

Rafael Port da Rocha, Sonia Caregnato

Grupo Temático 4: Arquitetura e Plataformas de Gestão de Acervos Digitais

Corpo de Conhecimento em Digitalização na Implantação do Centro de Documentação e Acervo Digital da Pesquisa - CEDAP

RESUMO. A preservação e o compartilhamento dos dados de pesquisa são fatores determinantes para a produção da ciência no mundo atual, pois permitem o sua evolução em um processo contínuo, diversificado, compartilhado, distribuído e transparente. O CDAP foi criado para promover a preservação e o acesso dos dados de pesquisa da UFRGS. Uma das atribuições do CEDAP é realizar a digitalização de documentos usados em pesquisa, incluindo obras raras, fotos, microfimes. Esse artigo aborda o conhecimento como ativo do CEDAP para realizar digitalização de documentos. Apresenta um Corpo de Conhecimento em digitalização necessário para o CEDAP desenvolver suas atividades; desenvolve uma estrutura para esse Corpo de Conhecimento a partir da análise dos Guias de Corpos de Conhecimento, PMBOK, BPM BOK e SWEBOK; e analisa infraestrutura para representação e construção coletiva desse Corpo de Conhecimento. Apresenta como solução um Corpo de Conhecimento estruturado em áreas de conhecimento ligadas a conceitos, organizações, projetos, processos, outras áreas, fontes de informação; com textos desenvolvidos coletivamente e semanticamente descritos por ontologia, sendo escolhida como plataforma a ferramenta Semantic Media Wiki, após análise envolvendo também as ferramentas OntoWiki e Fabrico.

Palavras-chave: Digitalização; Corpo de Conhecimento; Gestão de Acervos Digitais ; Semantic Media Wiki

ABSTRACT. Today, the preservation and sharing of research data are determining factors for the production of the science, giving a diversified, shared, distributed and transparent development process. CDAP is a center created to promote the preservation and access of research data in UFRGS. One of the tasks of CEDAP is to digitize documents used in research, including rare books, photographs and microfilms. This article discusses the knowledge as asset of CEDAP to perform document scanning. The article presents a Body of Knowledge in Digitalization necessary to the CEDAP develop its activities. It develops a structure for the Body of Knowledge analyzing PMBOK, BPM BOK and SWEBOK, known guides of Bodies of Knowledge, and it discuss an infrastructure to the representation and collective construction of the Body of Knowledge. It presents as solution a Knowledge Body structured in knowledge areas, which are related to concepts, organizations, projects, processes, other areas and information sources, all represented as texts which are collectively developed and semantically described by ontology. As platform, we chose the Semantic Media Wiki tool, after also analyzing Ontowiki and Fabrico.

Keywords: Digitalization; Body of knowledge; Digital Assets Management; Semantic Media Wiki;

1 Introdução

Dados da Pesquisa são aqueles materiais reconhecidos pela comunidade científica, que foram registrados durante uma investigação, e que servem para certificar os resultados da investigação que se realiza. São também considerados dados de pesquisa, documentos que registram criações intelectuais ou artísticas de interesse para pesquisas. Normalmente, dados de pesquisa encontram-se dispersos em diferentes unidades acadêmicas, sob a custódia de indivíduos ou grupos de pesquisa, ou como acervos de particulares ou de instituições públicas e privadas.

