e séries documentais.

Ao aplicar uma abordagem arquivística para solucionar problemas do tratamento informacional numa rede tecnológica, objetiva-se criar diálogos entre as áreas de tecnologia e da informação. Em relação à tecnologia aborda-se a *web* 3.0 e quanto à informação, esta é tratada tanto em sua estruturação como prevista pela Ciência da informação quanto ao compartilhamento da informação arquivística segundo parâmetros da Arquivologia. Assim, pretende-se incentivar a criação de políticas e uso de padrões de descrição e compartilhamento de informação em meio digital visando a *web* semântica e aprimorando a gestão e a distribuição do conhecimento presente na informação arquivística.

Desta forma foi realizada revisão da literatura dos conceitos fundamentais para as áreas de descrição arquivística e ontologias, e a seguir foram comparados os elementos descritivos presentes nas normas ISAD(G) (International Standard Archival Description – General) e NOBRADE com a estrutura do *Dublin Core*, esquema de metadados proposto como estrutura básica de organização de dados disponíveis na *web*. Com este estudo comparativo apresenta-se a aplicação do produto da descrição arquivística em estrutura de dados que pode ser expressa em linguagem de marcação pertencente ao contexto da *web* semântica.

Além disso, apresenta a ontologia como linguagem para informação digital (apesar desta não ser uma abordagem uníssona na área de Ciência da Informação), mas aderente à ideia de que "as ontologias, assim como as linguagens documentárias, na condição de instrumentos de representação do conhecimento, permitem descrever alguns fenômenos e atuam, desse modo, como linguagem" (Moreia & Santos Neto, 2014, p. 51) que pode ser aplicada à arquivos. Não é objetivo descrever neste artigo todos os componentes da *web* semântica, mas indicar alguns elementos que podem ser utilizados para a interoperabilidade de dados e interoperabilidade semântica da descrição arquivística em ambientes digitais.

Metodologia

O presente artigo é de natureza exploratória e comparativa. A exploração temática e conceitual se dá na revisão bibliográfica embasando a comparação dos elementos de descrição arquivística e os elementos utilizados pela *web* 3.0. Assim, exercita-se um diálogo com a literatura selecionada das áreas de Tecnologia da Informação, Ciência da Informação e Arquivologia, abordando elementos técnicos, informacionais, linguísticos e arquivísticos. Ao compreender a função da descrição de arquivos no contexto da *web* 3.0 prospecta-se uma abordagem arquivística para internet.

A técnica de análise traz na primeira parte do trabalho os conceitos fundamentais dos campos científicos e já procura criar relações com a Descrição Arquivística (uma das chamadas funções arquivísticas¹) e a *Web* 3.0. A partir das reflexões da revisão de conceitos na segunda parte desta pesquisa analisam-se dois padrões de descrição de informações, o padrão *ISAD(G)* e seu correspondente brasileiro NOBRADE para arquivos e o *Dublin Core* para informação digital.

Neste trabalho aborda-se a informação arquivística como aquela que resulta da descrição multinível dos arquivos e das instituições que os organiza com a abordagem analítico-funcional. Isto é uma característica dos arquivos contemporâneos e principalmente dos arquivos digitais, sendo que

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

Desta forma, a abordagem citada por Thomassem de "metodologia arquivística atual", que parte da avaliação do contexto orgânico e serial para definir seus padrões guia este trabalho e o que esta identificado como informação arquivística nesta pesquisa. Além disso, o levantamento de dados partiu da conceitualização, passou pela comparação de padrões e procurou identificar na rede mundial *web* pelo menos um projeto que já utilize algum padrão descritivo arquivístico e uma ontologia.

