



Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação

ISSN: 1518-2924

adilson.pinto@ufsc.br

Universidade Federal de Santa Catarina  
Brasil

Port da ROCHA, Rafael; CAREGNATO, Sonia; GABRIEL JUNIOR, Rene Faustino  
Aspectos de inovação na implantação de um centro de digitalização e gestão de dados da pesquisa  
Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, vol. 23, núm. Esp.1, 2018, -Junho, pp. 1-15  
Universidade Federal de Santa Catarina  
Brasil

DOI: <https://doi.org/10.5007/1518-2924.2018v23nespp1>

Disponível em: <https://www.redalyc.org/articulo.oa?id=14762635001>

- Como citar este artigo
- Número completo
- Mais informações do artigo
- Site da revista em [redalyc.org](http://redalyc.org)

UABEM [redalyc.org](http://redalyc.org)

Sistema de Informação Científica Redalyc  
Rede de Revistas Científicas da América Latina e do Caribe, Espanha e Portugal  
Sem fins lucrativos acadêmica projeto, desenvolvido no âmbito da iniciativa  
acesso aberto

ARTIGO

Recebido em:  
28/04/2017

Aceito em:  
09/04/2018

*Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 23, n. esp., p. 1-15, 2018.  
ISSN 1518-2924. DOI: 10.5007/1518-2924.2018v23nespp1

## Aspectos de inovação na implantação de um centro de digitalização e gestão de dados da pesquisa

*Innovation aspects for the implementation of a digitizing and research data management center*

Rafael Port da ROCHA ([rafael.rocha@ufrgs.br](mailto:rafael.rocha@ufrgs.br))\*

Sonia CAREGNATO ([sonia.caregnato@ufrgs.br](mailto:sonia.caregnato@ufrgs.br))\*

Rene Faustino GABRIEL JUNIOR ([rene.gabriel@ufrgs.br](mailto:rene.gabriel@ufrgs.br))\*

\* Professores Doutores do Departamento de Ciência da Informação da UFRGS.

### Resumo

Objetiva prover serviços de curadoria e gestão de dados da pesquisa com inovação em P&D a um centro de digitalização e gestão de dados da pesquisa. Caracteriza o ecossistema de P&D em gestão e curadoria de dados da pesquisa, e observa os tipos de processos de inovação que casam com características desse ecossistema. Propõe e cria um ambiente de inovação para o centro, desenvolvendo uma ferramenta fundamentada em trabalho colaborativo, corpos de conhecimento e processos de inovação. Apresenta resultados obtidos a partir do uso do ambiente.

**Palavras-chave:** Inovação. Dados da Pesquisa. Curadoria Digital. Colaboração. Ciência da Informação.

### Abstract

It aims to provide research data curation and management services to a digitization and research data management centre, with innovation in R&D. It characterizes the ecosystem that involves R&D in research data curation and management, and observes types of innovation processes that match with characteristics of this ecosystem. It proposes and creates an innovation environment for the centre, developing a tool founded on collaborative work, bodies of knowledge and innovations processes. It presents results obtained from the use of the environment.

**Keywords:** Innovation. Research Data. Digital Curation. Collaboration. Information Science.



v. 23, n. esp., 2018  
p. 1-15  
ISSN 1518-2924



Esta obra está licenciada sob uma [Licença Creative Commons](https://creativecommons.org/licenses/by/4.0/).

## 1 INTRODUÇÃO

O cenário atual da pesquisa demanda das universidades e de centros de pesquisa a oferta de novos serviços de apoio ao pesquisador, como os serviços de dados da pesquisa, que auxiliam pesquisadores no gerenciamento dos dados de suas pesquisas, e promovem a preservação e o compartilhamento desses dados. Segundo Lord e Macdonald (2003), a preservação e o compartilhamento de dados da pesquisa trazem benefícios como:

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, muitos pesquisadores, além de apresentarem os resultados de suas pesquisas através de publicações científicas, desenvolvem planos para gerir os dados dessas pesquisas e publicam esses dados em repositórios. Essas práticas foram impulsionadas por políticas adotadas por instituições de fomento à pesquisa e por periódicos científicos, que passam a exigir ou recomendar a inclusão, em suas pesquisas, do planejamento da gestão de dados e a publicação dos dados em repositório.

