

Web Semântica e preservação digital: o padrão de metadados PREMIS na proposta do *Linked Data*

Felipe Augusto Arakaki

Universidade Estadual Paulista “Júlio Mesquita Filho” – UNESP / Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Campus Itapetininga, Email: felipe.arakaki@unesp.br

Paula Regina Ventura Amorim Gonçalves

Universidade Federal do Espírito Santo - UFES, Email: paula.goncalvez@ufes.br

Caio Saraiva Coneglian

Universidade Estadual Paulista “Júlio Mesquita Filho” – UNESP, Email: caioconeglian@gmail.com

Plácida Leopoldina Ventura Amorim da Costa Santos

Universidade Estadual Paulista “Júlio Mesquita Filho” – UNESP, Email: placidasantos@gmail.com

José Eduardo Santarém Segundo

Universidade de São Paulo - USP, Email: santarem@usp.br

Resumo: A partir do significativo aumento da produção de conteúdos digitais, ações visando a recuperação e a preservação de objetos digitais a longo prazo tornam-se de importância capital em nossos dias, caso contrário, haverá um rompimento no conhecimento devido a perda ou insuficiência na representação das informações disponíveis no ambiente Web. Nesse contexto, a Curadoria Digital se configura como uma área interdisciplinar que gerencia informações com a proposta e o desafio de preservação e transmissão de objetos informacionais, atuando sobre os fatores de risco como: tecnologia obsoleta, fragilidade tecnológica, falta de entendimento sobre o que é uma boa prática digital e recursos inadequados. Nesse cenário, foi feito um recorte sobre conceitos e funções dos metadados para preservação digital. Assim objetiva-se verificar como o padrão de metadados *Preservation Metadata: Implementation Strategies* (PREMIS) pode ser utilizado para atender as demandas de preservação dos objetos informacionais digitais. Optou-se como procedimento metodológico a revisão da literatura visto que, é uma reflexão sobre a representação da informação tendo como pano de fundo a Web semântica. Como resultado abordou-se a estrutura e as características do padrão metadados PREMIS e sua relação com a Web Semântica.

Palavras-chave: Metadados de Preservação. Preservation Metadata: Implementation Strategies. PREMIS. Web Semântica.

INTRODUÇÃO

Com o desenvolvimento e avanço das tecnologias, o número de recursos informacionais é cada vez maior. Nesse contexto, bibliotecas, repositórios e arquivos digitais necessitam que seja feita a representação e a organização de tais recursos de maneira adequada, possibilitando sua posterior recuperação, localização e acesso.

Assim, há uma urgência de estudos para a criação de processos para a descrição e consequentemente e preservar os recursos informacionais, por um longo prazo. Nesse sentido, o presente trabalho aborda os metadados e a preservação digital no contexto do domínio bibliográfico, com o foco no padrão PREMIS, dessa forma o objetivo é apresentar um panorama de preservação e de curadoria digital focando nos conceitos e funções dos metadados para preservação digital; abordar a estrutura e as características do padrão PREMIS e destacar a relação do PREMIS com a Web Semântica.

Como procedimentos metodológicos optou-se pelo estudo bibliográfico e documental acerca dos aspectos da preservação do objeto digital e de ações do ciclo de vida da curadoria digital, em que são atribuídos os metadados nas ações que deverão ser feitas durante todo ciclo de vida da curadoria Digital; nas ações sequenciais para que sejam cumpridas as funções básicas de prover informações a partir do documento. Nessa linha foi feita a revisão bibliográfica dos temas *Web Semântica, Linked Data* e do padrão de metadados em questão.

2 PRESERVAÇÃO E CURADORIA DIGITAL

A preservação das informações a longo prazo tornou-se um dos grandes desafios do século XXI visto que, somente por meio da preservação, informações propiciarão a compreensão da história humana, caso contrário, a perda, a destruição dessas representações, conduzirá ao rompimento do conhecimento.

Nesse contexto reter as características únicas da informação desde o momento de sua criação e conservá-la por um longo tempo, garantindo sua confiabilidade, sua integridade, originalidade e acesso se configura como o maior desafio da preservação digital. Frente a tais questões a preservação alcança novo foco, pois à consciência do armazenamento, do acesso, da comunicação, da disseminação, do uso e reuso da informação, no presente e no futuro, serão capazes de contribuir para a identidade cultural da sociedade que os produziu.

Ainda que a capacidade de registrar informações no ambiente Web tenha aumentado de maneira acelerada, em contraponto, vivenciamos a diminuição da longevidade dos meios usados para que a informação seja armazenada, justificando a preocupação e o desafio para com a preservação digital devido à acelerada obsolescência tecnológica, digital, à fragilidade do objeto e à manutenção dos dados. (GONÇALEZ, 2017)

Para Hedstrom (1998, p. 191, tradução nossa), “[...] nossa capacidade de criar, acumular e armazenar materiais digitais excede em muito a nossa capacidade de preservá-los.” Assim, temos urgência de preservar os recursos informacionais que foram e estão sendo

produzidos e armazenados, utilizando-se das tecnologias disponíveis e acessíveis.