Web Semântica e Ontologias

A *web* semântica, como evolução da *web* sintática já é uma realidade colocada por meio de sua evolução estrutural, sendo um campo importante de pesquisa e posicionada numa linha de tempo

... a partir do final da década de 1990, começaram a formalizar-se pesquisas relacionadas ao desenvolvimento de uma nova geração da *Web*, com o objetivo de possibilitar a incorporação de ligações semânticas aos recursos informacionais, de modo que os próprios computadores possam "compreendê-las" de forma

¹ Segundo Rousseau e Couture são 7 as funções arquivísticas: Produção, Avaliação, Aquisição, Conservação, Classificação, Descrição e Difusão.

automatizada. Machine understandable information, com esta sucinta expressão Berners-Lee (1998) impulsionou os primeiros estudos em direção ao projeto da *Web Semântica* (Ramalho, Vidotti e Fujita, 2007, p.1)

A questão da *web* semântica cristaliza-se de maneira importante, fundamentalmente quando Robredo define o surgimento da *web* semântica citando seu criador Tim Berners-Lee et al (2001)

[...] a Web Semântica não é uma Web separada, mas uma extensão da atual, na qual é dada à informação um significado bem definido, o que permite às pessoas e aos computadores trabalharem em cooperação. Os primeiros passos para entrelaçar a Web Semântica na estrutura da Web existente, já estão sendo dados. Num futuro próximo, esses desenvolvimentos deverão conduzir a significativas novas funcionalidades à medida que as máquinas irão se tornando bem mais capazes de processar e —entender os dados que, por enquanto, se limitam somente a apresentar. A propriedade essencial da World Wide Web é sua universalidade. A força do hipertexto é que □ alguma coisa pode ligar-se a outra coisa (Robredo, 2010, p 25).

Aqui, vemos que a semântica, o sentido dado à informação, é o que diferencia a *web* sintática da *web* 3.0 e que esta é uma melhoria e não uma ruptura, o que concede seu potencial de interoperabilidade, visível em estruturas de rede “Como a Internet, a *Web Semântica* será tão descentralizada quanto possível. [...] A descentralização implica compromissos: a *Web* deve atingir uma consistência total em suas interconexões [...]” (Robredo, 2010, p 25). O desafio de estruturar a informação para compartilhar é que direciona os projetos de *web* semântica.

Nesta pesquisa entende-se que a produção de informações disponíveis na rede representa um desafio em estruturar e apresentar resultados, revendo técnicas e gerando conhecimento, sendo assim necessário planejar a disponibilização da informação através de ambientes digitais, afinal a produção exponencial de informação na *web* vem

conduzindo a um estágio em que os modelos clássicos de representação e recuperação de informações precisam ser (re)pensados sob diferentes perspectivas, pois considerando a representação como elemento fundamental, para a garantia de qualidade na recuperação” (Ramalho, Vidotti & Fujita, 2007, p.1)

Cabe destacar que, no âmbito da *web* 3.0 e considerando a organização da informação, parte de taxonomias e modelos descritivos alimentam os *cérebros artificiais* capazes de entender as estruturas e ontologias, afinal

dos bilhões de documentos que formam a *World Wide Web* e dos *links* que os mantêm juntos, os cientistas do computador e número crescente de empreendedores estão encontrando novos caminhos para garimpar a inteligência humana. A finalidade é acrescentar uma camada de significado no topo da *Web* atual que faça dela menos um catálogo e mais um guia – e ainda forneça os fundamentos para que os sistemas possam raciocinar de forma semelhante aos humanos (Robredo, 2010, p. 19).

Assim, fica claro que a forma de estruturar a *web* semântica será definida pelas taxonomias complementadas por ontologias (estruturas para representação de conteúdo semântico que são hierarquizadas como as linguagens documentárias ou vocabulários controlados), em conjunto com

os formatos de representação descritiva de informação, como são as técnicas de descrição arquivísticas realizadas e disponibilizadas em ambientes digitais.

As linguagens documentárias, representadas no âmbito da *web* como as ontologias no formato OWL e vocabulários no formato SKOS são elementos importantes que operam com outros elementos na *web* 3.0, fazendo parte de uma de suas camadas, prevista já na formulação do conceito da *web* semântica que foi concebido como um conjunto de tecnologias relacionadas, de modo que no ano de 2000 o *World Wide Web Consortium* (W3C)

divulgou publicamente a primeira proposta de arquitetura da *web* semântica, com base em uma série de camadas sobrepostas, onde cada camada ou tecnologia deveria obrigatoriamente ser complementar e compatível com as camadas ao mesmo tempo em que não deveria depender das camadas superiores, possibilitando assim uma estrutura idealmente escalonável, que indicasse os passos e as tecnologias necessários para a concretização do projeto *Web Semântica*. (Ramalho, Vidotti & Fujita, 2007, p. 2)