O sucesso da implantação e da operação de serviços de dados da pesquisa em uma instituição está associado a uma estratégia que considera Pesquisa e Desenvolvimento (P&D), e Inovação. P&D compreende nas atividades a serem desenvolvidas para criar produtos ou serviços, ou para incrementar produtos ou serviços existentes. Isso implica em empreendimentos que consideram: equipes de engenharia ou de cientistas aplicados, não depender somente do retorno a curto prazo, e assumir riscos e incertezas. A inovação envolve introduzir algo novo ou modificar substancialmente o existente. Segundo a Lei da Inovação (BRASIL, 2004), alterada pela Lei nº. 13.243, inovação é

A introdução de novidade ou aperfeiçoamento no ambiente produtivo e social que resulte em novos produtos, serviços ou processos ou que compreenda a agregação de novas funcionalidades ou características a produto, serviço ou processo já existente que possa resultar em melhorias e em efetivo ganho de qualidade ou desempenho (BRASIL, 2016).

P&D e Inovação são aspectos importantes a serem considerados para o sucesso da oferta de serviços de dados da pesquisa, visto que os temas envolvidos são recentes, em que a busca e a experimentação de soluções ocorrem, em grande parte, no âmbito da pesquisa. Preservação digital, dados da pesquisa e digitalização focada na preservação são exemplos de temas com essas características.

Muitos resultados de pesquisa em digitalização, preservação digital e em dados de pesquisa ainda não foram largamente experimentadas na oferta de serviços de dados da pesquisa completos e integrados, no âmbito de uma Instituição. Como esses resultados normalmente estão direcionados a partes específicas do problema, um desafio está em compô-los, na oferta de serviços mais abrangentes. Por exemplo, resultados de pesquisas em planos de gestão de dados, como os de Sallans e Donnelly (2012), e em planos de preservação digital, como os de Beker (2009), poderiam ser combinados nas ofertas de serviços de planejamento da gestão de dados da pesquisa integrados com o planejamento da preservação.

O desenvolvimento da pesquisa atual ocorre de forma muito diversificada, dispondo de uma grande variedade de técnicas para coleta, análise, representação e armazenamento de dados, e ocorrendo, em muitos casos, de forma interinstitucional e colaborativa. Muitas pesquisas inovam nos métodos e nas técnicas para obter, analisar, representar e armazenar seus resultados. Por exemplo, em pesquisas na área biomédica, recursos como ontologias tornam-se necessários para organizar e

produzir anotações colaborativas sobre genes (BLAKE, 2013). Esse cenário da pesquisa demanda por uma estratégia para desenvolver serviços de dados da pesquisa fortemente conectada com os pesquisadores e suas novas estratégias de pesquisas, buscando, junto a estes, novos serviços ou modificações em serviços atuais.

A implantação de serviços de dados da pesquisa por uma instituição requer por uma equipe que tenha capacidade dominar de forma continuada o estado arte em dados da pesquisa e em temas relacionados. Para ter esse domínio, é necessário que a equipe possua um entendimento comum sobre os conceitos desenvolvidos, e disponha de instrumentos para organizar e registrar os conhecimentos e as práticas adquiridas, tanto como decorrência de estudos do estado da arte, quanto de experimentações e pesquisas próprias. Também demanda de uma profunda interação com os pesquisadores no exercício de suas atividades de pesquisa, para identificar conjuntamente demandas e propor novas soluções.

O Centro de Documentação e Acervo Digital da Pesquisa (CEDAP) é um Órgão Auxiliar da Faculdade de Biblioteconomia e Comunicação (FABICO/UFRGS) e tem por finalidades dar suporte à pesquisa científica e tecnológica através da digitalização, gestão e curadoria de ativos digitais de pesquisa, assim como avançar o estado da arte em digitalização e curadoria de ativos digitais, estimulando a produção científica e o compartilhamento e a disseminação de novos conhecimentos. Os recursos para a construção da plataforma de digitalização foram disponibilizados pela FINEP, através do Edital CT-INFRA 2010.

Esse artigo discute a estratégia adotada pelo CEDAP para prover serviços de dados de pesquisa com P&D e inovação. O artigo está organizado da seguinte forma. A próxima seção apresenta, define e contextualiza serviços de dados da pesquisa. A seção seguinte apresenta os objetivos e a estratégia adotada para desenvolver serviços de dados da pesquisa no CEDAP. As seções posteriores apresentam e discutem as etapas da estratégia, que envolvem: identificar um ecossistema de P&D em dados de pesquisa, preservação digital e digitalização; identificar tipos/gerações de processos de inovação característicos para esse ecossistema; e propor um ambiente de funcionamento do CEDAP em que os serviços de dados da pesquisa são desenvolvidos com P&D e inovação, considerando o ecossistema identificado. Por último, são discutidos benefícios já obtidos pelo CEDAP, através da utilização desse ambiente.