Lord et al (2003), em um estudo sobre dados de pesquisa para o Governo Britânico, apontaram as seguintes razões para o compartilhamento e preservação de dados da pesquisa: permitir reuso em novas pesquisas; manter dados de observação; disponibilizar mais dados para novos projetos; cumprir requisitos legais; permitir a validação de resultados de pesquisa; utilizar os dados no ensino e para o bem público. Hoje, o compartilhamento e a preservação de dados da pesquisa são preocupações de órgãos financiadores americanos e europeus. A Comunidade Europeia, através do programa de pesquisa e inovação, Horizon 2020, espera melhorar o acesso à informação científica e aumentar os benefícios do investimento público. Segundo o Projeto Horizon 2020 (EUROPEAN COMMISSION, 2013), o compartilhamento e a preservação dos dados da pesquisa possibilitam: construir sobre os resultados de pesquisas anteriores (melhoria da qualidade dos resultados); fomentar a colaboração e evitar a duplicação de esforços (maior eficiência); acelerar a inovação; e envolver os cidadãos e a sociedade (aumento da transparência do processo científico). National Science Foundation (NSF), órgão americano que financia projetos de pesquisa, estabelece que pesquisadores devam facilitar e promover o compartilhamento dos dados científicos, e determina que propostas de pesquisas devam incluir planos de gestão de dados (NATIONAL SCIENCE FOUNDATION, 2013).

Em sintonia com essa forma de desenvolver a ciência, o Centro de Documentação e Acervo Digital da Pesquisa (CEDAP) visa dar suporte às pesquisas da Universidade Federal do Rio Grande do Sul (UFRGS) através de serviços para preservação e compartilhamento de dados de pesquisa, incluindo a digitalização de documentos. O CEDAP está em fase de implantação, em que estão sendo especificados e adquiridos os equipamentos de digitalização, configuradas as

instalações, e definidos os processos e as formas de atuação. Nessa fase de implantação, um ativo fundamental para o desenvolvimento do CEDAP é o conhecimento. É crucial para um órgão que está sendo criado, desenvolver, desde seu nascimento, uma infraestrutura para esse ativo, viabilizando sua aquisição, organização, preservação e disseminação.

A identificação de um Corpo de Conhecimento necessário para o desenvolvimento do CEDAP permite um melhor direcionamento nos esforços em adquirir, desenvolver e ampliar esse conhecimento, em um processo de construção continuado e organizado, evitando estudos desnecessários.

Esse artigo aborda o conhecimento como ativo para desenvolver a digitalização no CEDAP. A busca do Corpo de Conhecimento em Digitalização para o CEDAP envolveu os seguintes procedimentos metodológicos: (1) estudo de como são estruturados guias de Corpos de Conhecimento, a identificação da natureza de um Corpo de Conhecimento para o CEDAP e da estratégia para a elaboração desse corpo (seção 3); (2) definição da estrutura do corpo de conhecimento do CEDAP, tendo com referência os corpos de conhecimento estudados (seção 4); e (3) a definição da infraestrutura para representação e construção colaborativa do Corpo de Conhecimento em digitalização para o CEDAP (seção 5). A próxima seção apresenta o CEDAP.

2 O CEDAP

O CEDAP é um órgão da Faculdade de Biblioteconomia e Comunicação (FABICO) da UFRGS que provém infraestrutura para o compartilhamento e curadoria de dados de pesquisa da UFRGS, digitais natos e decorrentes da digitalização. Seus equipamentos foram adquiridos através de recursos da FINEP, Edital CT-INFRA 2010. O CEDAP tem como parceiro o Centro de Processamento de Dados da UFRGS.

Os objetivos do CEDAP são:

- Dar suporte à pesquisa científica, tecnológica, artística e cultural, através da curadoria de ativos digitais de pesquisa, visando o seu reuso, validar resultados, manter dados de observação, utilizar os dados no ensino e para o bem público.
- Promover digitalização e a curadoria de ativos digitais de pesquisa.

- Avançar o estado da arte em digitalização e curadoria de ativos digitais, por meio de uma abordagem interdisciplinar das áreas da Ciência da Informação, Comunicação e Informática, desenvolvendo projetos de investigação, participando de redes, estimulando a produção científica e promovendo o compartilhamento e a disseminação de novos conhecimentos.

Hoje, o cluster de digitalização do CEDAP está sendo implantado, composto por scanners de diversos tamanhos e finalidades. O cluster é composto por scanner planetário A2 (OS 12002), scanner de microfilme (Delta Plus), scanners de produção de documentos A3 (DR-G1100 e mesa, e DR-3030C) e scanner de fotografia e filmes fotográficos A3 (ScanMaker 1000XL Plus), além de infraestrutura de armazenamento via servidor lâmina para Blade. Está situado junto aos laboratórios da FABICO de arquivologia, biblioteconomia, museologia, bases de dados e restauro, cuja proximidade foi planejada para facilitar e promover a realização de ações conjuntas.

