



XVII Encontro Nacional de Pesquisa em Ciência da Informação (XVII ENANCIB)

GT 8 – Informação e Tecnologia

UM SISTEMA BASEADO NA CONSTRUÇÃO COLETIVA DE CORPO DE CONHECIMENTO PARA APOIAR AS ATIVIDADES DE UM CENTRO DE DIGITALIZAÇÃO

A SYSTEM BASED ON THE COLLECTIVE CONSTRUCTION OF A BODY OF KNOWLEDGE TO GUIDE THE WORKS OF A DIGITIZINZ CENTRE

Rafael Port da Rocha¹

Modalidade da apresentação: Comunicação Oral

Resumo: O desenvolvimento de um centro de digitalização requer que seus profissionais detenham domínio compartilhado sobre conhecimentos e práticas em digitalização. Esse artigo apresenta uma abordagem baseada em corpo de conhecimento, para permitir que membros de um centro de digitalização registrem e compartilhem conhecimentos e práticas adquiridos a partir de estudos e experimentos em digitalização. Destaca que a construção desse corpo possui uma natureza diferente, por ser feita coletivamente pelos membros do centro, quando comparada com corpos tradicionais, como BPM CBOK e PMBOC. Apresenta a estrutura do corpo de conhecimento, em que os recursos são descritos através de uma ontologia. Destaca a estrutura básica da ontologia; apresenta as áreas de conhecimento definidas para digitalização e processos ligados a essas áreas; e mostra como o software Semantic Media Wiki é usado para dar suporte à construção coletiva do corpo de conhecimento. Conclui apresentando o andamento da construção de um corpo de conhecimento para um centro de digitalização.

Palavras-chave: Digitalização. Sistemas de Construção Coletiva do Conhecimento. Ontologia. Web Semântica.

Abstract: *The development of a digitization centre requires from its professionals to have a shared domain of knowledge and digitizing practices. This paper presents an approach based on body of knowledge, to allow members of a digitization centre to record and share knowledge and practices, acquired from studies and digitizing experiments. It points out that the construction of this body of knowledge has a different nature, when compared to traditional bodies of knowledge (such as PMBOK and BPM CBOC), since it is collectively constructed by the members of the centre. The paper depicts the structure of the body of knowledge, which involves ontology. It highlights the basic*

¹ Professor do Departamento de Ciência da Informação da Universidade Federal do Rio Grande do Sul. Graduação, Mestrado e Doutorado em Ciência da Computação pela Universidade Federal do Rio Grande do Sul.

structure of the ontology. It presents the knowledge areas defined for digitizing, and the processes related to these areas; and it shows how the software Semantic Media Wiki is used to support the collective construction of the body of knowledge. It concludes, showing the progress of the construction of a body of knowledge in a digitizing centre.

Keywords: *Digitization. Collective Knowledge Systems. Ontology. Semantic Web.*

1 INTRODUÇÃO

O desenvolvimento e operação de um centro de digitalização requer, por parte de seus profissionais, domínio comum e compartilhado sobre conhecimentos e práticas em digitalização. Isso envolve, por exemplo, conhecer e executar procedimentos de seleção dos materiais a serem digitalizados, de configuração das imagens digitais, de construção do fluxo de digitalização, de geração dos pacotes que contém os arquivos da digitalização, de armazenamento seguro e por longo prazo dos documentos digitais, e de produção de metadados descritivos, administrativos, técnicos e estruturais. Um dos grandes desafios para um centro de digitalização está em desenvolver estratégias para organizar, capturar, construir e utilizar essas práticas e conhecimentos, em um processo coletivo e evolutivo. Esse artigo aborda o desenvolvimento de um corpo de conhecimento para dar apoio a um centro de digitalização na realização de suas atividades.

Com o objetivo em auxiliar seus profissionais no exercício de suas atividades, comunidades e associações profissionais de diversas áreas desenvolvem corpos de conhecimento. Um corpo de conhecimento contém o conhecimento estruturado que é usado por membros de uma disciplina para guiar suas práticas e trabalhos (ÖREN, 2005). Esse corpo representa a agregação de conhecimento estabelecida para uma área particular que é esperada ser aprendida a fundo por um indivíduo para ser considerado ou certificado como um praticante (ÖREN, 2005).

Os corpos de conhecimento são desenvolvidos por especialistas consagrados, que trazem seus conhecimentos e suas experiências; e são patrocinados por associações ou comunidades profissionais. No caso do centro de digitalização, a estratégia a ser adotada é desenvolver um corpo de conhecimento para apoiar seus membros no desenvolvimento e disseminação de seus conhecimentos e práticas em digitalização. Esse corpo de conhecimento possui algumas características particulares, quando comparado aos corpos de conhecimento tradicionais, visto que sua construção não é feita por meio de especialistas ligados a associações ou comunidades voltadas à digitalização. O desenvolvimento desse corpo é realizado de forma coletiva e evolutiva pelos membros do centro de digitalização, no decorrer de suas atividades, a

partir da busca do conhecimento em digitalização, da experimentação prática, assim como da produção de novos conhecimentos e práticas, ampliando o corpo. O foco é dar apoio a um centro de digitalização através da construção, operação e manutenção de um corpo de conhecimento, a fim de permitir que esse centro realize suas ações de acordo com as melhores práticas, conectadas com a pesquisa, assim como registrar e disseminar os conhecimentos adquiridos.

O desenvolvimento do corpo de conhecimento proposto para o centro de digitalização passa por desafios e questionamentos como: Qual é a estrutura para o corpo de conhecimento? Quais são as áreas de conhecimento que envolvem a digitalização? Como desenvolver um coletivamente um corpo de conhecimento, por membros de um centro de digitalização, a partir da busca de conhecimento e da experimentação na prática? Que ferramentas poderiam ser usadas para dar apoio à construção desse corpo de conhecimento? Que estruturas e ferramentas poderiam ser usadas para registrar e disseminar esse corpo de conhecimento?

Tendo como base essas questões, a construção do corpo de conhecimento observou os seguintes objetivos e procedimentos metodológicos:

- Propor uma estrutura para o corpo de conhecimento em digitalização, a partir da análise das estruturas de corpos de conhecimento existentes, identificando como esses corpos são organizados e quais são os seus principais componentes;
- Identificar as principais áreas de um corpo de conhecimento em digitalização, a partir da busca na literatura das principais publicações em digitalização, tendo como enfoque livros didáticos que abordam a digitalização em um aspecto sistêmico e voltados à prática, assim como publicações científicas que aprofundam esses conhecimentos;
- Identificar ferramentas de apoio para a construção coletiva do corpo de conhecimento, a partir da investigação de ambientes colaborativos, enfocando aspectos de cooperação, coordenação e comunicação, conforme o modelo 3C (FUCKS et al, 2007)
- Descrever e organizar formalmente, de forma compreensível por máquinas (computadores), os recursos que compõem o corpo de conhecimento, através da construção de uma ontologia, para a plataforma da Web Semântica, enfatizando a reutilização de ontologias pelo método MIREOT (TORNIAI et al., 2013).