## 2 SERVIÇOS DE DADOS DA PESQUISA

Serviços de dados da pesquisa ocorrem ao longo de todo o ciclo de vida do dado da pesquisa. Segundo o reconhecido serviço dados da pesquisa, *UK Data Service* (2017), o ciclo de vida do dado da pesquisa envolve as seguintes fases: **planejar dados** (desenhar a pesquisa; planejar a gestão de dados, o consentimento de compartilhamento, o processamento e os protocolos de coleta; explorar fontes de dados existentes), **coletar dados** (coletar dados; capturar dados com metadados; obter dados de terceiros); **processar e analisar dados** (entrar, digitalizar, transcrever e traduzir, checar, validar, limpar anonimizar, derivar e documentar dados; gerenciar e armazenar dados; analisar e interpretar dados; produzir saídas da pesquisa; citar fontes de dados); **publicar e compartilhar dados** (estabelecer copyright; criar documentação de usuário; criar metadados de descoberta; acesso apropriado aos dados; publicar/compartilhar dados; promover dados); **preservar dados** (migrar dados; armazenar e criar cópia de segurança; criar a documentação de preservação; preservar e curar dados) e **Reusar Dados** (conduzir análises secundárias, acompanhamentos, revisões da pesquisa; examinar os achados; usar dados para ensino e aprendizado).

Segundo Carlson (2014), serviços de dados da pesquisa podem ser serviços de gestão de dados da pesquisa e serviços de curadoria de dados da pesquisa. A gestão de dados da pesquisa compreende na

[...] organização dos dados, desde a sua entrada no ciclo de pesquisa, até a disseminação e arquivamento dos seus valiosos resultados. Visa garantir uma verificação fiável dos resultados e permitir novas e inovadoras pesquisas baseadas em informações existentes (WHYTE; TEDDS, 2011).

Os serviços de gestão de dados da pesquisa ocorrem durante os estágios ativos do ciclo de vida dos dados, quando os pesquisadores estão gerando e usando os dados em suas pesquisas, e incluem o desenvolvimento de planos de gestão de dados, treinamentos, documentação e organização dados, e desenvolvimento ferramentas ou recursos para armazenar os dados (CARLSON, 2014).

Serviços de curadoria de dados da pesquisa ocorrem tanto no momento em que os dados são geridos pela pesquisa original, quanto na fase de curadoria, em que os dados passam a estar amplamente disponíveis para outras pessoas, sendo muitas vezes administrados por terceiros, como um gestor de repositório (CARLSON, 2014).

Serviços de dados da pesquisa envolvem digitalização e preservação digital. Tendo como base o ciclo de vida do dado da pesquisa (UK DATA SERVICE, 2017), a digitalização está associada à fase de coleta. O Quadro 1 apresenta um paralelo entre o ciclo de vida da curadoria digital de Higgins (2008) e o ciclo de vida do dado da pesquisa de *UK Data Service* (2017), destacando fases que atuam na curadoria e fases que atuam na preservação. Analisando o quadro, observamos que as fases de curadoria do ciclo de vida do dado da pesquisa compreendem, basicamente, em especializações das etapas equivalentes do ciclo de vida da curadoria digital, com o enfoque sendo curadoria de dados.

**Quadro 1:** Paralelo entre ciclos de dado da pesquisa e da curadoria digital

Ciclo de Vida do Dado da Pesquisa	Ciclo de Vida da Curadoria Digital	Fases
Planejando Dados	Conceitualização	Curadoria Digital
Coletando Dados	Criação ou Recebimento	
Processando e Analisando Dados	Avaliação e Seleção	
Preservando dados	Ingestão	Preservação Digital
	Ação de Preservação	
	Armazenamento	
Publicando e Compartilhando Dados	Acesso, Uso e Reuso	Curadoria Digital
Reusando dados	Transformação	

**Fontes:** Adaptado de Higgins (2008) e UK Data Service (2017).

### 3 ESTRATÉGIA PARA PROVER SERVIÇOS DE DADOS DA PESQUISA NO CEDAP

Para desenvolver de serviços de dados da pesquisa com P&D e inovação, o CEDAP busca atuar combinando a aquisição e a produção de conhecimento, com a aquisição e a troca de experiências com os beneficiários desses serviços. Os seguintes objetivos foram traçados para o desenvolvimento dos serviços do CEDAP:

- O acompanhamento e o domínio, por parte da equipe do CEDAP, do estado da arte da pesquisa e das práticas em dados da pesquisa, e em assuntos relacionados, como digitalização e preservação digital.
- O desenvolvimento e a oferta de serviços de dados de pesquisa com base em observações e trocas realizadas com pesquisas em andamento e na aplicação de conhecimentos adquiridos.
- A aquisição e compartilhamento do conhecimento e das práticas de forma compartilhada e colaborativa.
- O registro e organização do conhecimento e das práticas em estruturas conceitualmente bem formadas, a fim de proporcionar uma convergência conceitual na atuação da equipe e um entendimento comum sobre os temas.