3 Corpo de Conhecimento em Digitalização para o CEDAP

O primeiro passo para a elaboração do Corpo de Conhecimento em Digitalização do CEDAP compreendeu em estudar como são estruturados guias de Corpos de Conhecimento. Para o estudo, foram utilizados os guias PMBOK (PROJECT MANAGEMENT INSTITUTE, 2013), BPM CBOK (BENEDICT et ali, 2013) e SWEBOK (BOURQUE et al, 2014), por serem consolidados e envolverem temáticas relacionadas às atribuições do CEDAP e à digitalização, como gerenciamento de projetos, gerenciamento de processos e desenvolvimento de software, respectivamente.

Segundo ÖREN (2005, tradução nossa) Corpo de Conhecimento é (1) “conhecimento estruturado que é usado por membros de uma disciplina para guiar suas práticas e trabalhos” e (2) “agregação estabelecida de conhecimento em uma área particular que é esperada ser aprendida a fundo por um indivíduo para ser considerado ou certificado como um praticante”. Um Corpo de Conhecimento identifica áreas de conhecimento relevantes para guiar trabalhos ou práticas de uma disciplina, podendo incluir a especificação ciclos de vidas, fluxos e processos ligados a essas práticas.

O PMBOK (Project Management Body of Knowledge) é um guia para Corpo de Conhecimento em gerência 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 (quadro 1); define processos de gerenciamento para áreas de conhecimento; especifica as entradas, as saídas e as ferramentas e as técnicas para esses processos; e classifica esses processos em de iniciação, planejamento, execução, controle e encerramento. Por exemplo, na área de conhecimento “Gerenciamento do Escopo do Projeto”, PMBOK especifica o processo de “Coleta de Requisitos”, que é um processo de planejamento, e o processo de “Controle do Escopo”, que é de controle. Para o processo de “Coleta de Requisitos”, PMBOK estabelece, como entradas, o termo de abertura do projeto e o registro das partes interessadas (stakeholders); como saídas, a documentação dos requisitos e o plano de gerenciamento de requisitos; e apresenta, como ferramentas e técnicas para execução do processo, entrevistas, dinâmicas de grupos, oficinas, técnicas de criatividade em grupo, entre outras. PMBOK constitui-se em um vocabulário comum para gestão de projetos.

Quadro 1 - Áreas de Conhecimento de PMBOK e tipos dos processos

Áreas de Conhecimento	Tipos de Processos
1.Integração	1.Iniciação
2.Escopo	2.Planejamento
3.Tempo	3.Execução
4.Custo	4.Monitoramento e Controle
5.Qualidade	5.Encerramento
6.Recursos Humanos	
7.Comunicações	
8.Riscos	

Fonte: PROJECT MANAGEMENT INSTITUTE, 2013

BPM CBOK (Business Process Management Common Body of Knowledge) é um guia desenvolvido pela Association of Business Process Management, que estabelece um Corpo de Conhecimento em processos de negócio. BPM CBOK também está estruturado em áreas de conhecimento (quadro 2). 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 ali., 2013)

Quadro 2 – Áreas de Conhecimento de BPM CBOK

Áreas de Conhecimento de BPM CBOK
1. Gerenciamento de Processos de Negócio
2. Modelagem de Processos
3. Análise de Processos
4. Desenho de Processos
5. Gerenciamento de Desempenho de Processos
6. Transformação de Processos
7. Organização do Gerenciamento de Processos
8. Gerenciamento Corporativo de Processos
9. Tecnologias de BPM