Este artigo discute essas questões, a partir de resultados obtidos no desenvolvimento de um corpo de conhecimento (chamando neste artigo de CCDig) para o Centro de Documentação e Acervo Digital da Pesquisa (CEDAP) da Universidade Federal do Rio Grande do Sul. Como esse corpo de conhecimento é desenvolvido de forma coletiva e evolutiva, suas estruturas, métodos, estratégias e conhecimentos estão em constante evolução. Esse artigo reflete esse corpo em seu primeiro ano de desenvolvimento, em que as prioridades são: a definição de uma estrutura para o corpo; a identificação das principais áreas de conhecimento que envolvem a digitalização; a identificação dos processos relacionados à essas áreas; e definição e implantação de um ambiente computacional para dar suporte a esse corpo e ao seu desenvolvimento continuado.

Esse artigo está organizado da seguinte forma: a seção 2 apresenta três corpos de conhecimento consagrados, enfatizando suas estruturas, que serviram de base para a definição da estrutura do corpo de conhecimento para o centro de digitalização. A seção 3 discute e apresenta características do corpo de conhecimento em digitalização considerando seu estado de desenvolvimento atual. A seção 3.1 apresenta como o CCDig está estruturado; a seção 3.2 apresenta as áreas de conhecimento em digitalização identificadas para compor o corpo de conhecimento; a seção 3.3 apresenta processos de planejamento e execução até então identificados como práticas de cada área de conhecimento; e a seção 3.4 aborda a ferramenta computacional que dá suporte à representação e ao desenvolvimento coletivo do CCDig. Considerações finais são trazidas na seção 4.

2 CORPOS DE CONHECIMENTO

Em muitas áreas, a preocupação em ter profissionais que apliquem na prática aquilo que se demonstrou o mais avançado, combinando a prática e a pesquisa, se concretizou através da elaboração de guias, chamados de corpos de conhecimento. Um corpo de conhecimento identifica áreas de conhecimento relevantes para guiar trabalhos ou práticas de uma disciplina, podendo incluir a especificação de ciclos de vidas, fluxos e processos ligados a essas práticas. PMBOK (PROJECT MANAGEMENT INSTITUTE, 2013), BPM CBOOK (BENEDICT et al., 2013), SWEBOOK (BOURQUE et al., 2014) são exemplos de corpos de conhecimento desenvolvidos por associações profissionais a fim de possibilitar que seus profissionais desenvolvam suas práticas conectadas com que existe de mais atual. Esses três corpos de conhecimento, respectivamente, foram desenvolvidos para as áreas de gestão de projetos, gestão de processos e engenharia de software.

O PMBOK (Project Management Body of Knowledge) é um guia para corpo de conhecimento em gerência de projetos. Constitui-se em um vocabulário comum para gestão de projetos, desenvolvido pela PROJECT MANAGEMENT INSTITUTE (PMI), instituição sem fins lucrativos, composta por profissionais em gestão de projetos. O guia PMBOK descreve áreas de conhecimento em gestão de projetos, e processos de gerenciamento para cada área de conhecimento. Os processos são classificados em iniciação, planejamento, execução, controle e encerramento; e a especificação de cada processo envolve a definição das entradas, das saídas e das ferramentas e técnicas usadas para gerar saídas a partir das entradas.

Por exemplo, os processos **Coleta de Requisitos** e **Controle do Escopo** fazem parte da área de conhecimento **Gerenciamento do Escopo do Projeto**, e são classificados em processos de planejamento e controle, respectivamente. O processo de **Coleta de Requisitos** tem, como entradas, o termo de abertura do projeto e o registro das partes interessadas (*stakeholders*). As saídas desse processo são: a documentação dos requisitos e o plano de gerenciamento de requisitos, e suas ferramentas e técnicas são: dinâmicas de grupos, oficinas, técnicas de criatividade em grupo, entre outras.

BPM CBOK (Business Process Management Common Body of Knowledge) estabelece um corpo de conhecimento em processos de negócio. Foi desenvolvido pela Association of Business Process Management e está estruturado em áreas de conhecimento. BPM CBOK tem como propósito ser

uma referência básica para profissionais de BPM [Business Process Management]. O propósito primário é identificar e fornecer uma visão geral das áreas de conhecimento necessárias para a prática de BPM. Inclui papéis e estruturas organizacionais, bem como provisões para conduzir uma organização orientada por processos. Fornece para cada área de conhecimento uma visão geral, uma lista de tópicos comuns associados, links e referências para outras fontes de informação que fazem parte do corpo de conhecimento mais amplo sobre BPM. Também pretende ser um ponto de referência para discussões entre profissionais de BPM (BENEDICT et al., 2013)

SWEBOK (Software Engineering Body of Knowledge) é um corpo de conhecimento para Engenharia de Software, publicado pela IEEE Computer Society. É composto por áreas de conhecimento em Engenharia de Software e por áreas de conhecimento relacionadas à Engenharia de Software. Cada área decomposta em tópicos e subtópicos. SWEBOK também apresenta, em anexo, padrões da IEEE e ISO/IEC que dão suporte à Engenharia de Software. O quadro 1 apresenta as principais estruturas desses três corpos de conhecimento.

Quadro 1 – Principais elementos de MBOK, BPM CBOK e SWEBOK.

PMBOK	BPM CBOK	SWEBOK
Áreas de Conhecimento Processos (entradas, saídas, ferramentas, técnicas)	Papéis, Estrutura Organizacional, Áreas de Conhecimento (referências) Tópicos	Áreas de Conhecimento Tópicos Áreas Relacionadas Padrões Relacionados

Fonte: Autor, compilada de PMBOK, BPM CBOC e SWEBOK

3 CORPO DE CONHECIMENTO PARA O CENTRO DE DIGITALIZAÇÃO

Essa seção discute e apresenta características do corpo de conhecimento em digitalização (CCDig), considerando seu estado de desenvolvimento atual, enfatizando sua estruturação, áreas de conhecimento, processos e ferramenta computacional de suporte.