Assim, é necessário que se criem instrumentos que forneçam o sentido lógico- semântico aos computadores, sendo o uso de linguagens documentárias uma dessas camadas inseridas como atributos nas linguagens de marcação, e aplicando valores por meio de *links*, afinal como diz Ramalho et al. (2007, p.2) "estes instrumentos são os *links* semânticos" representando a estrutura de linguagens disponível para a *web* semântica, que possibilita a criação de sentido aos *links* e consequentemente proporcionando uma melhor experiência na organização, recuperação da informação, e acesso do usuário, que conforme o W3C, é composta, entre outras, pelas estruturas indicadas no Quadro 1.

Quadro 1: Estrutura de recursos de linguagens da *Web Semântica*.

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

Fonte: W3C Brasil <http://www.w3c.br/Padroes/WebSemantica>

Na comparação entre a organização da informação da *web* sintática com a *web* 3.0, destaca-se a importância da ontologia, pois para que a camada ontológica OWL possa repassar conhecimento é necessário contextualizar a informação dentro de um cenário concreto. Nesta

camada é onde se define uma ontologia que se trata de conceituações de determinado domínio linguístico, de área do conhecimento, de um conjunto de atividades funcionais, ou outras, pois

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

Tanto uma camada ontológica, como uma camada com um sistema de organização do conhecimento, precisam inserir seus atributos em uma linguagem de marcação de proposição geral, que apresenta as especificações sintáticas e descritivas. Assim, vamos analisar a partir destes princípios os elementos descritivos propostos pela NOBRADE como forma de identificar *links*, fornecendo informação arquivística para o contexto 3.0, que neste sentido

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

Em analogia, podemos afirmar que as URIs da *web* 3.0 possuem elementos descritivos, como fichas de metadados, que descrevem e dão sentido a um elemento do objeto informacional. O conjunto a que se refere Berners-Lee pode ser considerado uma ficha descritiva, com as relações representativas criadas pelos elementos, aplicados aos *links*.

Metadados e Interoperabilidade

Os arquivos, assim como bibliotecas e museus, sempre produziram metadados, por meio de seus catálogos, dossiês, índices entre outros instrumentos descritivos. Agora eles estão digitalizados com potencial de tornarem-se interoperáveis (através do movimento de dados abertos – LOD: *Linked Open Data*), lincáveis e carregando sentido através de seu esquema de metadados. São recentes as iniciativas que unificam estes metadados, como a EUROPEANA que opera este modelo para bibliotecas públicas na Europa, ou do projeto ReLoad (Repository for Linked Open Archival Data) que

"è un progetto realizzato dall'Archivio Centrale dello Stato, dall'Istituto per i Beni culturali dell'Emilia Romagna e da Regesta.exe con l'obiettivo di sperimentare le metodologie del semantic web e le tecnologie standard per i linked open data (LOD) per favorire la condivisione di informazioni archivistiche provenienti da una molteplicità di fonte." (ReLoad, 2014)

A base da interoperabilidade é a informação adicionada sobre o *link*, ou a meta-informação (metadados) do objeto informacional, criando assim formas de descrever o objeto. Como afirma Patricio (2012) "o desafio da *Web* Semântica é propiciar uma linguagem que expresse dados e regras lógicas para raciocinar sobre esses dados, de forma interoperável na *web*. Adicionar lógica à

Web significa utilizar regras para fazer inferências, escolher ações e responder a questões. As tecnologias da *Web Semântica* são o XML e o RDF. O XML permite que qualquer pessoa adicione etiquetas para anotar as suas páginas *web* ou secções de texto nessas páginas”. Logo, metadados são representações de informação digital, associados a um recurso *online*, possibilitando seu tratamento técnico.

Com base em Pires (2012), afirma-se que os metadados podem "ser descritos separadamente em um sistema de armazenamento de informações (bancos de dados), como podem estar embebidos nos arquivos digitais”, portanto a padronização e a normatização descritiva garantem dados semelhantes para troca e interoperabilização.

Este conjunto é fundamental para a garantia da interoperabilidade, que ocorre por máquina no processo de navegação, a padronização do formato para equivalência sintática a outros formatos, e a normatização do conteúdo descrito para a equivalência semântica.