Para atender a esses objetivos, a seguinte estratégia foi estabelecida:

- Caracterizar o ecossistema de P&D em dados da pesquisa, digitalização e preservação digital, com base na literatura e em serviços de dados da pesquisa disponíveis na web.
- Investigar processos de inovação existentes, com base na literatura, e identificar aqueles que mais caracterizam o ecossistema identificado.
- Criar um ambiente para desenvolver e operar serviços de dados da pesquisa no CEDAP, com foco nos processos de inovação e ecossistema identificados.

A criação desse ambiente teve como estratégia:

- A identificação dos atores envolvidos.
- A busca de um modelo de colaboração tendo como base os atores, o ecossistema e os processos de inovação identificados.
- A definição das estruturas para organizar e representar o conhecimento e as práticas.
- A busca de uma ferramenta de colaboração que atenda ao modelo de colaboração e às estruturas para organizar e representar o conhecimento.

#### **4 ECOSISTEMA DE P&D EM SERVIÇOS DE DADOS DA PESQUISA**

Com base na literatura e na análise de serviços e recursos presentes na web, identificou-se que a P&D em dados da pesquisa ocorre em um ecossistema que envolve uma rede aberta de pessoas e entidades que cooperam e trocam experiências. Também envolve digitalização e preservação digital, pois a digitalização é uma forma de processamento de dados, e o ciclo de vida do dado pode ser visto como uma espécie de ciclo de vida da curadoria digital em que os objetos sob curadoria são dados da pesquisa (Quadro 1).

O ecossistema identificado é composto por atores que são: pesquisadores ligados a instituições de pesquisa, que disseminam e validam seus resultados através de canais de publicação científica; órgãos de instituições de ensino e pesquisa que desenvolvem e oferecem serviços de digitalização, preservação digital (como DPOC) e dados da pesquisa (como *UK Data Service*); instituições nacionais e regionais que promovem o compartilhamento de dados em suas jurisdições (como DANS, da Holanda, e ANDS, da Austrália); associações e alianças compostas por pesquisadores e instituições que focam o desenvolvimento e o compartilhamento de conhecimento em dados da pesquisa (como RDA) e na preservação digital (como *Digital Preservation Coalition*); e órgãos que promovem padrões, recomendações, modelos de referência e guias (como *DDI Alliance* e DSA). Além disso, participam desse ecossistema, instituições governamentais de apoio à pesquisa (como a *National Science Foundation*, dos Estados Unidos, e Horizonte 2020, da Europa), fortalecendo o ecossistema através do estabelecimento de políticas para abertura e publicação de dados da pesquisa.

O ecossistema opera em um mundo aberto, com o enfoque na livre troca do conhecimento. Usa padrões abertos com o intuito de disseminar conhecimentos e promover a interoperabilidade entre os ambientes criados. Explora padrões, modelos de referência, recomendações e guias abertos para seleção de material para digitalização; configurações para representações digitais de objetos digitalizados; pacotes para representar objetos digitais; metadados descritivos, estruturais e administrativos; identificação persistente objeto recursos na *web*; citação; interoperabilidade de informação; ciclos de vida; entre outros.

O ecossistema também é rico na produção, uso e reuso de *software* livre. *Softwares* livres são usados para verificar e validar formatos e arquivos, para apoiar o desenvolvimento de planos de gestão de dados e de preservação digital, para repositórios de dados, e para apoiar pesquisadores nas fases do ciclo de vida. O ecossistema identificado possui as seguintes características:

- Inter-relaciona práticas e pesquisas em dados de pesquisa, digitalização e preservação digital.
- É composto por diversos tipos de atores engajados em desenvolver e promover pesquisa e serviços, em um ambiente de cooperação, aberto, com a formação de alianças, associações e centros de referência.
- Está voltado a promover e disseminar os resultados em um ambiente aberto e de livre acesso.
- Busca a padronização, através do desenvolvimento e uso de modelos de referência, normas, certificações e guias abertos.
- Está focado na produção, uso e combinação de *software* livre, para repositórios, planos de gestão de dados, ciclos de vida de dados da pesquisa, etc.