3.1 ESTRUTURA DO CORPO DE CONHECIMENTO

A primeira questão a ser abordada na construção do CCDig é definir como o corpo de conhecimento será estruturado. Os principais elementos do CCDig foram trazidos a partir de uma análise das estruturas de PMBOK, BPM CBOK e SWEBOC. A escolha desses três corpos de conhecimento deve-se aos mesmos serem consagrados, e terem ligações com a digitalização, pois tratam de aspectos de planejamento, definição de processo e software.

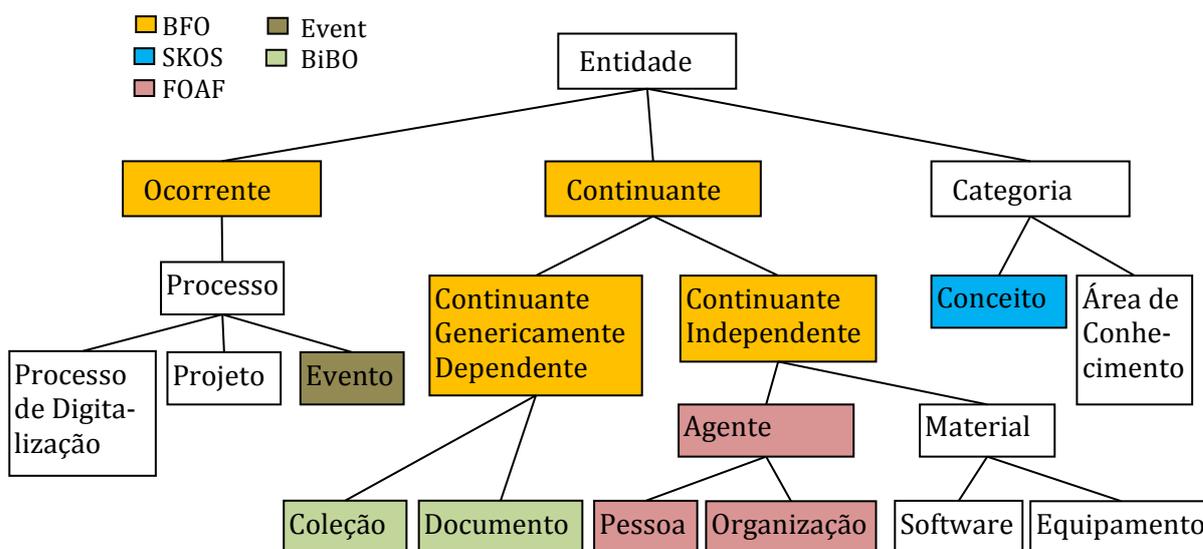
Os três principais elementos do CCDig são: área de conhecimento, processo e conceito. Áreas de conhecimento compreendem nas áreas que os profissionais de um centro de digitalização devem dominar. Os processos correspondem ações ligadas a cada área de conhecimento, a serem desempenhadas por esses profissionais. Os processos são classificados planejamento, execução, controle, indicado atuação em planejar algo, em executar algo ou em controlar o andamento dessa execução. Os conceitos representam tópicos que envolvem a digitalização.

Uma ontologia está sendo desenvolvida para identificar e organizar os recursos abordados pelo corpo de conhecimento. Isso permite a representação formal e compreensível por máquina dos elementos envolvidos, possibilitando o desenvolvimento de recursos computacionais inteligentes, permitindo assim uma melhor compreensão sobre a natureza dos recursos envolvidos. A ontologia que está sendo desenvolvida representa não somente os principais elementos que fazem parte do corpo de conhecimento, mas também todos aqueles

recursos que envolvem um centro de digitalização. Fazem parte dessa ontologia, os instrumentos básicos do CCDig (área de conhecimento, processo, conceito), recursos que operacionalizam a digitalização (equipamentos, scanners, pessoas, software, suportes, gêneros e tipos documentais, coleções, guias, etc.) e recursos externos que dão apoio à digitalização (projetos de digitalização, centros de digitalização, bases de dados, guias, publicações, etc.).

Essa ontologia, ao buscar reconhecer aquilo que está envolvido com a digitalização, traz para o corpo de conhecimento recursos como scanners, guias, métodos, softwares, projetos, cujo conhecimento sobre os mesmos é adquirido/produzido coletivamente e de forma evolutiva. O desenvolvimento dessa ontologia tem como requisitos: (a) ter como base uma ontologia de alto nível, a fim de produzir uma ontologia bem fundamentada, com facilidades para extensões; (b) reutilizar ontologias populares, visando interoperabilidade; e (c) atuar na plataforma da web semântica, visando permitir a interligação de dados (Linked Data). Nesse sentido, o desenvolvimento da ontologia envolve o uso da ontologia de fundamentação Basic Formal Ontology (SPEAR, 2006); e são reutilizadas ontologias como VIVO-ISF (CORSON-RIKERT, J., 2012; TORNIAI et al., 2013), para recursos acadêmicos; FOAF (BRICKLEY; MILLER, 2014), para pessoas e organizações; SKOS (WORLD WIDE WEB CONSORTIUM, 2012), para conceitos; Event (RAIMOND; ABDALLAH, 2007), para eventos; BiBO (D'ARCUS; GIASSON, 2009), para recursos bibliográficos; entre outras. Para a reutilização das ontologias, MIREOT (TORNIAI et al., 2013) é o método adotado, pois permite a incorporação seletiva de classes das ontologias reutilizadas. A figura 1 apresenta uma o esboço resumido das classes mais genéricas da ontologia. A figura 1 apresenta, resumidamente, as principais classes da ontologia.

Figura 1 – Hierarquia de classes resumida da Ontologia



3.2 ÁREAS DE CONHECIMENTO

A digitalização de acervo é uma área que possui uma base sólida e bem definida, com práticas e conhecimentos sendo transmitidos a profissionais através da publicação de melhores práticas. Nesse sentido, analisando pesquisas, guias e melhores práticas, observamos que as seguintes áreas estruturam o conhecimento em digitalização: seleção, conversão, controle de qualidade, produção das imagens, metadados, armazenamento e acesso, planejamento e gerenciamento da digitalização.

Essas áreas foram desenvolvidas a partir da pesquisa na literatura científica e em sites de projetos e centros de digitalização de acervos. Elas compõem o corpo de conhecimento em seu estado atual de desenvolvimento, caracterizando um ponto de partida para seu aprimoramento via construção coletiva. Isso significa que mudanças podem ocorrer ao longo do desenvolvimento coletivo de CCDig. As áreas de CCDig estão apresentadas no quadro 2. Esse quadro também apresenta também outras áreas que dão suporte à digitalização.