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

Logo, são dois os recursos de estruturação da *web* semântica, o modelo XML e um esquema de metadados para representação descritiva, como o Dublin Core, pois um importante passo para essa estruturação foi o desenvolvimento da XML (Extensible markup language), que enfatiza não mais a aparência, porém o conteúdo das páginas *web*. Uma segunda e igualmente fundamental contribuição vem da Dublin Core Metadata Initiative (Trzesniak & Koller, 2005, p. 188). Estes dois recursos tecnológicos são básicos para a interoperabilização de dados.

A isso se soma o modelo de expansão, ditado pelo movimento *Open Archives*, arquivos abertos, que possibilita o intercâmbio, coleta e distribuição de conteúdo presente em metadados estruturados. Continuam os autores esclarecendo que

A terceira componente é devida ao movimento dos arquivos abertos (open archives initiative, OAI: www.openarchives.org): consiste no protocolo para coleta de metadados (protocol for metadata harvesting, PMH), através do qual toda a informação na *web*, que esteja adequadamente descrita por metadados, pode ser coletada e apropriadamente recuperada de repositórios construídos por provedores de dados (Trzesniak & Koller, 2005, p. 188).

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

Nesse sentido, a interoperabilidade semântica, segundo Sayão e Marcondes (2008), “está relacionada com o significado ou semântica das informações originadas de diferentes recursos e é solucionada pela adoção de ferramentas comuns ou/e mapeáveis de representação da informação, como esquemas de metadados, classificações, tesouros e mais recentemente, ontologias”.

Para facilitar a interoperabilidade semântica é necessária a estruturação dos atributos de representação da informação nas camadas que neste primeiro será analisada através da comparação dos elementos descritivos da NOBRADE, que é um desdobramento da ISAD-G, com o Dublin Core.

O Dublin Core e a Descrição Arquivística

O Dublin Core é um modelo de metadados que possui um esquema que pode ser usado por todos aqueles que se propõem a descrever informação, sem a necessidade de seja um profissional da área de informação ou tecnologia. Foi criado para ser capaz de padronizar elementos comuns de descrição dos objetos informacionais, afinal

as publicações na *Web* crescem de forma exponencial e não existem profissionais de informações suficientes para descrevê-los todos, permitir que autores eles mesmos descrevam seus documentos ao publicá-los eletronicamente “na fonte”, conforme uma proposta bem antiga da Biblioteconomia, retomada por Weibel (1995). Para isso, é necessário, portanto, que descrições/representações de recursos informacionais – os agora chamados metadados - possam seguir um padrão de descrição que seja simples o suficiente para que um autor possa descrever seu documento. Esta é a proposta da iniciativa Dublin Core. (Marcondes 2005, p.5)

Então podemos entender que um esquema de metadados registra a ficha do objeto informacional (direitos de propriedade, autorais, responsabilidade, entre outros) e que serve de registro da vida pregressa e atual do objeto, são dados como o tamanho em dimensões e bytes dos documentos, o seu formato para gerenciá-lo e utilizá-lo, fornecendo assim sua qualificação e informações orgânicas. O Dublin Core fornece as informações consideradas orgânicas e, portanto, potencialmente arquivísticas, além das informações de contexto de criação, autoridade, sendo muito importante, pois vários tipos de metadados devem ser acompanhados de meta-informação sobre seu criador, sua data de criação, bem como o método de criação, pois

Linked Data deve ser publicado juntamente com vários tipos de metadados, a fim de aumentar a sua utilidade para os consumidores de dados. Permitindo que os clientes avaliem a qualidade dos dados publicados e determinem se eles querem confiar nesses dados, os dados devem ser acompanhados com meta-informação sobre seu criador, sua data de criação, bem como o método de criação (Hartig, 2009). A proveniência básica da meta-informação pode ser fornecida usando termos Dublin Core (Heath, Hepp & Bizer, 2008, p.8).

Aborda-se, portanto, o uso da NOBRADE como padrão descritivo procurando relacioná-lo ao Dublin Core, validando a possibilidade de haver informação arquivística estruturada para distribuição na *web* 3.0 utilizando os padrões tanto da Arquivologia quanto da *web*.