Fonte: BENEDICT et al, 2013

SWEBOK (Software Engineering Body of Knowledge) estabelece um Corpo de Conhecimento para Engenharia de Software, que foi desenvolvido por vários profissionais da área e publicado pela IEEE Computer Society. SWEBOK é composto por áreas de conhecimento em Engenharia de Software e por áreas de conhecimento relacionadas à Engenharia de Software (quadro 3). Em SWEBOK, cada área de conhecimento é introduzida e brevemente definida, e uma visão geral sobre seu escopo e suas relações com outras áreas são apresentadas. A base da descrição de cada área é sua decomposição em tópicos e subtópicos. Para cada tópico/subtópico, é dada uma breve descrição e são apresentadas uma ou mais referências (BOURQUE et al, 2014). SWEBOK também apresenta, em anexo, padrões da IEEE e ISO/IEC que dão suporte à Engenharia de Software.

PMBOK, BPM CBOK e SWEBOK são Corpos de Conhecimento desenvolvidos para guiar profissionais em trabalhos e no exercício de práticas das respectivas áreas. Foram criados por especialistas nessas áreas, em processos promovidos por associações de profissionais. Resultam em publicações de guias, com processo evolutivo focado pela edição de novas versões.

Já o Corpo de Conhecimento em Digitalização do CEDAP (CEDAP DBOK) possui uma natureza diferente. Os guias resultam da reunião, sistematização e organização do conhecimento trazido por profissionais experientes em trabalhos e práticas, enquanto que o desenvolvimento CEDAP DBOK envolve a aquisição, o registro e a organização do conhecimento, que são desenvolvidas a partir das práticas e dos estudos realizados por membros do CEDAP ao longo dos projetos de digitalização.

Quadro 3 – Áreas de Conhecimento de SWEBOK

Áreas de Conhecimento	Áreas de Conhecimento Relacionadas
Requisitos de Software Projeto Técnico de Software Construção de Software Teste de Software Manutenção de Software Gestão de Configuração de Software Gestão de Engenharia de Software Processo de Engenharia de Software Ferramentas e métodos de Engenharia de Software Qualidade de Software	Engenharia de Computação Ciência da Computação Gestão Matemática Gestão da Qualidade Ergonomia de Software Engenharia de Sistemas

Fonte: BOURQUE et al, 2014

A estratégia de publicação de guias (e suas versões) não é a mais adequada para o CEDAP DBOK. O desenvolvimento do CEDAP DBOK envolve um processo de construção coletiva, em que novos conhecimentos são trazidos por membros, socializados, experimentados, discutidos e aprovados pela equipe, com o registro desse processo e de seus resultados. Esse registro deve ser formalmente especificado, permitindo sua recuperação baseada em significado e apoiada pelo computador. Em função disso, o desenvolvimento do CEDAP DBOK deve ser apoiado por ambientes semânticos e colaborativos. A web semântica apresenta uma infraestrutura para dar suporte semântico à representação do conhecimento, e na Web 2.0 temos técnicas adequadas para promover a construção coletiva.

O desenvolvimento do CEDAP DBOK envolve aspectos de organização e registro do corpo de conhecimento, assim como a implantação de um ambiente que

apoia a construção coletiva desse corpo. Com relação a esses aspectos, o desenvolvimento do CEDAP DBOK observa a seguinte estratégia:

- Identificar as estruturas para a representação de Corpos de Conhecimento, analisando os guias PMBOK, BPM CBOK e SWEBOK, escolhidos por serem amplamente reconhecidos e envolverem temáticas próximas às práticas do CEDAP ;
- Identificar requisitos para a construção colaborativa do Corpo de Conhecimento, com base na infraestrutura da Web 2.0, caracterizada por O'REILLY (2007);
- Identificar requisitos para a representação e organização do conhecimento, com base na arquitetura da Web Semântica proposta por BERNERS-LEE, HENDLER, J e LASSILA (2001);
- Identificar, escolher e implantar uma ferramenta existente, de código aberto, para o registro e desenvolvimento colaborativo do CEDAP DBOK;
- Especificar, na ferramenta, a estrutura inicial para o registro do conhecimento, dando início ao processo colaborativo de desenvolvimento do CEDAP DBOK e da evolução da infraestrutura para o desenvolvimento do CEDAP DBOK

A primeira estratégia é analisada na seção 4, e as estratégias seguintes são abordadas na seção 5.