Ainda segundo o autor supracitado por preservação entende-se “assegurar proteção à informação de valor permanente para acesso pelas gerações presentes e futuras (HEDSTROM, 1998, p. 191). O autor ainda pontua sobre a necessidade de incluir a preservação digital em planejamentos e ações que serão realizadas pelas organizações ao afirmar que é necessário o “[...] planejamento, alocação de recursos e aplicação de métodos e tecnologias para assegurar que informações digitais de valor contínuo permaneça acessível e utilizável.” (HEDSTROM, 1998)

Nesse sentido, as Tecnologias de Informação e Comunicação (TIC) foram impactadas com as formas de materialização da informação, visto que, grande parte de sua gênese se encontra em ambientes informacionais digitais, tornando um desafio a bibliotecários e arquivistas a preservação digital de longa duração.

Segundo Lyman (2002), mais de 90% (noventa por cento) dos objetos digitais são produzidos nos ambientes informacionais digitais, ou seja, são natos digitais, tornando-os suscetíveis a alterações, podendo ser corrompidos devido à sua fragilidade ou ainda devido à dependência tecnológica, pois se são vulneráveis à rápida obsolescência a que a tecnologia está sujeita. Nesse sentido é necessário que informações sejam transportadas para esquemas de metadados.

Nessa linha Siebra et al. (2013) nos apontam que a fragilidade digital, frente à evolução tecnológica nos impõe o desafio de preservá-la e torná-la acessível para as futuras gerações. O que se vislumbra resolver com o surgimento da Curadoria Digital, a qual é vista como um processo mais completo que trata do planejamento, da avaliação e da reavaliação das ações em prol da curadoria do objeto digital que engloba a preservação digital como parte do seu ciclo (SIEBRA et al., 2013).

Ainda que o foco da curadoria digital seja a gestão por todo ciclo de vida do material digital, para que seja possível sua recuperação e acesso, a curadoria amplia a capacidade dos dados serem recuperados por serem expressos por metadados, que configurados como ferramentas significativas para os procedimentos de controle e autenticação (HIGGINS,2011).

Sayão e Sales (2012, p. 180) apontam que o conceito de Curadoria Digital surge para que seja feita a gestão ativa dos dados de pesquisa e seu principal desafio está na necessidade de,

[...] preservar não somente o conjunto de dados, mas de preservar, sobretudo, a capacidade que ele possui de transmitir conhecimento para uso futuro das comunidades interessadas. Isto significa que os ativos genuínos da

pesquisa científica devem permitir que futuros usuários reanalisem os dados dentro de novos contextos.

Destaca-se que a Curadoria Digital reflete uma abordagem holística para o processo de gerenciamento do objeto digital. Assim no ano de 2012, o *Digital Curate Centre - DCC* propôs um modelo de Ciclo de vida da Curadoria, que tem como elementos-chave: os dados, os objetos digitais e o banco de dados.

No centro do ciclo de vida e das demais ações, está o dado digital, o qual se configura como qualquer informação codificada em formato binário, podendo tanto ser dado nascido digital, como também poderá ser migrado a processo de digitalização.

Partindo do centro do Ciclo de Vida da Curadoria Digital, que se encontram os dados, os objetos digitais e o banco de dados o *Digital Curation Centre (2008)* apontam as ações que deverão ser realizadas em todo o ciclo de vida, ações sequenciais e ações ocasionais, explicitadas no quadro 1 a seguir:

Quadro 1 - Ações do Ciclo de Vida da Curadoria Digital

Para todo ciclo de Vida	
Descrição e representação da informação	Coletar e atribuir informações que representem a informação, atribuindo metadados administrativos, descritivos técnicos, estruturais e de preservação, usando padrões apropriados, para garantir a descrição e controle adequado sobre a longo prazo.
Planejamento da preservação	Planejar a preservação em todo o ciclo de vida curadoria do objeto digital; incluindo planos para a gestão e administração de todas as ações de ciclo de vida de curadoria.
Acompanhamento e participação da comunidade	Ter atenção com as atividades que são desenvolvidas nas comunidades comprometidas com a curadoria, com também a necessidade de participar no desenvolvimento de padrões de ferramentas e de softwares adequados aos problemas e que possam ser compartilhados.
Curadoria e preservação	Estar em alerta continuamente e realizar ações administrativas e gerenciais planejadas para promover a curadoria e preservação ao longo do ciclo de vida da curadoria.
Ações Sequenciais	
Conceitualização	Conceber e planejar a criação de dados, incluindo os métodos de captura e as opções de armazenamento.
Criação e recebimento	Criar dados, incluindo metadados administrativos, descritivos, estruturais e técnicos; metadados de preservação poderão também ser adicionado no momento da criação; receber dados, em conformidade com políticas documentadas de seus criadores, que poderão ser: de outros arquivos, de repositórios ou de centros de dados e, se necessário atribuir metadados apropriados para a curadoria e preservação dos dados recepcionados.
Avaliação e seleção	Avaliar os dados e selecionar qual será objeto dos processos de curadoria e de preservação por longo prazo, manter-se aderente tanto às boas práticas, quanto às políticas pertinentes como também às exigências legais.
Arquivamento	Transferir os dados para um arquivo, repositório, centro de dados ou outro custodiante; aderir à