A descrição arquivística, aplicada através de modelos como a ISAD(G) e a NOBRADE é o processo em que o arquivista cria representações de um determinado acervo arquivístico, explicitando o contexto e conteúdo deste acervo num instrumento de pesquisa ou de referência, de forma que os

"instrumentos arquivísticos de referência são os produtos do processo de descrição arquivística, que se ocupam de criar representações para o acervo ou parcelas deste". Ou seja, a estruturação da informação presente nos arquivos representada digitalmente, conhecida como informação arquivística, é foco deste processo, desta função arquivística (Andrade, 2009, p. 6).

A padronização da descrição arquivística é um fato recente para a área, contemporânea ao

conceito de *Web 3.0* e focada no processo de criação de instrumentos descritivos, que servem, para dar acesso e permitir a pesquisa nos acervos. Este conjunto de regras gerais para a descrição arquivística faz parte de um processo que visa:

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

Há quatro regras fundamentais que devem ser aplicadas para estabelecer a relação hierárquica entre as descrições segundo o Conselho Internacional de Arquivos que são: a descrição do geral ao particular, apresentando uma relação hierárquica entre as partes e o todo; a informação relevante para o nível de descrição, apresentando informações que devem ser apropriadas para o nível que está sendo descrito; relação entre descrições que identifica o nível de descrição; e a não repetição de informação para não repetir as informações em níveis diferentes de descrição.

A NOBRADE, utilizada somente no Brasil, é resultante da tropicalização da norma mundial ISAD(G). Ao entender a estrutura de uma, entendese a estrutura de ambas, que possuem abordagem multinível, ou seja, “a estrutura hierárquica de fundos, séries e dossiês, avalizada pela norma ISAD(G), está baseada no princípio da proveniência, caro à gestão arquivística; acervos arquivísticos são mantidos obedecendo a esse princípio” (Marcondes, 2012, p. 183).

No entanto, deve-se considerar todo processo longo e revisado por pares de formação das primeiras regras de descrição de arquivos, inéditas até o fim da década de 90 do século passado

Normas de descrição arquivística são baseadas em princípios teóricos aceitos. Por exemplo, o princípio de que a descrição arquivística procede do geral para o particular é uma consequência prática do princípio de respeito aos fundos. Este princípio deve ser claramente enunciado caso se deseje construir uma estrutura de aplicação geral e um sistema de descrição arquivística, manual ou automático, não dependente de instrumentos de pesquisa (Conselho Internacional de Arquivos, 2000, p. 2).

A falta de padrão isola o conteúdo e a informação gerada nos serviços de informação como museus e arquivos, pois os mesmos muitas vezes possuem uma abordagem singular, cada serviço destes possui um contexto específico, portanto, em se tratando de dados abertos interligados (*Linked Open Data*), devemos considerar que

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

Por isso a utilização de regras para padronização como a ISAD(G) pode colaborar no processo de criação de uma *web* semântica com dados arquivísticos, pois disponibiliza de forma padronizada os dados descritivos dos acervos e dos geradores de acervos, mantendo a

possibilidade de descrição multinível, o que só pode enriquecer o acesso à informação.

A utilização de um padrão de descrição como ISAD(G) aumenta a visibilidade da informação arquivística, pois pode ser empregado como um padrão de definição de elementos a serem processados pela *web* semântica.

A ISAD (G) possui os seguintes elementos, organizados por áreas que agrupam os elementos descritivos por proximidade de uso. Aqui não será detalhado cada elemento, para isso recomenda-se a leitura da norma, serão indicados para contextualização. A norma possui sete áreas como apresentado no Quadro 2.

Quadro 2 – ISAD (G) – Conarq

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

Fonte: http://www.conarq.arquivonacional.gov.br/Media/publicacoes/isad_g_2001.pdf

A NOBRADE possui, além dos elementos da ISAD (G) mais uma Área (área 8 – Pontos de Acesso) e mais dois Elementos de Descrição (os elementos 6.2 – Notas sobre conservação e 8.1 – Pontos de acesso e indexação de assuntos).

Ao aplicar a NOBRADE, além de aplicar um modelo internacional já adotado por outras iniciativas de Open Archives (pois a NOBRADE contém a ISAD-G), estamos também ampliando o potencial de interoperabilidade de dados, pois o oitavo elemento da NOBRADE abre a possibilidade do uso de vocabulários e ontologias temáticas, enriquecendo a indexação e melhorando a relevância na recuperação da informação.