4 Organização do Corpo de Conhecimento em Digitalização

Analisando as estruturas dos Guias PMBOK, PMI CBOK e SWEBOK, foram observados aspectos relevantes, não necessariamente presentes em todos os guias, como:

- Contextualização do Corpo de Conhecimento;
- A estrutura básica do corpo focada nas Áreas de Conhecimento que compõem o corpo;
- Definição da Área de Conhecimento, seu escopo, referências e fontes de informação relacionadas;
- Áreas de Conhecimento desmembradas em tópicos e subtópicos;

- Processos, relacionados a Áreas, com a identificação de saídas, entradas, fluxos e documentação necessária;
- Relacionamentos com outras Áreas de Conhecimento;
- Relacionamento com padrões e normas;
- Glossários e definições

A partir da identificação desses aspectos, a estrutura do CEDAP DBOK foi estabelecida. Essa estrutura observa características especiais do CEDAP e da digitalização, que envolvem: um processo de construção coletiva; a existência de diretrizes, recomendações e padrões de digitalização; a necessidade em conhecer e operar recursos como equipamentos, softwares, bases de dados, formatos de arquivo; e a existência de iniciativas, grupos, organizações e projetos que visam desenvolver e promover a digitalização. A seguinte estrutura inicial foi estabelecida para o CEDAP DBOK:

- Áreas do Conhecimento em Digitalização;
- Áreas do Conhecimento relacionadas com a Digitalização;
- Conceitos ligados às Áreas de Conhecimento;
- Fontes de conhecimento para as Áreas;
- Diretrizes, normas padrões estabelecidas para Digitalização;
- Processos e fluxos associados às Áreas de Conhecimento;
- Recursos ligados à digitalização, como modelos de scanners, gerenciadores de bases de dados, softwares para captura, descrição
- Instituições, grupos, iniciativas e projetos ligados à digitalização.

Como o CEDAP DBOK é desenvolvido a partir da construção coletiva, essa estrutura pode mudar, adquirir novas formas. Analisando as principais publicações e guias em digitalização, as foram estabelecidas as seguintes Áreas de Conhecimento para CEDAP BOK, apresentadas no quadro 4.

Quadro 4 – Áreas de Conhecimento de DCAP DBOK

Áreas de Conhecimento	Áreas de Conhecimento Relacionadas
Seleção Conversão Controle de Qualidade Fluxo de Digitalização Metadados	Processamento Imagem Gerenciamento de Projetos Gerenciamento de Direitos Autorais Ciência da Informação Arquivologia

Preservação Digital Armazenamento e Acesso Planejamento da Digitalização	Museologia
--	------------

Fonte: CDAP DBOK

5 Infraestrutura para Representação e Construção do Corpo de Conhecimento

Para o desenvolvimento do CDAP CBOK, foram estabelecidos os seguintes requisitos para construção colaborativa e a representação e organização do conhecimento:

- Permitir a construção colaborativa, focada em um modelo 3C (FUCKS, 2007), que desenvolve a Cooperação, a Coordenação e a Comunicação, permitindo: a produção cooperativa de textos e descrições semânticas; a mediação/controlar do material produzido com registro das mudanças (histórico do que foi desenvolvido, com identificação de alterações); a discussão, através de fóruns de trocas de mensagens;
- Permitir o registro do conhecimento em forma textual, assim como sua descrição semântica, compreensível por máquina, compatível com a infraestrutura da Web Semântica (uso de ontologias, e representação de descrições em RDF)

Em função desses requisitos, foram buscadas e analisadas ferramentas que combinam técnicas e estruturas a Web 2.0 e da Web Semântica, isto é que dão suporte à construção colaborativa (Web 2.0) e à representação formal e compartilhada do conhecimento (Web Semântica). Foram analisadas as ferramentas Semantic Media Wiki (VÖLKEL et ali, 2006), OntoWiki (AUER; DIETZOLD e RIECHERT,2006) e Fabrico/Ciência (ROCHA, 2014).