	orientação documentada, políticas ou requisitos legais.
Preservação	Empreender ações para garantir a preservação a longo prazo e retenção do dado de natureza oficial; ações de preservação devem garantir que o dado permanece autêntico, confiável e útil, mantendo a sua integridade; as ações incluem a limpeza de dados, validação, atribuição de metadados de preservação e de informação de representação e a garantia de estruturas de dados ou formatos de arquivos aceitáveis.
Armazenamento	Armazenar os dados de forma segura, mantendo a aderência aos padrões relevantes.
Acesso uso e reuso	Certificar que os dados estão acessíveis para a sua comunidade-alvo ou pelos demais usuários interessados pelo reuso dos dados; isto pode ser na forma de disposição das informações publicadas; o controle de acesso e procedimentos de autenticação podem ser aplicados.
Transformação	Criar novos dados a partir do original, por exemplo, por migração para formatos diferentes ou por meio da criação de um subconjunto - feita pela seleção ou consulta, para criar resultados que podem ser publicados.
Ações Pontuais	
Disponibilizar	Disponibilizar os dados, que não tenha sido selecionado para curadoria e preservação à longo prazo de acordo com as políticas documentadas, orientações ou requisitos legais.
Reavaliar	Retornar dados em que houve falha procedimentos de validação para nova avaliação e possível seleção para curadoria.
Migrar	Migrar dados para um formato diferente; isto pode ser feito para compatibilizá-lo com o ambiente de armazenamento ou para assegurar a imunidade dos dados devido a obsolescência de hardware ou software.

Fonte: Adaptado de Higgins (2008)

A partir do quadro acima, podemos ter uma visão sobre todo processo que envolve a curadoria digital bem como, quais às ações que devem ser realizadas para a preservação dos objetos digitais. Com o foco nas ações "Descrição e representação da informação" e na "Curadoria e preservação" a próxima seção irá discutir a questão dos metadados no contexto da preservação digital, abordando nesse contexto o *PREservation Metadata: Implementation Strategies* (PREMIS).

2 METADADOS E PREMIS

É indissociável falar em preservação digital sem falar em metadados. A importância, princípios, fundamentos e funções dos metadados foram trabalhados por diversos autores como Baca (1998, 2016), Mendez Rodrigues (2002), Duff e McKemmish (2000), Duval et al. (2002), Zeng e Qin (2008, 2016), Alves (2005; 2010), Alves e Santos (2013), Simionato (2012; 2015), Pomerantz (2015) e entre outros.

Segundo Alves (2010, p. 47), os metadados são

[...] atributos que representam uma entidade (objeto do mundo real) em um

sistema de informação. Em outras palavras, são elementos descritivos ou atributos referenciais codificados que representam características próprias ou atribuídas às entidades; são ainda dados que descrevem outros dados em um sistema de informação, com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação.

Segundo o documento “Understanding metadata: what is metadata, and what is it for?” publicado pela NISO e redigido por Riley (2017) classifica os metadados em: **Metadados descritivos** que são utilizados para representar um recurso com o intuito de dar acesso, localizar e encontrar um recurso. **Metadados administrativos** que são subdivididos em: **Metadados técnicos** que são utilizados para decodificar e renderizar arquivos; **Metadados de preservação** que auxiliam no gerenciamento a longo prazo; e **Metadados de direitos** que estão relacionados à propriedade intelectual do conteúdo. **Metadados estruturais** que fazem as relações de partes de recursos entre si. E por fim, a **linguagem de marcação** que integra metadados e sinalizadores para outros recursos estruturais ou semânticos no conteúdo.

Com a capacidade do armazenamento no meio digital crescendo exponencialmente, há uma preocupação latente em como garantir que esses recursos possam ser recuperados a curto e longo prazo. Pensando nesse cenário, foi criado o grupo de trabalho ‘PREMIS’ que significa ‘PREservation Metadata: Implementation Strategies’ (Metadados de preservação: Estratégias de Implementação), pela *Online Computer Library Center* (OCLC) e pelo *Research Libraries Group* (RLG) no período de 2003 a 2005. O resultado da reunião deste grupo foi um relatório que recebeu o nome de *PREMIS Data Dictionary for Preservation Metadata* (PREMIS Dicionário de Dados para Preservação de Metadados), e é incluso um dicionário de dados e um pouco de narrativa sobre preservação de metadados. (CAPLAN, 2009).