O uso de padrões aplicados para descrição de arquivos traz uma mudança de atuação dos profissionais da área, e a necessidade de se adotar mudanças conceituais que aprofundam o controle da estruturação das informações (descritivas e temáticas com as ontologias), a economia de recursos (uso de padrões), intercâmbio de dados (XML e RDF) e a melhoria da qualidade metodológica com ISAD(G) e NOBRADE, além da definição de estratégias, políticas institucionais e de relacionamentos mais amplos, visando o compartilhamento das informações com instituições nacionais e internacionais. Há como manter-se a custódia (posse da informação), partilhando os metadados e fornecendo o acesso aos objetos informacionais digitais.

Resultados

Devido às complexidades impostas pelo mundo digital, e principalmente aos profissionais da área de informação e de arquivos, a *web* semântica exigirá a um grande esforço para a estruturação de dados visando evitar a redundância e garantir a interoperabilidade, de modo que deixa de ser foco o compartilhamento de documentos para focar o compartilhamento de dados estruturados, como proposto pela *web* 3.0.

Os metadados do Dublin Core (DCMI) foram analisados e comparados com as áreas de informação da NOBRADE. A partir disso através do método dedutivo os dados foram distribuídos os elementos DCMI nas áreas da NOBRADE.

O padrão DCMI é composto de 15 elementos descritivos, simples e autoexplicativos portanto, dada a facilidade de aplicação deste padrão e a recomendação do consórcio W3C, foi realizada a comparação com o padrão da NOBRADE para sugerir usos e identificar se existe a necessidade de criar outros elementos que comportam os dados das URI.

O formato Dublin Core é composto por 15 elementos, indicados no Quadro 3 sendo que o detalhamento pode ser acessado no *site* do DCMI.

Quadro 3 – Elementos do Dublin Core

Elementos
Title (Título)
Creator (Criador)
Subject (Assunto)
Description (Descrição)
Publisher (Editor)
Contributor (Contribuidor)
Date (Data)
Type (Tipo)
Format (Formato)
Identifier (Identificador)
Source (Origem)
Language (Idioma)
Relation (Relação)
Coverage (Abrangência)
Rights (Direitos)

Fonte: http://www.niso.org/apps/group_public/download.php/10256/Z39-85-2012_dublin_core.pdf

Alguns elementos podem ter qualificadores que têm a função de refinar ou tornar mais específico o recurso descrito, portanto, é decisão do profissional da informação sobre qual formato

utilizar (simples ou qualificado). Com a prerrogativa de poder utilizar qualificadores para o Dublin Core é possível criar elementos específicos para arquivos, o que pode atender, por exemplo, a área de notas não coberta pelos 15 elementos básicos.

Assim, foram relacionados os elementos NOBRADE detalhados anteriormente com os elementos simples do Dublin Core utilizando-se de um elemento qualificado apresentado no Quadro 4.

Quadro 4: Comparação dos elementos NOBRADE e Dublin Core

Agrupamento de Elementos NOBRADE	Elementos Dublin Core
Área de identificação	Title (Título); Creator (Criador); Date (Data); Identifier (Identificador); Source (Origem)
Área de contextualização	Coverage (Abrangência); Rights (Direitos)
Área de conteúdo e estrutura	Description (Descrição); Contributor (Contribuidor); Type (Tipo); Format (Formato)
Área de condições de acesso e de uso	Rights (Direitos); Language (Idioma); Coverage (Abrangência)
Área de fontes relacionadas	Relation (Relação);
Área de notas	Text (Texto) - elemento qualificado;
Área de controle da descrição	Publisher (Editor);
Área de Pontos de Acesso e Indexação de Assuntos	Subject (Assunto);

O elemento qualificador que pode ser utilizado para notas é o elemento TEXT – Texto, que segundo o DCMI são textos para serem lidos e incluem livros, cartas, dissertações, poemas, jornais, artigos, arquivos de listas de discussão, além de imagens de textos e cópias, segundo sua definição: “A resource consisting primarily of words for reading. Comment: Examples include books, letters, dissertations, poems, newspapers, articles, archives of mailing lists. Note that facsimiles or images of texts are still of the genre Text” (Dublin Core Metadata Applications Initiative 2014).