Quadro 2 – Áreas de Conhecimento em Digitalização

Áreas de Conhecimento em Digitalização	Áreas Relacionadas
<ul style="list-style-type: none"> • Seleção • Conversão • Controle de qualidade • Produção das Imagens • Metadados • Armazenamento e Acesso • Planejamento e Gerenciamento da Digitalização 	<ul style="list-style-type: none"> • Direito Autoral • Processamento de Imagens • Gerenciamento de Projetos • Gerenciamento de Processos de Negócio • Bases de Dados • Preservação e Curadoria Digital

Fonte: Autor

A **seleção** compreende em identificar as necessidades dos usuários, e conhecer os atributos dos documentos e a infraestrutura técnica (CORNELL UNIVERSITY LIBRARY, 2003; KENNEY; RIEGER, 2000). Envolve também identificar os direitos relacionados com a obra e sua digitalização. Nessa área, ocorre a elaboração de critérios de seleção (gerais ou específicos para um determinado gênero de documento, como livro raro), e políticas de direito autoral. Hoje, conhecimentos e práticas dessa área estão registrados em guias, como guia de seleção desenvolvido pelo CLIR - Council on Library Information Resources (HAZEN; HORRELL; MERRILL-OLDHAM, 1998) e o guia de seleção de obras raras desenvolvido pela IFLA - International Federation of Library Associations and Institutions (IFLA RARE BOOK AND SPECIAL COLLECTIONS SECTION, 2014). Essa área está relacionada ao conhecimento em direito autoral.

A **conversão** envolve caracterizar os documentos a serem digitalizados de acordo com a forma com que os originais foram produzidos (como texto impresso, manuscrito, mapa) e realizar a configuração da imagem digital para os tipos de documentos originais caracterizados (definir resolução, modo, profundidade, compactação, formato de arquivo). Na digitalização, várias versões digitais são geradas, para propósitos distintos, como armazenamento a longo prazo (chamadas de matrizes), acesso (derivadas de acesso), índice (miniaturas) e impressão (derivadas de impressão). A conversão também se apoia na área de computação gráfica, ao usar métodos de processamento de imagem.

Conhecimentos e práticas dessa área estão registrados em guias e recomendações gerais, como o guia americano (FEDERAL AGENCIES DIGITIZATION INITIATIVE - STILL IMAGE WORKING GROUP, 2015), as recomendações brasileiras para digitalização de documentos arquivísticos do Conselho Nacional de Arquivos (CONSELHO NACIONAL DE ARQUIVOS, 2010), o guia do órgão alemão de fomento à pesquisa DFG (DEUTSCHE

FORSCHUNGSGEMEINSCHAFT, 2013) e as recomendações mínimas para digitalização elaboradas pela Associação Americana de Bibliotecários (BOGUS et ali, 2013). Conhecimentos e práticas também são registrados por gêneros específicos de documento, como as do guia digitalização de para manuscritos do projeto Europeana Regia (BACHER, R; FABIAN, C.; IKAS, W.; SCHREIBER, C., 2011).

Controle de qualidade envolve estabelecer parâmetros, critérios e rotinas para avaliar a qualidade da imagem durante o processo de digitalização. Na literatura, encontramos guias e ferramentas (*checklists*) para verificar a qualidade, como instrumento Digitisation Disposal Policy Toolkit (QUEENSLAND STATE ARCHIVES, 2014), assim como critérios de aceitação.

A **produção de imagens** envolve a definição do fluxo de digitalização, assim como a execução e o gerenciamento desse fluxo. Esse fluxo compreende na preparação do material, na captura do material, no controle de qualidade do material digitalizado (incluindo redigitalização), no empacotamento dos arquivos gerados e na submissão desse pacote à ambientes de armazenamento e acesso. Nessa área, observamos que instituições desenvolvem suas próprias soluções, criando fluxos que integram várias ferramentas, como os projetos das universidades de Leiden (VAN BERGEM, 2014) e Utah (NEATROUR, 2014). Essa área apoia-se na gestão de processos de negócio (PM CBOK) para gerenciar os processos de digitalização. Associações voltadas à certificação de profissionais que trabalham com a gestão da informação e imagem em organizações (focadas para a gestão de documentos correntes) desenvolveram manuais de apoio à certificação que incluem recomendações de captura de imagens, como os manuais para as certificações CDIA+ (MOORE, 2014) e AIIM-CPF (WEISSMAN E MANCINI, 2012).

Produzir **metadados** compreende em registrar e armazenar informações sobre os objetos digitalizados. Metadados são geralmente classificados (CAPLAN, 2003) em: descritivos (descrevem informações sobre o conteúdo intelectual/artístico do recurso), estruturais (descrevem como os elementos físicos e intelectuais que formam um recurso são agrupados), administrativos (descrevem informações para ajudar a gerenciar o recurso, como preservação digital, direitos de acesso, processo de digitalização, aquisição) e técnicos (metadados administrativos que descrevem aspectos técnicos do documento, como resolução). Área de metadados obteve grandes avanços referentes à padronização, com a definição e adoção de

vários esquemas padrões de metadados, como MODS² e EAD³ (descritivos), PREMIS⁴ (administrativo/preservação digital), MIX⁵ (administrativo/técnico) e METS⁶ (estrutural).

Na digitalização, metadados estruturais assumem um papel fundamental, pois são usados para montar as estruturas dos documentos digitalizados, assim como para agregar os outros tipos de metadados usados. O padrão METS possui elementos para a descrição dos arquivos e das estruturas lógicas dos documentos, assim como para encapsular ou relacionar metadados descritivos (em Dublin Core ou MODS, por exemplo), técnicos (em MIX, por exemplo) e administrativos (como PREMIS). Para o corpo de conhecimento, é crucial dominar conhecimentos e práticas tanto no uso e combinação de metadados descritivos (HEERY e PATEL, 2000; CHANG e ZENG 2006); quanto nas estratégias de aplicação de METS, como as observados nos estudos de Guenter (2008), para combinar METS e PREMIS, e de Veermaten (2010), que auxiliam a definição de perfis em METS, e os perfis de uso do METS, como o perfil australiano (PEARCE, 2008) e o recomendado pelo DFG (DEUTSCHE FORSCHUNGSGEMEINSCHAFT, 2013).