Semantic Media Wiki (SMW) é uma extensão da ferramenta MediaWiki (MW). MW é a ferramenta usada pela Wikipédia para a construção coletiva dos seus verbetes. Permite a edição colaborativa de textos, registra o histórico de cada texto (identificando seus autores e contribuições), possibilita o estabelecimento de mecanismos de moderação e de comunicação através de fóruns de discussão. É uma ferramenta amplamente utilizada, e várias extensões foram desenvolvidas e estão disponibilizadas no site de MW.

SMW é uma extensão de MW que incorpora ao MW recursos semânticos, de acordo com a arquitetura da Web Semântica. Permite a adição de semântica aos textos. No SMW, podemos definir, através de uma ontologia, classes (como Área de Conhecimento, Conceito, Modelo de Scanner, Processo, Organização) e propriedades (como o nome de uma organização, a resolução de um modelo de scanner, os processos de uma área de conhecimento).

Os textos wiki são tipificados de acordo com as classes definidas na ontologia. Por exemplo, no contexto do CDAP DBOK, o texto wiki “Seleção” seria do tipo “Área de Conhecimento”, pois representaria uma área de conhecimento; o texto wiki “Scanner OS1002” seria do tipo Modelo de Scanner, pois representaria um modelo de scanner, e o texto Wiki “Configuração do Representante Digital” representaria um processo.

No SMW, propriedades permitem que sejam atribuídas características a um recurso, assim como relacionar recursos. As propriedades de um recurso são definidas no próprio texto wiki do recurso. Por exemplo, no texto wiki do processo “Configuração do Representante Digital”, poderia haver um link para o texto da Área de Conhecimento “Conversão”, relacionando o processo com sua área. Esse link poderia ser especificado no SMW como a propriedade “área do processo”. Como os textos wiki e os links são construídos a partir de ontologias, temos um wiki semântico.

OntoWiki (AUER, DIETZOLD e RIECHERT, 2006) é uma ferramenta para autoria, apresentação e gestão de bases de conhecimento, inspirada nas facilidades oferecidas pelos ambientes wiki para desenvolver o trabalho colaborativo. Em OntoWiki, usuários produzem colaborativamente sentenças, isto é, descrições semânticas na linguagem RDF e de acordo com ontologias. Segundo Heino (2009), 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. Os conteúdos da base de conhecimento são editados e navegados através de telas que são geradas automaticamente a partir da ontologia e das descrições semânticas.

Fabrico/Ciência é um anotador baseado em ontologias construído de acordo com a arquitetura da Web Semântica e seguindo o modelo de negócio de sistemas da Web 2.0 de O’Reilly (2007).

[...] utiliza técnicas da Web 2.0 para capturar contribuições dos usuários (via anotações semânticas) e para auxiliar a exploração dos dados. Incorpora um ambiente wiki, um fórum de discussão, funcionalidades de agregação típicas de folksonomias, e um mecanismo que informa as últimas contribuições e que relaciona usuários que realizaram contribuições envolvendo mesmos recursos. Fórum e wikis podem ser criados para discutir e documentar qualquer recurso descrito, inclusive sentenças (via reificação), ou classes (no ambiente de construção de ontologias) (ROCHA, 2014)

SMW, OntoWiki e Fabrico/Ciência são ambientes que operam na plataforma da Web Semântica (via ontologias e descrições em RDF). Para fomentar a construção coletiva, oferecem recursos como fóruns e registram o histórico/mudanças. Enquanto OntoWiki e Fabrico/Ciência têm as sentenças semânticas (sentenças RDF) como elementos base, em SMW os textos wiki compõem a base, que passaram a representar conceitos semânticos, como classes, propriedades e instâncias. SMW possui recursos mais restritos para representar ontologias e descrições que OntoWiki e Fabrico/Ciência, entretanto, possui mais facilidades para a construção de textos.