Pode-se perceber que as áreas informacionais da NOBRADE têm correspondentes no esquema Dublin Core, seja de identificação, contexto, conteúdo, estrutura, acesso, fontes relacionadas, e controle. Logo é possível concluir que os metadados Dublin Core podem ser utilizados para representar dados de descrição arquivística.

A comparação da NOBRADE com o DUBLIN CORE mostra que pode ser criado um tipo de ontologia descritiva para aplicar na *web* 3.0, desde que ancorada numa base XML e que pode, portanto, ser empregada também em RDF, que opera em ontologias leves, sendo de fácil aplicação em códigos de ambientes digitais.

Por fim, foi realizada uma busca exploratória na *web* utilizando-se as palavras Ontologia e Descrição Arquivística, o que possibilitou a identificação do projeto italiano ReLOAD (ReLOAD, 2014) que mostra que a proposta deste trabalho já é uma realidade. O projeto disponibiliza em seu endereço eletrônico uma Ontologia ISAD (G). O objetivo do projeto é testar os métodos de *web* e padrão de tecnologias semânticas para dados abertos ligados para facilitar o compartilhamento de informações de arquivo a partir de uma variedade de fontes.

Conclusões

Partiu-se da necessidade de conhecer o conceito de *web* semântica relativo à linguagem e sua estrutura para compreender o papel dos serviços de informação e da informação digital gerada

nos arquivos neste ambiente da *web* 3.0. A partir daí, foi observado que a estrutura organizacional mantém a linguagem documentária como elemento chave e que esta pode ser uma ontologia.

Observou-se que a informação arquivística gerada nos processos de descrição utilizando-se da ISAD (G) e NOBRADE pode ser uma forma de estruturar ontologias descritivas na *web*, podendo-se utilizar do padrão Dublin Core na codificação. A partir daí, foi possível aferir que a formação de padrões descritivos para arquivos através do uso de normas visa facilitar o compartilhamento da informação arquivística, gerando significado aos *links* criados nos acervos digitais. Cabe ainda destacar o papel da informação arquivística e sua organização nesta pesquisa.

Para a compreensão da informação arquivística, esta foi definida como aquela originada em sistemas arquivísticos que tem como principal característica ser a informação contida no conjunto de documentos arquivísticos, ou seja, a que é considerada como informação orgânica.

Por tratar-se de um conjunto descritivo multinível, como previsto na ISAD(G), deve-se considerar a informação arquivística como resultado da estruturação de um conjunto documental de mesma origem orgânica, ou um fundo, sendo a

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

Ou seja, a informação orgânica de séries, subséries, seções e subseções, além da própria informação de fundo e de autoridade representam o que pode ser determinado como informação arquivística, pois

Desta maneira, em nossa concepção, há duas "possibilidades informacionais" no campo dos arquivos. Uma, diz respeito ao que está "dentro" do documento, o seu conteúdo. A outra é aquela que está "fora" do documento, isto é, quem o produziu, por que, para quê, quando, onde e como, além de seu trâmite e seus vínculos. Ou seja, o "material de arquivo" tal como argumentado por Bellotto (2002a, p. 11). Todas essas informações que estão "fora" do documento são, por nós consideradas, portadoras dos elementos que caracterizam este "material de arquivo" e tributam à Arquivologia a sua especificidade frente a outras áreas do saber. (Schmidt, 2013).

Assim, pode-se considerar a informação orgânica, ou o material de arquivo, (produto do funcionamento de sistemas informacionais de arquivos, que apoiam a realização de atividades), como a informação arquivística. Claramente utiliza-se informação arquivística e informação orgânica como sendo sinônimos, pois, o foco é a informação descritiva e extraída através da aplicação de normas, como a ISAD(G) e a NOBRADE.

O uso de padrões descritivos como a ISAD(G) cria uma normalização de elementos informacionais comuns a todos os arquivos (metadados), extrai-se deles os contextos orgânicos variados, gerados numa realidade única para cada arquivo, mas também se cria um padrão de interoperabilidade e a possibilidade de aplicar a visão da *web* 3.0 nos arquivos, isto é, que utiliza a informação arquivística, a partir dos instrumentos descritivos, como mais uma ontologia de descrição arquivística.