## **5 PROCESSOS INOVAÇÃO EM DADOS DA PESQUISA NO CEDAP**

O processo de inovação, ao longo de sua história, passou por vários modelos. Rotwell (1995) identificou que o processo de inovação ao longo dos tempos se desenvolveu em cinco gerações. Na primeira geração, o desenvolvimento tecnológico era o principal promotor da inovação, sendo o mercado um mero receptor, em um modelo sequencial, partindo da ciência básica, passando pela engenharia e pela manufatura, e chegando ao marketing e às vendas. A inovação era empurrada pela tecnologia. Na segunda geração, o processo de inovação segue um fluxo inverso, em que as necessidades do mercado são os fatores que movem a inovação. Nesse caso, a necessidade do mercado leva o desenvolvimento tecnológico, industrialização e vendas, em um fluxo sequencial. A inovação era puxada pelo mercado. Na terceira geração ocorre uma acomodação entre a pesquisa e o desenvolvimento, demandado pelas necessidades de racionalização. Nela, o processo de inovação representa a confluência entre as capacidades tecnológicas e as necessidades do mercado, em um fluxo sequencial bidirecional, inserido em uma rede complexa de comunicação, que liga a empresa à comunidade científica e tecnológica e a outras empresas do mercado.

Na quarta geração, o processo de inovação desenvolve-se em um modelo em que o foco é a integração e o desenvolvimento em paralelo. Nessa geração, diferentes departamentos trabalham no projeto simultaneamente, mas integrados, buscando rapidez na inovação, em função da mudança tecnológica ocorrer em grande intensidade e dos produtos terem ciclos de vida curtos. A quinta geração decorre do aperfeiçoamento da quarta geração, com o desenvolvimento do modelo de trabalho em rede, pela integração de sistemas e uso de sistemas de informação e de projetos. Essa geração está focada na velocidade, na eficiência e na flexibilidade no desenvolvimento de seus produtos.

Inovação fechada é aquela adotada por empresas que obtém vantagens de mercado ao manterem internamente complexos laboratórios e infraestrutura de

P&D. Chesbrough (2006) introduz o processo de inovação aberta, que se opõe à inovação fechada. Segundo Chesbrough (2006), o processo de inovação aberta:

- Assume que as empresas podem e devem usar tantas ideias externas, quanto internas, visando avançar em sua tecnologia.
- Utiliza modelos de negócios para definir os requisitos para arquiteturas e sistemas que se valem de ideias internas e externas.
- Trata P&D como um sistema aberto, assumindo que ideias podem vir (e ir para o mercado), tanto de dentro como de fora da empresa.
- Ocorre muitas vezes como metodologias abertas de desenvolvimento de *software*, incorporando um modelo de negócio focado tanto no valor da criação (como no *software* aberto), quanto no valor da captura.
- Assume que o conhecimento útil é amplamente distribuído, e que mesmo as organizações de P&D mais capacitadas devem identificar, conectar e alavancar as fontes de conhecimento externas como um processo central na inovação.

No cenário atual, outro modelo de inovação bastante discutido é a inovação interativa. Segundo Ganzer et al. (2014), o processo de inovação interativa:

- É composto de *feedbacks* entre os diversos estágios do desenvolvimento da inovação, entre os polos científico, tecnológico e de mercado.
- Observa que a inovação não é resultado de um processo linear que se inicia com a pesquisa básica, passa pela pesquisa aplicada e termina com o desenvolvimento de um novo produto ou processo que é ofertado ao mercado.
- Observa que o processo inovativo não é determinista e não segue uma fórmula pronta; ele é socialmente construído pelos atores envolvidos ou interessados na geração da inovação.
- Observa que a inovação não é consequência de desenvolvimentos da ciência e da tecnologia exógenos ao sistema econômico e social, como coloca a abordagem empurrada pela tecnologia, mas tampouco é resultado exclusivamente de uma demanda revelada, como defende a teoria puxada pelo mercado.

Considerando o ecossistema identificado, observamos que o desenvolvimento dos serviços de dados da pesquisa do CEDAP passa por um processo de inovação que adquire características de inovação aberta e de inovação interativa. Se desenvolvidos considerando essas características, esses serviços introduzirão novidades ou aperfeiçoamentos no ambiente produtivo e social, resultando em melhorias e efetivo ganho de qualidade ou desempenho, isto é, trarão inovação.