Quadro 2 - Evolução da atualização do PREMIS

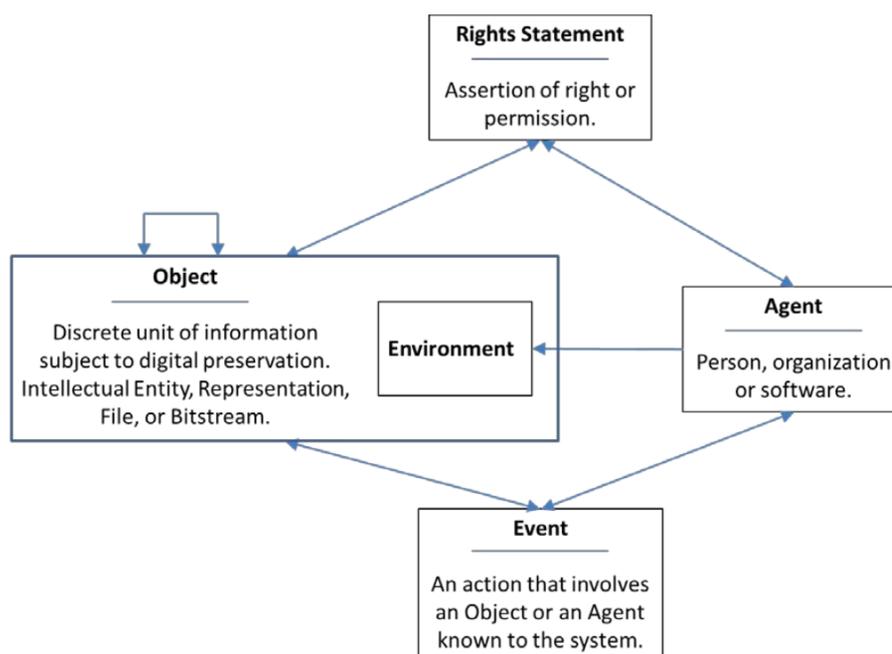
Versão	Data de lançamento
Versão 1.0	Mai de 2005
Schema Version 1.0	17 Maio de 2005
Schema Version 1.1	27 Setembro de 2005
Versão 2.0	Março de 2008
Schema Version 2.0	17 Julho de 2008
Versão 2.1	Janeiro de 2011
Schema Version 2.1	6 Janeiro de 2011
Versão 2.2	Julho de 2012
Schema Version 2.2	15 Maio de 2012
Schema Version 2.3	4 Agosto de 2014
Versão 3	Junho de 2015

Fonte: Premis Data Dictionary for Preservation Metadata (2015, p. 4, tradução nossa).

O padrão PREMIS foi desenvolvido para a preservação dos recursos informacionais em ambientes digitais e está baseado no modelo *Open Archival Information System* (OAIS). De acordo com o Conselho Nacional de Arquivos (2014, p. 22, grifo do autor) “[...] apresenta um conjunto básico (core) de elementos de metadados de preservação para apoiar sistemas que gerenciam objetos digitais. [...] e seu principal documento de referência é o PREMIS *Data Dictionary for Preservation Metadata*.”

O PREMIS define cinco entidades sobre as quais é possível falar, a saber, *Environment* (Suporte), *Object* (Objeto), *Event* (Evento), *Agent* (Agente), e *Rights Statement* (Declaração de direitos), entidades estas que possuem papéis diretamente relacionados a questão da preservação digital, como mostra a ilustração a seguir.

Figura 1 - Modelo de Dados do PREMIS 3.0



Fonte: Premis Data Dictionary for Preservation Metadata (2015, p. 6).

O Premis *Data Dictionary for Preservation Metadata* (2015, p. 7, grifo do autor, tradução nossa) exhibe a definição para as cinco entidades:

Objeto (ou Objeto Digital): uma unidade discreta de informações sujeita a preservação digital. A versão 3 introduz a noção de que este pode ser um suporte usado como parte do processo de preservação.

Suporte: Tecnologia (software ou hardware) que suporta um Objeto Digital de alguma forma (por exemplo, renderização ou execução). Suportes podem

ser descritos como Entidades Intelectuais e capturados e preservados no repositório de preservação como Representações, Arquivos e / ou Bitstreams.

Evento: uma ação que envolve ou afeta pelo menos um objeto ou agente associado ou conhecido pelo repositório de preservação.

Agente: pessoa, organização ou programa / sistema de software associado a eventos na vida de um objeto ou com direitos associados a um objeto. Também pode ser relacionado a um ambiente Objeto que atua como um Agente.

Declaração de direitos: afirmação de um ou mais direitos ou permissões pertencentes a um objeto e / ou agente.

Caplan (2009, p. 4, tradução nossa) esclarece que o dicionário PREMIS:

[...] define um conjunto básico de unidades semânticas [...] que os repositórios devem saber, a fim de desempenhar as suas funções de preservação. As funções de preservação podem variar de um repositório para outro, mas geralmente inclui ações para garantir que os objetos digitais permanecem viáveis (isto é, pode ser lido em qualquer mídia) e renderizável (ou seja, pode ser exibido, reproduzido ou de outra forma interpretado por software de aplicação), bem como para garantir que os objetos digitais do repositório não sejam inadvertidamente alteradas, e que as mudanças legítimas para objetos são documentados.

No entanto, não é intenção do dicionário definir todos os elementos de preservação de metadados desta maneira, são definidos somente àqueles que a maioria dos repositórios necessitam saber na maior parte do tempo, assim, as categorias de metadados que foram excluídas são:

Metadados de formatos específicos. Os metadados que pertence a apenas um formato de arquivo ou classe de formatos, como áudio, vídeo ou gráficos vetoriais.

Metadados específicos de aplicação em regras e negócios. Metadados que descreve as políticas e práticas de um repositório individual, tais como a forma como ele fornece acesso a materiais.

Metadados descritivos. Embora o recurso de descrição seja pertinente para a preservação, muitas normas independentes podem ser utilizadas para este fim, tal como MARC21, MODS, e Dublin Core.

Informações detalhadas sobre a mídia ou hardware. Mais uma vez, embora claramente relevante para a preservação, esta categoria de metadados foi deixada para que outras comunidades a definam.