Referências

- Andrade, R. S., & Silva, R. R. (2009). Uma nova geração de instrumentos arquivísticos de referência: a publicação dos produtos das descrições arquivísticas em meio eletrônico. *Simpósio Baiano de Arquivologia*, 2.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*.
- Borges, M. A., & Morales, A. B. (2013). Representação ontológica da memória organizacional da midiatisação da interação educacional. *Colóquio Luso-Brasileiro de Educação a Distância e E-learning*, (pp. 1-18). Lisboa.
- Conselho Internacional de Arquivos. (2000). *ISAD(G): norma geral internacional de descrição arquivística* (2. ed ed.). Rio de Janeiro.
- Conselho Nacional de Arquivos. (2006). *Nobrade: Norma Brasileira de Descrição Arquivística*. Rio de Janeiro: Arquivo Nacional.
- Dublin Core Metadata Applications Initiative. (2014). *DCMI*. Fonte: <http://dublincore.org/specifications/>
- Heath, T. H. (2008). Special Issue on Linked Data. *International Journal on Semantic Web and Information Systems International*.
- Lumsden, J., Hall, H., & Cruickshank, P. (2011). Ontology definition and construction, and epistemological adequacy for systems interoperability: a practitioner analysis. *Journal of Information Science*, 37, p. 246.
- Marcondes, C. H. (2005). Metadados: descrição e recuperação de informações na web. In: C. H. Marcondes, *Bibliotecas digitais: saberes e práticas*. , BA : , 2005. p. (pp. 77-143). Salvador; Brasília: EDUFBA; IBICT.
- Marcondes, C. H. (2012). “Linked Data” dados interligados e interoperabilidade entre arquivos, bibliotecas e museus na web. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 17(34), pp. 171-192.
- Moreira, W., & Santos Neto, M. F. (2014). A formação do conceito de ontologia na ciência da informação: uma análise nos periódicos Scire e Ibersid. *Scire*, 20(2), pp. 49-54.
- Moreiro González, J. A. (2011). Evolução ontológica das linguagens documentárias. Relato de uma experiência de curso organizado conjuntamente para o DT/SIBI-USP e o PPGCI/ECA. *InCID: Revista de Ciência da Informação e Documentação*, 2(1), pp. 143-164.
- Patricio, H. (2012). A Europeana e a agregação de metadados na web: análise dos esquemas ESE/EDM e da aplicação de standards da web semântica a dados de bibliotecas. *Actas*, 11.
- Pires, D. (2012). Uso do Dublin Core na descrição de obras raras na web: a coleção da biblioteca brasileira digital. *I Encontro Nacional de Catalogação* (pp. 1-13). Fundação Biblioteca Nacional .
- Ramalho, R. A., & Vidotti, S. A. (2007). Web semântica: uma investigação sob o olhar da Ciência da Informação. *DataGramaZero*, 8(6).
- Robredo, J. (2010). Ciência da informação e Web semântica: linhas convergentes ou linhas paralelas? In: J. Robredo, & M. B. (orgs), *Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento* (pp. 12-47). Brasília: IBICT.

- Santos, V. B. (2011). *A teoria arquivística a partir de 1898: em busca da consolidação, da reafirmação e da atualização de seus fundamentos*.
- Sayao, L. F., & Marcondes, C. H. (2008). O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. *TransInformação*, 20(2), pp. 133-148.
- Schmidt, C. M. (2013). Entre o documento de arquivo e a informação arquivística: reflexões acerca do objeto científico da Arquivologia. *XIV Encontro Nacional de Pesquisa em Ciência da Informação - ENANCIB*.
- The Open Archives Initiative Protocol for Metadata Harvesting. (2014). *Protocol Version 2.0*.
- Thomassem, T. (2006). Uma primeira introdução à Arquivologia. *Arquivo & Administração*, 5(1), pp. 5-16.
- Trzesniak, P., & Koller, S. (2005). Tornar-se semântico: adequando nossa revista ao tempo de informação por toda parte. *Revista Interamericana de Psicología*, 39(2), pp. 185-188.