A inovação no CEDAP adquire característica de inovação aberta, pois o ecossistema identificado atua em contexto do acesso livre, com as trocas de ideias ocorrendo em canais abertos de comunicação, sendo promovido por alianças e associações, e enfatizando desenvolvimento de padrões, guias e recomendações abertos. Nesse ecossistema, o desenvolvimento e o uso de soluções informatizadas ocorrem na forma de *software* livre.

A inovação no CEDAP adquire características de inovação interativa, pois o ecossistema identificado é composto por uma grande variedade de atores engajados em socialmente construir a inovação. Também ocorrem no CEDAP *feedbacks* entre nos estágios de desenvolvimento da inovação, como na estratégia da equipe do



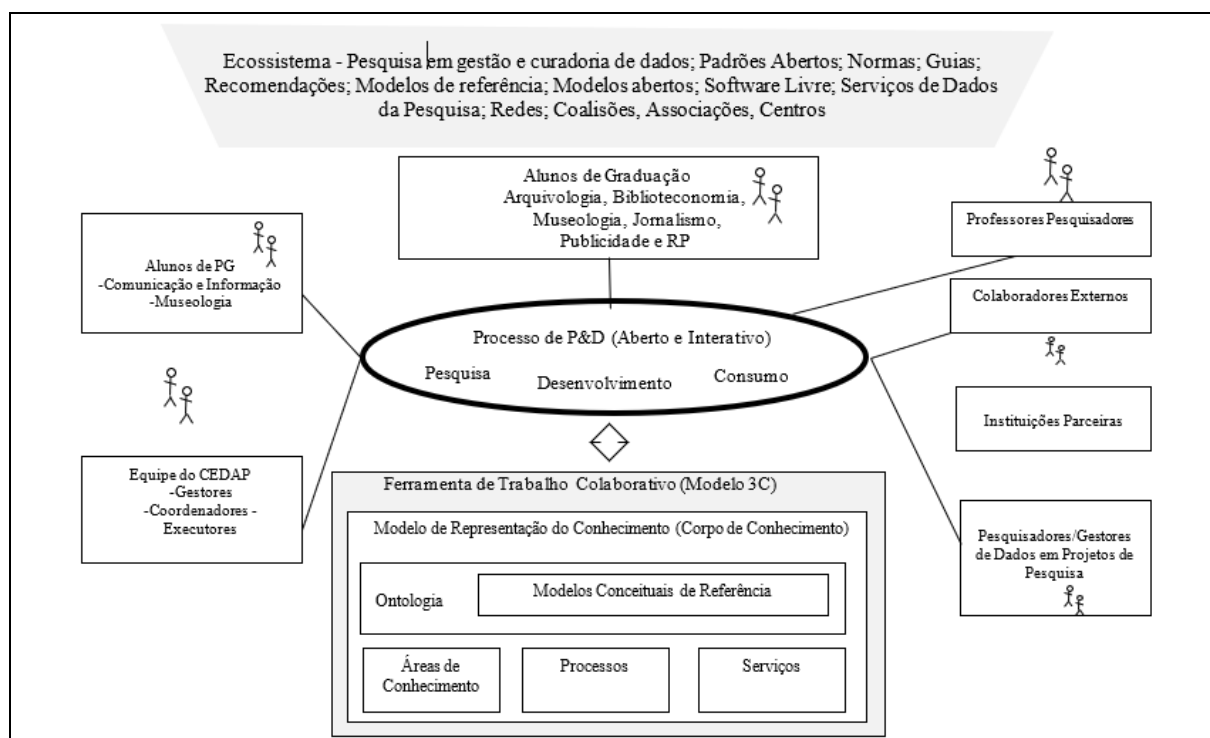
CEDAP de acompanhar o desenvolvimento de pesquisas em andamento, na busca conjunta de soluções e serviços inovadores.

## 6 AMBIENTE DE INOVAÇÃO EM DADOS DA PESQUISA NO CEDAP

Para apoiar um processo de inovação focado na inovação aberta e interativa, e no contexto do ecossistema identificado, um ambiente colaborativo foi desenvolvido para o CEDAP. Esse ambiente permite capturar, sistematizar, gerar e disseminar conhecimentos e práticas, assim como formalizar e institucionalizar os processos, os serviços e os produtos em digitalização, preservação digital e em dados da pesquisa. Também permite descrever elementos externos e internos do ecossistema, assim como o registrar conhecimentos capturado sobre eles. Os fundamentos básicos desse ambiente estão baseados em dois modelos: um modelo de colaboração e um modelo de representação do conhecimento.