Informações sobre agentes (pessoas, organizações ou software) com exceção do mínimo necessário para a identificação.

Informações sobre os direitos e permissões, exceto aquelas que afetam diretamente as funções de preservação. (CAPLAN, 2009, p. 4-5, tradução nossa, grifo nosso).

Ao se pensar em quais metadados serão necessários por uma organização para a preservação de objetos digitais, o PREMIS define somente àqueles que são necessários para

efetuar funções de preservação destes objetos.

Como o foco este trabalho está na atribuição dos metadados que visam cumprir a função básica de prover informações sobre o recurso informacional para preservação a longo prazo, a próxima seção irá abordar como o PREMIS pode adequar às tecnologias da *Web Semântica*.

3 WEB SEMÂNTICA E PREMIS OWL

No contexto das novas tecnologias, a *Web Semântica* vem para agregar novas possibilidades de acesso, de identificação e principalmente de recuperação da informação. Segundo seus criadores Berners-Lee, Hendler e Lassila (2001, p. 2): “A *Web Semântica* não é uma *Web* separado, mas uma extensão da atual, em que a informação tem um significado definido, permitindo que computadores e pessoas trabalhem melhor em cooperação.”

Assim, a *web semântica* permite que agentes computacionais trabalhem em cooperação com humanos, facilitando as buscas pelo usuário, visto que não será preciso mais coletar dados em diferentes websites, para que suas necessidades informacionais sejam satisfeitas, logo os agentes computacionais ao entenderem a necessidade dos usuários, lhe entregam com eficiência aquilo que lhe interessa.

Seus criadores ainda pontuam que a *Web Semântica* já está a caminho e que “Em um futuro próximo, este desenvolvimento irá inaugurar novas funcionalidades significativas, de maneira que as máquinas se tornarão mais capazes de processar e ‘entender’ os dados do que simplesmente exibi-los.” (BERNERS-LEE; HENDLER; LASSILA, 2001, p. 2).

Nesse sentido podemos vislumbrar que a *Web Semântica* possibilitará que diferentes conteúdos serão compreendidos e gerenciados independentes do gênero que pertença, ou seja, este poderá ser sonoro, audiovisual, visual, publicações em mídias sociais, *wikis*, mas pelo valor semântico que possuem.

As tecnologias da *Web Semântica* já vêm sendo discutida por diversos autores como Santarem Segundo (2004); Alves (2005); Ramalho (2006); Silva, R. (2013); e Ferreira, J. (2014). Entretanto, Nowack (2009) apresenta uma nova forma de visualizar as tecnologias da *Web Semântica*, indicando conceitos e abstrações bem como suas respectivas especificações e soluções.

Nowack (2009) explica que a base está na Plataforma *Web* e que tem como solução o *Hypertext Transfer Protocol* (HTTP), protocolo que possibilita a comunicação entre computadores e o Unicode que é um esquema de codificação que permite computadores

representar e manipular, texto de qualquer sistema de escrita existente.

Para questões relacionadas aos formatos (*Formats*), o autor (NOWACK, 2009) destaca que foram desenvolvidas diversas soluções como a *eXtensible Markup Language* (XML) que é uma linguagem de marcação recomendada pelo W3C para criar padrões de comunicação entre sistemas de computadores; o *Terse RDF Triple Language* (Turtle) é um formato para expressar dados no modelo *Resource Description Framework* (RDF); e o *RDF in attributes* (RFDa) possibilita escrever triplas RDF na linguagem (X)HTML (SIKOS, 2015). A troca de informações (*Information Exchange*) é considerada por Nowack (2009) como a chave do processo e é estabelecida pelo modelo RDF. O RDF consiste em “[...] uma infraestrutura que permite a codificação, o intercâmbio e o reuso de metadados estruturados.” (FERREIRA, 2014, p. 45).

Nesse cenário, a composição de outras tecnologias (principalmente da troca de informações em RDF) permitem que a busca (Query) acesse os dados por meio do *SPARQL Query Language for RDF* (SPARQL). Para Santarem Segundo (2014, p. 3870) o SPARQL é “[...] um conjunto de especificações que fornecem linguagens e protocolos para consultar e manipular o conteúdo publicado em RDF na Web.”

Ainda na concepção de Nowack (2009), os modelos da Web Semântica estão relacionados à *Web Ontology Language* (OWL), que segundo Méndez e Greenberg (2012) consiste em uma linguagem para definir e instanciar Ontologias na Web. Em paralelo, o *Resource Description Framework Schema* (RDFs), surgiu como uma extensão do RDF e é caracterizado como uma linguagem de descrição de vocabulários que permite a criação de hierarquias, classes, propriedade, subclasses e subpropriedades. (BREITMAN, 2005; SILVA, R. 2013).

Há ainda o *Simple Knowledge Organization System* (SKOS), que de acordo com Catarino (2014) é “[...] um modelo para expressar a estrutura básica e o conteúdo de sistemas de organização do conhecimento.” Miles e Bechhofer (2009), Ramalho (2016, não paginado) complementam que o SKOS é “[...] um modelo de dados para a representação de Sistemas de Organização do Conhecimento no ambiente Web, com o objetivo de atender às novas demandas informacionais relacionadas ao compartilhamento e à interoperabilidade de vocabulários no ambiente Web.”