Na digitalização, são geradas várias representações para documentos originais (matriz e derivadas), e metadados estruturais são usados para descrever essas representações e as estruturas lógicas dos documentos (páginas, seções, capítulos). Por exemplo, a digitalização de um livro pode levar a produção de matrizes em TIFF, na forma de um arquivo TIFF por página; e de uma derivada de acesso, em que o livro é representado em um arquivo PDF. Metadados especificados em METS podem descrever essas representações, indicando também como os arquivos TIFF são organizados para compor as páginas e as seções do documento. Exemplos de projetos de digitalização que usam o padrão METS para estruturar os arquivos que compõem um documento digitalizado são: Europeana Regia (BACHER, R; FABIAN, C.; IKAS, W., SCHREIBER, C., 2011), Fragmenta Membranea (HAKALA, et al., 2012) e Digital Library of the Caribbean (DIGITAL LIBRARY OF THE CARIBBEAN, 2011). Para descrever estruturalmente um documento digitalizado, foram desenvolvidas ferramentas como: como Goobi/KiToDo⁷, RescartaToolkit⁸, ContentE⁹ e Sobek CM¹⁰.

² Metadata Object Description Schema - <http://www.loc.gov/standards/mods/>

³ Encoded Archival Description - <http://www.loc.gov/ead/>

⁴ Preservation Metadata - <http://www.loc.gov/standards/premis/>

⁵ NISO Metadata for Images in XML - <http://www.loc.gov/standards/mix//>

⁶ Metadata Encoding & Transmission Standard - <http://www.loc.gov/standards/mets/>

⁷ Goobi/Kitodo - <http://www.kitodo.org/>

⁸ Rescarta ToolKit <http://www.rescarta.org/>

⁹ ContentE <http://purl.pt/index/geral/PT/infoProfContentE.html>

¹⁰ Sobek - <http://sobekrepository.org/>

Armazenamento e acesso compreendem no empacotamento de um objeto digital decorrente da digitalização de um documento, e o envio deste pacote a um repositório digital confiável, que será responsável em prover o armazenamento e o acesso confiável e a longo prazo (preservação digital). Isso implica em estruturar o resultado da digitalização (DAPERT e ENDERS, 2008), gerando pacotes de submissão de informação para serem enviados a repositórios confiáveis (CENTER FOR RESEARCH LIBRARIES, ONLINE COMPUTER LIBRARY CENTER, 2007), que serão responsáveis pelo armazenamento e disseminação. Essa área se apoia nas áreas que desenvolvem conhecimentos em preservação digital e repositórios digitais confiáveis, à medida que os objetos digitalizados passam a ser de responsabilidade desses repositórios.

A área de **Planejamento e gerenciamento da digitalização** compreende em desenvolver e gerenciar um projeto de digitalização. A área de gestão de projeto já possui um corpo de conhecimento, denominado PMBOK. Com isso, o planejamento e o gerenciamento de um projeto de digitalização implica em aplicar PM BOOK, tendo como base as atividades e os fluxos de produção de imagem, assim como registrar e disseminar os conhecimentos e práticas adquiridos.

3.3 PROCESSOS

A estruturação do CCDig contempla a definição de processos, que indicam ações a serem desenvolvidas por profissionais de um centro de digitalização para as áreas de conhecimento especificadas. Esses processos são classificados em processos de planejamento, execução e controle, seguindo a estratégia de PMBOK.

Os processos de planejamento são aqueles que (a) visam estabelecer critérios e recomendações para serem usadas/seguidas na execução de projetos de digitalização; e (b) aqueles processos ligados ao planejamento de um projeto específico de digitalização. Esses processos estão presentes em todas as áreas de conhecimento em digitalização. Por exemplo, a área de **conversão** possui o processo **Definir recomendações para a configuração da imagem**. Esse processo tem, como resultado, guias que recomendam quais são as melhores alternativas para definir resolução, profundidade, compactação e formato de arquivo, para a digitalização de um documento original com determinadas características (como um manuscrito). Esses guias são usados no planejamento de projeto de digitalização, pelo processo de planejamento, **Definir configurações para as imagens em um projeto de digitalização**, que visa estabelecer quais serão as configurações adotadas para digitalizar os documentos do

projeto.

Os processos de execução são realizados quando um projeto de digitalização é desenvolvido, como por exemplo, o processo **Captura da Imagem**, que compreende na digitalização das imagens de um projeto, de acordo com as configurações estabelecidas. Processos de controle verificam a execução dos processos.

O mapeamento dos processos do centro de digitalização proporciona o desenvolvimento de um órgão em que suas ações são institucionalmente definidas, estabelecidas e reconhecidas, trazendo transparência e agilidade à sua gestão. Esses processos estão atrelados a áreas de conhecimentos em digitalização, e são desenvolvidos coletivamente pela equipe. Isso possibilita não somente o registro das práticas realizadas, mas também dos conhecimentos adquiridos, em um processo contínuo.

Assim como em PMBOK, em CCDig os processos possuem entradas, saídas e registram técnicas utilizadas para gerar saídas a partir de entradas. A ontologia de CCDig permite a descrição de recursos dessa natureza, como, por exemplo, de recomendações definidas pelo processo **Definir recomendações para a configuração da imagem** ou de guias da literatura usados por esse processo para produzir suas recomendações.

O quadro 3 apresenta os processos de planejamento e execução que estão sendo especificados para cada área de conhecimento em digitalização. Como o centro de digitalização está em fase de implantação, muitos desses processos ainda não foram executados. Muitos desses processos indicam um planejamento, e serão validados e aprofundados ao longo do desenvolvimento do centro.

Quadro 3- Processos

	Processos de Planejamento	Processos de Execução
Seleção	<ul style="list-style-type: none">• Definir critérios de seleção• Definir requisitos para digitalização• Analisar requisitos em projeto de digitalização• Aplicar critérios de seleção em projeto de digitalização	<ul style="list-style-type: none">• Preparação do Material
Conversão	<ul style="list-style-type: none">• Definir recomendações para configuração da imagem• Definir recomendações para OCR• Definir recomendações para escolha de um scanner• Definir de recomendações para	<ul style="list-style-type: none">• Captura da Imagem• Processamento da Imagem Capturada