Para chegar ao ambiente, as seguintes estratégias foram adotadas: (a) identificação dos atores envolvidos; (b) definição das estruturas do modelo de representação do conhecimento e das práticas, (c) busca de um modelo de colaboração apoiado pelo computador, (d) busca de uma ferramenta de colaboração que atenda aos modelos de colaboração e de representação do conhecimento especificados.

A Figura 1 apresenta o ambiente do CDAP. Os atores interagem de forma não linear e não determinística, colaborando tanto na pesquisa, como no desenvolvimento do produto/serviço. São atores, a equipe interna do CEDAP, a comunidade ligada à Faculdade de Biblioteconomia e Comunicação (alunos, professores/pesquisadores), equipes que realizam pesquisas, colaboradores externos e instituições parceiras.



**Figura 1:** Ambiente de inovação

**Fonte:** Elaborado pelos autores.

O modelo para representar o conhecimento foi desenvolvido para atender aos seguintes objetivos: ser capaz de representar tanto o conhecimento quanto as práticas relacionadas a esse conhecimento; ser capaz permitir a organização, a

descrição e o registro de conhecimento capturado; ser capaz de permitir que o conhecimento seja semanticamente organizado e também registrado de forma textual; ser capaz de permitir a descrição de elementos do ecossistema e o registro de conhecimentos adquiridos sobre esses elementos; e ser capaz de registrar a colaboração (colaboradores e mudanças ocorridas ao longo da colaboração). Para atender a esses objetivos, o modelo desenvolvido possui as seguintes características:

- Sua estrutura é baseada em Corpos de Conhecimento, a fim de permitir o registro textual de conhecimentos relacionados a práticas.
- Uma ontologia é usada para organizar e descrever formalmente os recursos do ecossistema, os elementos do corpo de conhecimento e os recursos do CEDAP (incluindo equipamentos, como scanners).

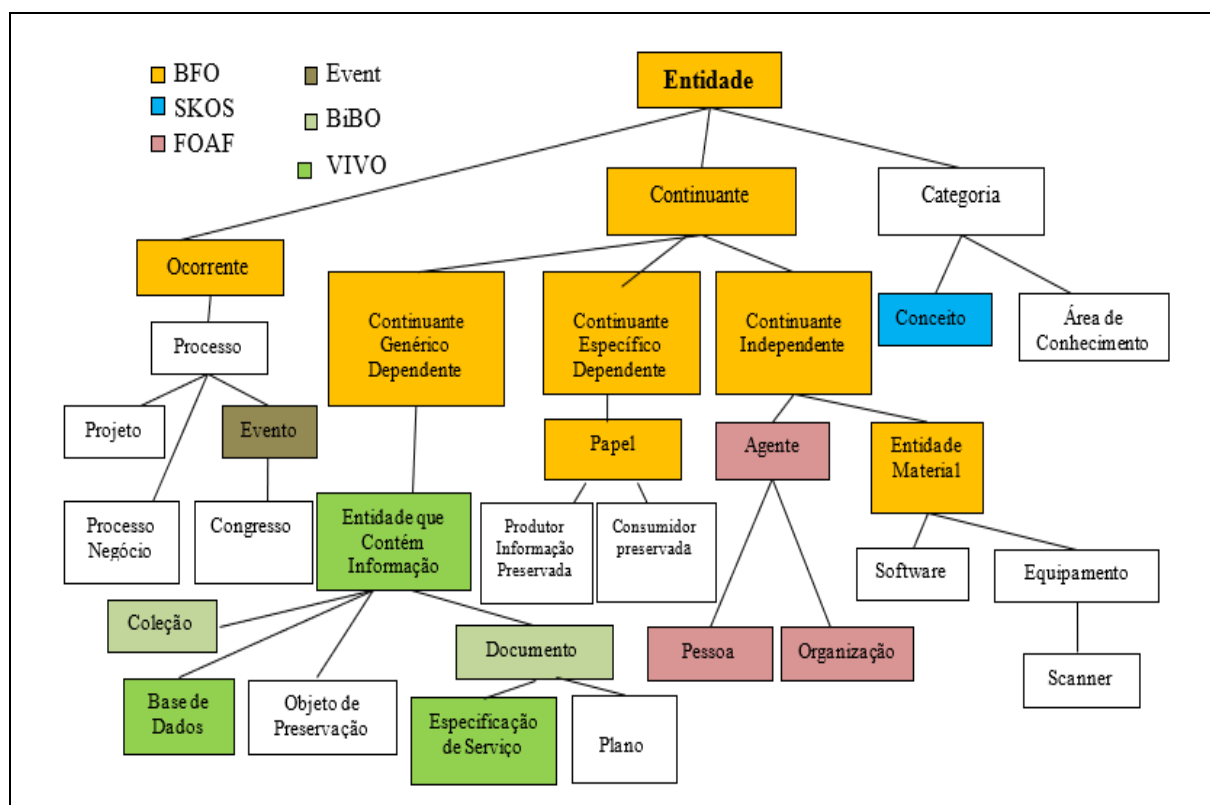
O modelo está baseado nas estruturas dos corpos de conhecimento, pois esses vão ao encontro de um dos objetivos do ambiente, que é representar tanto o conhecimento quanto às práticas relacionadas (ROCHA, 2016). Corpos de Conhecimento são instrumentos desenvolvidos por órgãos ligados a profissionais de uma disciplina (como associações profissionais em gestão de processos ou gestão de projetos). Um corpo de conhecimento contém o conhecimento estruturado que é usado por membros de uma disciplina para guiar suas práticas e trabalhos (ÖREN, 2005). Representa a agregação de conhecimento estabelecida para uma disciplina que é esperada ser aprendida por um indivíduo para ser considerado ou certificado como um praticante (ÖREN, 2005). PMBOK, BPM CBOK e SWEBOK são exemplos de corpos de conhecimento para gestão de projetos, gestão de processos e engenharia de *software*, respectivamente.