As Regras (*Rules*) da Web Semântica são estabelecidas por meio do *Rule Interchange Format* (RIF) que segundo Ramalho e Ouchi (2011, p. 68) “[...] é um padrão para intercâmbio de regras lógicas, possibilita maior interoperabilidade entre as regras utilizadas em diferentes

sistemas e baseia-se na concepção de que uma linguagem única não abrangeria todos os paradigmas de representação do conhecimento utilizados;” ou seja, é a base para estruturação da Lógica (*Logic*). Por sua vez, a segurança (*Security*) e a prova (*Proof*) são base para a estruturação da confiança (*Trust*). (ARAKAKI, 2016).

Nesse contexto, criado no ano de 2006 por Berners-Lee, o *Linked Data*, se utiliza de conceitos da Web Semântica, no qual insere significado nos dados da Web. Seu criador apresenta diretrizes para a implementação do *Linked Data* como a criação de bases de dados orientados por normas que facilitem a inserção de significado a esses dados.

Esse cenário propiciou que diferentes áreas do conhecimento publicassem seus dados adotando os princípios do *Linked Data*, principalmente no que se refere às publicações científicas, tem se percebido iniciativas para criação e compartilhamento da informação no que diz respeito à publicação dos dados que são provenientes dos resultados das pesquisas científicas, como artigos de periódicos e trabalhos apresentados em eventos (CONEGLIAN; LUZ; SANTARÉM, 2017).

As publicações de dados de pesquisa é um fator importante na divulgação do conhecimento científico o que é apontado por Lynch (2011) como uma ação que possibilitará a reprodução dos resultados oferecendo comprovação e validação para a qualidade do trabalho científico.

Para o sucesso na construção e manutenção dos *datasets* que disponibilizam e fazem a disseminação das pesquisas científicas, as instituições responsáveis por tais *datasets* devem estruturá-los de forma que a gestão ocorra de maneira eficiente bem como o acesso e a recuperação da informação para que, dessa maneira satisfaçam as necessidades informacionais de seu público alvo, composto quase que exclusivamente por pesquisadores e comunidades científicas.

No contexto do ambiente digital, o uso das tecnologias da Web Semântica tem papel fundamental para multidimensionar as possibilidades de representação e recuperação da informação e também para preservação digital. Nesse sentido, a *Library of Congress* (EUA) tem trabalhado na inserção do PREMIS no contexto da Web Semântica utilizando a OWL.

A ontologia PREMIS OWL é uma formalização semântica do dicionário de dados PREMIS 2.2 e define um modelo conceitual para a informação de preservação de um arquivo digital. A ontologia PREMIS OWL permitiu à comunidade interessada expressar metadados de preservação no RDF, usando o modelo conceitual do PREMIS-DD e, como tal, pode ser usado para publicar a informação de preservação como *Linked (Open) Data*. (DI IORIO; CARON, 2016, p. 2)

Segundo Coppens et al. (2015), para cada entidade do PREMIS foi estabelecida como uma classe OWL, configurando-se em: *IntellectualEntity*, *Object*, *Event*, *RightsStatement* e *Agent*. Essas classes estão relacionadas entre si usando as propriedades do objeto *hasObject*, *hasEvent*, *hasAgent*, *hasIntellectualEntity*, *hasRelationship* e *hasRightsStatement*.

No contexto da Web Semântica, as marcações das relações entre as entidades do PREMIS podem apresentar resultados significativos para buscas específicas e identificar relacionamentos entre entidades e recursos.

4 CONSIDERAÇÕES FINAIS

Em síntese, percebe-se que os metadados assumem um papel fundamental nas descrições e representações dos recursos informacionais no ambiente digital e possibilitam também a interoperabilidade dos sistemas de bibliotecas, arquivos e repositórios digitais, além do acesso do usuário ao conteúdo. A preservação digital possui uma grande importância mundial, pois tem o papel de armazenar as informações no meio digital para que qualquer pessoa tenha acesso à informação que desejam de onde estiver e a qualquer momento a longo prazo.

Nesse contexto, o PREMIS teve grande importância por fomentar discussões acerca da preservação digital desde 2005, quando foi lançado. Desde então, houve uma preocupação de sua constante atualização.

Com a inserção da Web Semântica, novas possibilidades de uso desses dados podem ser aplicados. As relações entre as entidades ganham semântica, possibilitando a ligação que antes não eram possíveis ou previstas, mas que no cenário atual tornaram-se imprescindíveis, o que nos leva a constatar que estudos devem ser realizados nas temáticas de preservação e metadados.

Semantic Web and digital preservation: the metadata standard PREMIS in the proposal of Linked Data

ABSTRACT

From the significant increase in the production of digital content, actions aimed at the recovery and preservation of digital objects in the long term become of capital importance in our day, otherwise there will be a break in knowledge due to loss or insufficiency in the representation of the In this context, Curatorial Digital is configured as an interdisciplinary area that manages information with the proposal and the challenge of preservation and transmission of informational objects, acting on the risk factors such as obsolete technology, technological fragility, lack understanding of what is good digital practice and inadequate resources. In this scenario, a clipping was made about concepts

and functions of metadata for digital preservation. This paper aims to verify how the metadata standard Preservation Metadata: Implementation Strategies (PREMIS) can be used to meet the preservation demands of digital information objects. The literature review was chosen as a methodological procedure, since it is a reflection on the representation of information against the background of the semantic Web. As a result we approached the structure and characteristics of the PREMIS metadata standard and its relation to the Semantic Web.