	<p>captura de imagem</p> <ul style="list-style-type: none"> Definir configurações para as imagens em um projeto de digitalização 	
Controle de qualidade	<ul style="list-style-type: none"> Elaborar recomendações para critérios de avaliação da qualidade da imagem Elaborar recomendações para rotinas de avaliação da qualidade da imagem Definir rotina para avaliar a qualidade da imagem em um projeto de digitalização Definir critérios de avaliação da qualidade da imagem um projeto de digitalização 	<ul style="list-style-type: none"> Verificar a qualidade da imagem
Produção das Imagens	<ul style="list-style-type: none"> Elaboração de recomendações para fluxo de digitalização definir o fluxo de digitalização em um projeto de digitalização 	<ul style="list-style-type: none"> Executar fluxo de digitalização
Metadados	<ul style="list-style-type: none"> Elaboração de recomendações para padrões de metadados Elaboração de recomendações para construção de Perfis de Aplicação Estabelecer os Padrões de Metadados para um projeto de digitalização Elaborar Perfis de Aplicação para um projeto de digitalização 	<ul style="list-style-type: none"> Descrição do Recurso
Armazenamento e Acesso	<ul style="list-style-type: none"> Elaboração de recomendações para empacotamento Definir serviço para armazenamento e acesso para um projeto (de acordo com o políticas e planos de preservação digital da instituição) 	<ul style="list-style-type: none"> Geração do Pacote de Submissão Ingestão Acesso
Planejamento e Gerenciamento da	<ul style="list-style-type: none"> Desenvolver o plano de gerenciamento do projeto (segundo PMBOK): Definir escopo do projeto; elaborar a estrutura analítica do projeto; definir, sequenciar e estimar as atividades do projeto; estimar custos do projeto; planejar a qualidade; desenvolver o plano de recursos humanos; planejar as comunicações; planejar riscos; planejar aquisições 	<ul style="list-style-type: none"> Gerenciar Projeto de Digitalização

3.3 CONTRUÇÃO E OPERAÇÃO DO CORPO DE CONHECIMENTO

O desenvolvimento de CCDig ocorre em um processo contínuo e coletivo, em que cada novo conhecimento ou prática desenvolvida é registrado. Isso implica em um ambiente de construção colaborativa. Para dar suporte computacional ao desenvolvimento de CCDig, foram estabelecidos os seguintes requisitos: (a) permitir a construção colaborativa, focada no modelo 3C (FUCKS et al, 2007), que envolve o desenvolvimento da Cooperação (produção cooperativa de textos), da Coordenação (mediação, controle e registro de mudanças/histórico) e da Comunicação (discussão, troca de mensagens, fóruns); (b) permitir o registro do conhecimento em forma textual e através de descrições semânticas (sentenças RDF, produzidas de acordo com ontologias).

Três ferramentas foram analisadas, para escolha daquela que mais se adequa para servir como ambiente para representação e construção do CCDig: Semantic Media Wiki (SMW) (VÖLKEL et al., 2006), OntoWiki (AUER; DIETZOLD; RIECHERT, 2006) e Fabrico/Ciência (ROCHA, 2014).

Semantic MediaWiki (VÖLKEL et al., 2006) é a ferramenta MediaWiki (utilizada na Wikipedia) estendida com recursos semânticos. Permite a edição colaborativa de textos, registra o histórico de cada texto, possibilita o estabelecimento de mecanismos de moderação e de comunicação, através de fóruns de discussão. Permite a especificação de classes e propriedades de ontologia, e as instâncias dessas classes são textos wiki que contém sentenças produzidas de acordo com as propriedades especificadas.

OntoWiki (AUER; DIETZOLD; RIECHERT, 2006) é uma ferramenta para autoria, apresentação e gestão de bases de conhecimento. Usuários produzem colaborativamente descrições semânticas na linguagem RDF e de acordo com ontologias. OntoWiki oferece recursos para edição e apresentação intuitiva dos dados; para gerar visões e agregações da base de conhecimento; para gerenciar mudanças; para promover a colaboração (envolvendo discussões, votações); para prover estatísticas e para distribuir informação (HEINO, 2009).

Fabrico/Ciência é um anotador baseado em ontologias construído de acordo com a arquitetura da Web Semântica e seguindo o modelo de negócio de O'Reilly (2007) para sistemas da Web 2.0. O ambiente usa técnicas da Web 2.0 para capturar contribuições dos usuários através de anotações semânticas, expressas em RDF e de acordo com ontologias. Incorpora wikis, fóruns, folksonomias, recursos de exploração de dados (coautoria, cotermos) (ROCHA, 2014).

SMW, OntoWiki e Fabrico/Ciência operam na plataforma da Web Semântica (RDF/OWL), fomentam a construção coletiva, oferecem recursos como fóruns e registram o histórico/mudanças. OntoWiki e Fabrico/Ciência têm as sentenças semânticas (sentenças RDF) como elementos base. Já em SMW, os elementos da base são os textos wiki, que representam conceitos semânticos, como classes, propriedades e instâncias. SMW foi escolhida para o CCDig, e o motivo principal dessa escolha foi a sua natureza básica ser focada na construção coletiva de textos.

O ambiente SMW foi configurado da seguinte forma, para representar e permitir o desenvolvimento do CCDig:

- As classes e propriedades da ontologia (figura1) foram especificadas em SMW;
- Páginas wiki são criadas para representar cada recurso do corpo do conhecimento (como uma área de conhecimento ou scanner), e cada página é descrita como uma instância de uma classe especificada na ontologia ;
- Propriedades especificadas na ontologia são usadas nos textos para indicar características dos recursos;
- Os textos e a ontologia são desenvolvidos coletivamente, com apoio de recursos do SMW como fórum de discussão, registro do histórico e moderação.

A figura 2 exemplifica um recurso do CCDig. A página wiki visualizada na figura 2 corresponde ao software Goobi Production. Nela, são registrados textualmente os conhecimentos e as práticas adquiridas pela equipe sobre o software. No contexto das ontologias, essa página é uma instância da classe Software (categoria: Software, na figura 2). Por ser uma instância de Software, além de textos, a página também contém as propriedades especificadas na ontologia para Software: nome, sistema operacional e licença. Essas propriedades são operadas pelo ambiente como sentenças RDF sobre o recurso. A construção coletiva desse recurso é apoiada por uma lista de discussão (aba “**Discussão**”, na figura 2) e pelo histórico. Na aba “Ver Histórico”, é possível ter acesso à evolução da construção de **Goobi Production** (colaborações e seus colaboradores).

Figura 2 – Página Wiki que representa o software Goobi Production

Página [Discussão](#) [Ler](#) [Editar](#) [Ver histórico](#) [★](#) [Mais ▾](#)

Goobi Production

Nome:	Goobi Production
Sistema Operacional:	Plataforma WWW
Licença:	GNU/GPL

Software usado para gerenciar a digitalização e para descrição estrutural dos recursos digitalizados utilizando a especificação [Goobi METS](#). Atualmente conhecido como [KIToDo](#). [Site](#) [↗](#).