SMW foi escolhida para o CDAP DBOK, e o motivo principal dessa escolha foi a sua natureza básica ser focada na construção coletiva de textos. Com base na estrutura para o CEDAP DBOK estabelecida na seção 4, a ferramenta SWM foi configurada da seguinte forma:

- Uma ontologia foi desenvolvida para representar os tipos de recursos da estrutura do CDAP DBOK (identificados na seção 4). As principais classes dessa ontologia são: Áreas do Conhecimento, Conceitos, Processos, Organizações, Projetos, Modelos de Scanner, Software, Formato de Arquivo, Esquema de Metadados. As classes da Ontologia também são páginas Wiki, e atributos e relacionamentos entre recursos são estabelecidos através de propriedades do SMW.
- Páginas foram criadas para representar os recursos especificados na ontologia, isto é, uma página foi criada para cada área de conhecimento (quadro 4), conceito, processo, organização, etc. Nos textos dessas páginas incluem propriedades especificadas na ontologia.
- Os textos e a ontologia são desenvolvidos coletivamente, com apoio de recursos do MW como fórum de discussão, registro do histórico e moderação.

6 Considerações Finais

A estratégia de buscar nos Corpos de Conhecimento subsídios para organizar o conhecimento em digitalização do CEDAP resulta em uma base de conhecimento com estruturas voltas não somente capturar o conhecimento, mas também apoiar o conhecimento aplicado a práticas e trabalhos.

Normalmente, Corpos de Conhecimento são desenvolvidos por especialistas para auxiliar profissionais em suas práticas e trabalhos, e são representados através de guias que evoluem através de novas versões. O CEDAP CBOK possui uma natureza diferente. O corpo de conhecimento do CEDAP é usado para auxiliar os membros do CEDAP no exercício de suas práticas em digitalização, e é construído pelos membros do CDAP à medida que conhecimentos e experiências são adquiridos. Em função disso, a estratégia utilizada para desenvolver esse Corpo de Conhecimento é a construção colaborativa de textos e a descrição semântica do conteúdo. Por ser focada na elaboração coletiva de textos e no uso de ontologias para dar significado a estes textos, a ferramenta SMW foi selecionada.

O CEDAP e seu Corpo de Conhecimento estão em fase de implantação e desenvolvimento. A ontologia e os textos principais (Áreas de Conhecimento, Conceitos, Modelos de Scanners, Processos, Formatos, etc) foram criados e estão sendo alimentados ao longo das experiências adquiridas. Por exemplo, experiências no uso dos scanners e de seus softwares de captura estão sendo registradas nos seus respectivos textos, assim como conhecimentos exigidos em função desses experimentos.

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.
- 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
- BERNERS-LEE, T.; HENDLER, J.; LASSILA O. The Semantic Web. **Scientific American**, may 2001.
- BOURQUE, Pierre et al. **Guide to the Software Engineering Body of Knowledge, Version 3.0.** IEEE Computer Society Press, 2014.

- EUROPEAN COMMISSION. **Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020**. Version 16. European Commission, December 2013.
- 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
- HEINO, N. et al. Developing semantic web applications with the ontowiki framework. In. **Networked Knowledge-Networked Media**. Berlin Heidelberg:Springer, 2009.
- LORD, P. et al. **E-Science curation report, Data curation for e-Science in the UK: An audit to establish requirements for future curation and provision**. Digital Archiving Consultancy Limited, 2003.
- NATIONAL SCIENCE FOUNDATION. Other Post Award Requirements and Considerations. In **Proposal and Award Policies and Procedures Guide**. National Science Foundation, jan. 2013.
- 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.
- PROJECT MANAGEMENT INSTITUTE. **A Guide to the Project Management Body of Knowledge, PMBOK Guide**. Project Management Institute, 2013.
- ROCHA. R. FABRICO/CIÊNCIA no Desenvolvimento de Ambientes Linked Data para a Ciência. In.: Linded Open Data Brasil. 2014, Florianópolis. **Anais...** Nov. 2014.
- 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