Keywords: *Preservation Metadata. Preservation Metadata: Implementation Strategies. PREMIS. Semantic Web.*

REFERÊNCIAS

- ALVES, R. C. V. **Web semântica: uma análise focada no uso de metadados.** 2005. 180 f. Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2005. Disponível em: <<http://repositorio.unesp.br/handle/11449/93690>>. Acesso em: 12 nov. 2017.
- ALVES, R. C. V. **Metadados como elementos do processo de catalogação.** 2010. 132 f. Tese (Doutorado em Ciência da Informação)–Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, 2010. Disponível em: <<http://hdl.handle.net/11449/103361>>. Acesso em: 12 nov. 2017.
- ALVES, R. C. V.; SANTOS, P. L. V. A. da C. **Metadados no domínio bibliográfico.** Intertexto: Niterói, 2013.
- ARAKAKI, F. A. **Linked data: ligação de dados bibliográficos.** 2016. 144 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, 2016. Disponível em: <<https://repositorio.unesp.br/handle/11449/147979>>. Acesso em: 12 nov. 2017.
- BACA, M. (ed.). **Introduction to Metadata: pathways to digital information.** Los Angeles, CA: Getty Information Institute, 1998.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific american**, v. 284, n. 5, p. 28-37, 2001.
- BREITMAN, K. **Web semântica: a Internet do Futuro.** LTC, 2005.
- CAPLAN, P. **Understanding PREMIS.** Washington, DC: Library of Congress Network Development and MARC Standards Office, 2009. Disponível em: <<http://www.loc.gov/standards/premis/understanding-premis.pdf>>. Acesso em: 12 nov. 2017.
- COPPENS, S. et al. **PREMIS OWL: A semantic long-term preservation model.** *Int J Digit Libr*, v. 15, n. 87, 2015. Disponível em: <<https://link.springer.com/article/10.1007/s00799-014-0136-9#citeas>>. Acesso em: 12 nov. 2017.
- CONEGLIAN, C. S.; LUZ, L. P. SANTAREM SEGUNDO, J. E. Boas práticas para publicação de dados na web: aplicação nos dados referentes aos resultados de pesquisa científica. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA

INFORMAÇÃO (UNESP), 18., 2017. Marília. **Anais...** Marília: UNESP, 2017. p. 1-21. Disponível em: <

<http://enancib.marilia.unesp.br/index.php/xviiienancib/ENANCIB/paper/viewFile/382/835>>

Acesso em: 10 nov. 2017.

CONSELHO NACIONAL DE ARQUIVOS. **Diretrizes para a implementação de repositórios digitais confiáveis de documentos arquivísticos**. Rio de Janeiro: Câmara Técnica de Documentos Eletrônicos, 2014. Disponível em:

<http://www.conarq.arquivonacional.gov.br/media/publicacoes/conarq_repositorios_2014.pdf>. Acesso em: 12 nov. 2017.

DI IORIO, A.; CARON, B. **PREMIS 3.0 Ontology**: Improving Semantic Interoperability of Preservation Metadata. EUA: Library of Congress, 2016. Disponível em:

<https://www.loc.gov/standards/premis/pif/2016/iPresDiIoria_Caron_iPRES_2016_paper_150.pdf>. Acesso em: 12 nov. 2017.

DUFF, W.; MCKEMMISH, S. Metadata and ISO 9000 Compliance. **Information Management Journal**, v. 34, n. 1, 2000. Disponível em:

<<http://www.infotech.monash.edu.au/research/groups/rcrg/publications/smckduff.htm>>. Acesso em: 12 nov. 2017.

DUVAL, E. et al. Metadata principles and practicalities. **D-Lib Magazine**, v. 8, n. 4, 2002.

Disponível em: <<http://www.dlib.org/dlib/april02/weibel/04weibel.html>>. Acesso em: 12 nov. 2017.

FERREIRA, J. A. **Wikis semânticos**: da Web para a Web Semântica. 2014. 131 f.

Dissertação (Mestrado em Ciência da Informação). Faculdade de Filosofia e Ciências – Universidade Estadual Paulista, Marília, 2014. Disponível em:

<<http://repositorio.unesp.br/handle/11449/108380>>. Acesso em: 12 nov. 2017.

GILLILAND-SWETLAND, A. Defining Metadata. In: BACA, M. (ed.), **Introduction to Metadata**: pathways to digital information. Los Angeles, CA: Getty Information Institute, 1998.

GONÇALEZ, P. R. V. A. **Repositórios arquivísticos digitais confiáveis**: identificação de requisitos com ênfase no acesso à informação. 2017. 166 f. Tese (Doutorado em Ciência da Informação)-Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2017.

HEDSTROM, M. Digital preservation: a time bomb for digital libraries. **Computer and the Humanities**, n. 31, p. 189-202, 1998. Disponível em:

<<https://doi.org/10.1023/A:1000676723815>>. Acesso em: 12 nov. 2017.