Exemplos [\[editar\]](#)

Universidade de Goettigen

- [Notitia trivm codicvm avtographorvm Iohannis Regiomontani in bibliotheca Christophori Theophili de Mvrr](#) [↗](#): Via [DFG-viewer](#) [↗](#) ou via [METS.xml](#) [↗](#)
- [Würzburger gelehrte Anzeigen im Jahre](#) [↗](#), volume: [Würzburger gelehrte Anzeigen im Jahre Volume 1.1786](#) [↗](#): Via [METS.xml](#) [↗](#) e [TOC](#) [↗](#)

Staatsbibliothek Berlin

- [Abschnitte aus 1001 Nacht](#) [↗](#): Via [DFG-viewer](#) [↗](#) e via [METS.xml](#) [↗](#)

Categoria: [Software](#)

4 CONSIDERAÇÕES FINAIS

O CCDig está sendo construído junto à implantação do CEDAP. Atualmente, o desenvolvimento do CEDAP encontra-se em uma fase que envolve a busca de conhecimentos e práticas em digitalização; a configuração/operação dos scanners adquiridos; e a realização de experimentações de digitalizações, envolvendo materiais originais com diversas características (manuscritos, obras raras, revistas, atas). O objetivo dessa fase é criar uma infraestrutura conceitualmente bem definida, fundamentada em conhecimentos e práticas sólidas, para prover a realização da digitalização em escala.

A solução CCDig vem demonstrando extremamente necessária ao centro de digitalização; tanto na formação e disseminação do conhecimento e de práticas em digitalização; quanto no auxílio à gestão/implantação do centro, à medida que os recursos e os procedimentos gerenciados pelo centro são conhecidos e descritos através do uso de ontologia.

Na implantação do centro de digitalização, houve uma grande pesquisa sobre os conhecimentos e práticas existentes, e o ambiente possibilitou que os resultados dessas pesquisas fossem imediatamente registrados e compartilhados. Isso ocorreu tanto na concepção da ontologia, quando no registro dos conhecimentos adquiridos, referentes aos recursos classificados pela ontologia.

O CCDig também trouxe grandes benefícios às atividades desenvolvidas de configuração dos scanners adquiridos e nas experimentações em digitalização. Possibilitou que

as experiências e os conhecimentos adquiridos sobre os scanners e seus softwares fossem imediatamente registrados. O fato de CCDig incorporar processos em sua ontologia, exigiu que as experimentações em digitalização fossem realizadas a partir de procedimentos definidos em CCDig. Atualmente o CCDig contém a seguinte quantidade de recursos, por classe da ontologia: Área de Conhecimento: 13 (incluindo áreas relacionadas), Computador: 5, Conceito: 24, Equipamento:10, Esquema de Metadados: 7, Formato de Arquivo: 5, Modelo de Scanner: 6, Organização: 1, Processo de Execução: 7, Processo de Planejamento: 18, Scanner: 5, Software: 25, Tipo Documental: 7.

SMW demonstrou-se adequado ao desenvolvimento de CCDig. Tem como base um wiki, fato que oferece agilidade, visto que o conhecimento é registrado de forma hipertextual e colaborativa. Por outro lado, ao usar ontologias, permite que páginas wiki sejam vistas como representantes de recursos de digitalização, que são definidos como instâncias de classes de uma ontologia, contendo propriedades especificadas para essas classes. Sobre esse aspecto, os recursos de digitalização do CCDig são semanticamente descritos através de sentenças RDF. As experiências da construção/desenvolvimento de CCDig via SMW estão sendo compartilhadas/trocadas com o Projeto UNBRAL (ROCHA; DORFMAN, 2015), que busca registrar práticas e conhecimentos na construção de uma base de dados, a partir da coleta de informações relevantes, presentes em outras bases de dados.

O ambiente e a estrutura do CCDig estão em um estágio de desenvolvimento suficiente para atender às próximas atividades do CEDAP, que envolvem tanto a busca de novos conhecimentos/práticas, quanto a execução de projetos de digitalização em escala maior.

REFERÊNCIAS

- AUER, S; DIETZOLD, S; RIECHERT, T. OntoWiki – a tool for social, semantic collaboration. In: THE INTERNATIONAL SEMANTIC WEB CONFERENCE, 5, 2005, Athens, USA. **Proceedings...** Berlin,Heidelberg:Springer, 2006.
- BACHER, R; FABIAN, C.; IKAS, W., SCHREIBER, C. **State of the Art in Image Processing**. Europeanaregia, 2013. Disponível em: <<http://www.europeanaregia.eu/sites/www.europeanaregia.eu/files/europeanaregia-state-of-the-art-in-image-processing.pdf>>. Acesso em: 20 jul. 2016
- BENEDICT, T. et al. **Guide to the Business Process Management Body of Knowledge**, BPM CBOK, Versão 3.0. Association of Business Process Management Professionals, 2013.
- BOGUS, Ian. **Minimum Digitization Capture Recommendations**. Chicago:American Library Association, 2013. Disponível em: <<http://www.ala.org/alcts/resources/preserv/minimum->

digitization-capture-recommendations>. Acesso em: 14 jun. 2015

BOURQUE, Pierre et al. Guide to the Software Engineering Body of Knowledge, Version 3.0. **IEEE Computer Society Press**, 2014.

BRICKLEY, D.; MILLER, L. **FOAF vocabulary specification 0.99**. Namespace Document Paddington Edition, jan. 2014. Disponível em: <<http://xmlns.com/foaf/spec/>>. Acesso em: 02 jan. 2016.

CAPLAN, Priscilla. **Metadata fundamentals for all librarians**. Chicago: American Library Association, 2003.

CENTER FOR RESEARCH LIBRARIES, ONLINE COMPUTER LIBRARY CENTER. **Trustworthy repositories audit & certification (TRAC) criteria and checklist**, Version 1, Centre for Research Libraries, 2007

CHAN, Lois Mai; ZENG, Marcia Lei. Metadata interoperability and standardization: A study of methodology, Part I. **D-Lib Magazine**, v. 12, n. 6, p. 3, 2006.

CONSELHO NACIONAL DE ARQUIVOS. **Recomendações para Digitalização de Documentos Arquivísticos Permanentes**. CONARQ, 2010. Disponível em: <http://www.conarq.arquivonacional.gov.br/images/publicacoes_textos/Recomendacoes_digitalizacao_completa.pdf>. Acesso em: 20 ago. 2016

CORNELL UNIVERSITY LIBRARY. **Theory into Practice Digital Imaging**. 2003. Disponível em: <<http://www.library.cornell.edu/preservation/tutorial/contents.html>>. Acesso em: 20 jul. 2016

CORSON-RIKERT, J. et. al. The VIVO ontology. In: BÖRNER, K; CONLON, M.; CORSON-RIKERT, J.; DING, Y. **VIVO: a semantic approach to scholarly networking and discovery**. US: Morgan & Claypool Publishers, 2012. Chapter 2.