HIGGINS, S. The DCC Curation Lifecycle Model. **International Journal of Digital**

Curation, v. 3, n. 1, p. 134-140, 2008. Disponível em: <<https://doi.org/10.2218/ijdc.v3i1.48>>.

Acesso em: 12 nov. 2017.

LYMAN, P. Archiving the world wide web. **Council on Library and Information Resources**, 2002. Disponível em: <<https://www.clir.org/pubs/reports/pub106/web.html>>. Acesso em: 12 nov. 2017.

MÉNDEZ RODRÍGUEZ, E. **Metadatos y recuperación de información**: estándares, problemas y aplicabilidad en bibliotecas digitales. Trea: Espanha, 2002.

MENDEZ, E. GREENBERG, J. Linked data for open vocabularies and HIVE's global framework. **El profesional de la información**, mayo-junio, v. 21, n. 3, 2012. Disponível em: <<http://recyt.fecyt.es/index.php/EPI/article/view/epi.2012.may.03/17916>>. Acesso em: 12 nov. 2017.

MILES, A.; BECHHOFFER, S. SKOS: **Simple Knowledge Organization System Reference**. W3C Recommendation, 2009. Disponível em: <<http://www.w3.org/TR/skos-reference/>>. Acesso em: 12 nov. 2017.

NATIONAL INFORMATION STANDARD ORGANIZATION. **Understanding Metadata**. Bethesda, MD: NISO Press, 2004. Disponível em: <<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>. Acesso em: 12 nov. 2017.

NOWACK, B. **The Semantic Web Technology Stack (not a piece of cake...)**. Linked Data Developer. [Online] 2009. Disponível em: <<http://linkeddatadeveloper.com/Projects/LinkedData/media/fig11.2.png>>. Acesso em: 12 nov. 2017.

POMERANTZ, J. **Metadata**. USA: The MIT press essential knowledge series, 2015.

PREMIS DATA DICTIONARY FOR PRESERVATION METADATA. **Data Dictionary for Preservation Metadata**: PREMIS version 3.0. [S.l.: s.n.], 2015. Disponível em: <<http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>>. Acesso em: 12 nov. 2017.

RAMALHO, R. A. S. **Web semântica**: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da ciência da informação. 2006. 120 f. Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2006. Disponível em: <<http://repositorio.unesp.br/handle/11449/93709>>. Acesso em: 12 nov. 2017.

RAMALHO, R. A. S. Análise dos modelos de dados SKOS e BIBFRAME: novas perspectivas de representação na era dos dados interligados. ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., **Anais...** Bahia, ANCIB; UFBA, nov. 2016. Disponível em: <<http://www.ufpb.br/evento/lti/ocs/index.php/enancib2016/enancib2016/paper/view/4111>>. Acesso em: 12 nov. 2017.

RAMALHO, Rogério Aparecido Sá; OUCHI, Marcos Teruo. Tecnologias Semânticas: Novas Perspectivas para a Representação de Recursos Informacionais. **Informação & Informação**, Londrina, v. 16, n. 3, p. 60-75, dez. 2011. ISSN 1981-8920. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/9829/10643>>. Acesso em: 12 nov. 2017. doi:<http://dx.doi.org/10.5433/1981-8920.2011v16n3p60>.

RILEY, J. **Understanding metadata**: what is metadata, and what is it for? Baltimore: NISO, 2017. Disponível em: <http://www.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf>. Acesso em: 12 nov. 2017.

SANTAREM SEGUNDO, J. E. **Recursos tecno-metodológicos para descrição e recuperação de informações na Web**. 2004. 157 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2004. Disponível em: <<http://repositorio.unesp.br/handle/11449/93618>>. Acesso em: 12 nov. 2017.

SANTAREM SEGUNDO, J. E. Web semântica: introdução a recuperação de dados usando SPARQL. ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 15., Minas Gerais. **Anais...** Minas Gerais, ANCIB; UFMG, 2014. Disponível em: <<http://enancib2014.eci.ufmg.br/documentos/anais/anais-gt8>>. Acesso em: 12 nov. 2017.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade**: Estudos, João Pessoa, v. 22, n. 3, p. 179-191, set./dez. 2012.

SIEBRA, S. A. et al. Curadoria digital: além da questão da preservação digital. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 14., 2013. Florianópolis. **Anais...** Rio de Janeiro. UFSC, 2013, p. 1-20. Disponível em: <<http://basessibi.c3sl.ufpr.br/brapci/index.php/article/view/0000013754/cfb64a34a6903340c12b82d9319d014>>. Acesso em: 12 nov. 2017.

SIKOS, L. F. **Mastering structured data on the Semantic Web**: from HTML5 microdata to Linked Open Data. EUA: Apress, 2015.

SILVA, R. E. da. **As tecnologias da web semântica no domínio bibliográfico**. 134 f. 2013. Dissertação (Mestrado em Ciência da Informação)–Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Filosofia e Ciências, Marília, 2013. Disponível em: <<http://repositorio.unesp.br/handle/11449/93653>>. Acesso em: 12 nov. 2017.

ZENG, M. L. QUIN, J. **Metadata**. New York: Neal-Schuman Publishers, 2008.