D'ARCUS, B.; GIASSON, F. **Bibliographic Ontology Specification**. 2009. Disponível em: <<http://bibliontology.com/>>. Acesso em: 05 jan. 2016.

DAPPERT, Angela; FARQUHAR, Adam. Modelling organizational preservation goals to guide digital preservation. **International Journal of Digital Curation**, v. 4, n. 2, p. 119-134, 2009

DEUTSCHE FORSCHUNGSGEMEINSCHAFT. **DFG Practical Guidelines on Digitisation**. DFG: Bonn, 2013. Disponível em: <http://www.dfg.de/formulare/12_151/12_151_en.pdf>. Acesso em: 14 jun. 2015

DIGITAL LIBRARY OF THE CARIBBEAN. **dLOC Training Manual**. 2011. Disponível em: <<http://dloc.com/AA00002865/00001>>. Acesso em: 05 jan. 2016

FEDERAL AGENCIES DIGITIZATION INITIATIVE - STILL IMAGE WORKING GROUP. **Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files**, Federal Agencies Digitization Initiative, DRAFT September 2, 2015. Disponível em: <http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image_Tech_Guidelines_2015-09-02_v4.pdf>. Acesso em: 05 jan. 2016

FUKS, H.; RAPOSO, A.; GEROSA, M.A.; PIMENTEL, M.; LUCENA, C.J.P. The 3C Collaboration Model. In: KOCK, N et ali. **The Encyclopedia of E-Collaboration**. Hershey:Information Science, 2007

GUENTHER, Rebecca. Battle of the buzzwords: Flexibility vs. interoperability when implementing PREMIS in METS. **D-Lib Magazine**, v. 14, n. 7/8, 2008.

HAKALA Juha, et al. **FRAGMENTA MEMBRANEA PROJECT 2009–2012**. Final Report. The National Librar of Finland. 2012 . Disponível em: <http://www.doria.fi/bitstream/handle/10024/98923/210314_fragmenta_membranea_978-951-51-0226-3.pdf?sequence=2>. Acesso em: 05 jan. 2016

HAZEN, Dan; HORRELL, Jeffrey; MERRILL-OLDHAM, Jan. Selecting research collections for digitization. **CLIR Report, Pub 74**, 1998.

HEERY, R PATEL, M. Application profiles: mixing and matching metadata schemas. **Ariadne**, n. 25, 2000.

HEINO, N. et al. Developing semantic web applications with the ontowiki framework. In. **Networked Knowledge-Networked Media**. Berlin Heidelberg:Springer, 2009.

IFLA Rare Book and Special Collections Section. **Guidelines for Planning the Digitization of Rare Book and Manuscript Collections**. Haia:International Federation of Library Associations and Institutions , 2014

KENNEY, A; RIEGER, O. **Moving theory into practice**: digital imaging for libraries and archives. Mountain View: Research Libraries Group, 2000.

MOORE, D et ali. **CompTIA CDIA+ Certification Handbook**. CompTIA Learning, 2014.

NEATROUR, Anna et al. The SIMP Tool: Facilitating Digital Library, Metadata, and Preservation Workflow at the University of Utah's J. Willard Marriott Library. **D-Lib Magazine**, v. 20, n. 7, p. 2, 2014.

O'REILLY, T. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. **International Journal of Digital Economics**, n. 65, March 2007.

ÖREN, T. Toward the body of knowledge of modeling and simulation. In: Interservice/industry Training, Simulation, and Education Conference, 2005, Orlando, Flórida. **Proceedings...** 2005.

PEARCE, Judith et al. The Australian METS profile-A journey about metadata. **D-Lib Magazine**, v. 14, n. 3, p. 5, 2008.

PROJECT MANAGEMENT INSTITUTE. **A Guide to the Project Management Body of Knowledge**, PMBOK Guide. Project Management Institute, 2013.

QUEENSLAND STATE ARCHIVES. **Digitization Disposal Policy Toolkit**: Quality Assurance Guidance, 2014. Disponível em <<http://www.archives.qld.gov.au/Recordkeeping/GRKDownloads/Documents/Digitisation-Disposal-Policy-Toolkit-Quality-assurance-guidance.pdf>>. Acesso em: 05 jan. 2016

RAIMOND, Y.; ABDALLAH, S. **The Event Ontology**. 2007. Disponível em: <<http://motools.sourceforge.net/event/event.html>>. Acesso em: 07 jan. 2016.

ROCHA, R.; DORFMAN, A. O UNBRAL fronteiras em busca de boas práticas na divulgação e preservação da produção acadêmica. In: **Anuário Unbral das fronteiras brasileiras 2014**. Projeto Unbral, 2015.

ROCHA, R. FABRICO/CIÊNCIA no Desenvolvimento de Ambientes Linked Data para a Ciência. In.: LINDED OPEN DATA BRASIL. 2014, Florianópolis. **Anais...** Nov. 2014.

SPEAR, A. D. **Ontology for the twenty first century: an introduction with recommendations**. Saarbrücken, Germany: IFOMIS, 2006. Disponível em: < <http://ifomis.uni-saarland.de/bfo/documents/manual.pdf> >. Acesso em: 05 jan. 2016.

TORNIAI, C. et al. Finding common ground: interating the eagle-I and VIVO ontologies, In International Conference on Biomedical Ontology, 4, Montreal, Canada, July 7-12, 2013. Proceedings... **CEUR Workshop Proceedings**, vol 1060, 2013, p. 46-49

VAN BERGEN, Saskia. Connecting Systems for Better Services Around Special Collections. **D-Lib Magazine**, v. 20, n. 9, p. 7, 2014.

VERMAATEN, Sally. A Checklist and a Case for Documenting PREMIS-METS Decisions in a METS Profile. **D-Lib Magazine**, v. 16, n. 9, p. 2, 2010.

VÖLKEL, M., KRÖTZSCH, M., VRANDECIC, D., HALLER, H., & STUDER, R. Semantic wikipedia. International conference on World Wide Web, 15, 2006, Edinburgh, Scotland. **Proceedings...** New York: ACM, 2006

WEISSMAN, S; MANCINI, J. **AIIM Certified Information Profession Study Guide**. Association for Information and Image Management, 2012

WORLD WIDE WEB CONSORTIUM - W3C. **Introdution to SKOS**. 2012. Disponível em: <<http://www.w3.org/2004/02/skos/intro>>. Acesso em: 14 jun. 2015.