O corpo de conhecimento do CEDAP possui algumas diferenças quando comparados com os corpos de conhecimento tradicionais. Corpos de Conhecimento tradicionais são desenvolvidos por especialistas consagrados, patrocinados por organizações profissionais, e destinados a praticantes. Já no CEDAP, o desenvolvimento do corpo é realizado de forma colaborativa pelos membros do centro, no decorrer de suas atividades, a partir da busca do conhecimento e da experimentação prática (ROCHA, 2016).

Corpos de conhecimento são estruturados em áreas de conhecimento, tópicos/glossário, áreas relacionadas, padrões e técnicas (ROCHA, 2016). A relação com a prática ocorre através dos processos que são associados às áreas de conhecimento. Esses processos identificam entradas, saídas, ferramentas e técnicas. O corpo de conhecimento do CEDAP é composto por áreas de conhecimento, processos, modelos conceituais, serviços e ontologia. O **modelo conceitual** descreve formalmente e define os conceitos usados, com a finalidade de compreensão e comunicação. Visa convergência conceitual na atuação da equipe, para que todos desenvolvam os conhecimentos e as práticas a partir de um entendimento comum e compartilhado. As **áreas de conhecimento** delimitam e registram conhecimentos necessários para o corpo, permitindo um estudo segmentado e direcionado. Os **processos** representam práticas realizadas nas áreas. Os **serviços** são os serviços a serem oferecidos pelo CEDAP e são desenvolvidos a partir da composição de processos.

A **ontologia** representa formalmente os elementos do corpo do conhecimento (como área de conhecimento, processo de negócio, conceito, objeto de preservação, assim como documentos associados às áreas de conhecimento, como especificações de serviços, planos, políticas, estratégias, estudos e guias), os recursos do CEDAP (como *software*, equipamento, *scanner*, pessoas, serviços, projetos e documentos decorrente da execução de serviços e projetos) e os elementos do ecossistema (como pessoas, produtores e consumidores de informação, organizações, pessoas, projetos, conferências, serviços de bases de dados, periódicos, assim como documentos que representam guias, planos, políticas,

estratégias, padrões, modelos de referência, artigos, etc.). A Figura 2 apresenta, de forma resumida, classes da ontologia e sua hierarquia. Nessa Figura, a cor atribuída a uma classe indica a ontologia a partir da qual a classe foi seletivamente reutilizada.



**Figura 2:** Ontologia

**Fonte:** Elaborado pelos autores.

O desenvolvimento da ontologia teve como requisitos: ser voltada para Web Semântica; reutilizar ontologias populares e permitir extensões, isto é, permitir que a mesma evolua, sendo capaz representar novas entidades que venham a surgir no corpo de conhecimento ou no ecossistema. A ontologia utiliza como base a ontologia de alto nível *Basic Formal Ontology* (BFO) e reutiliza, de forma seletiva, ontologias populares como VIVO-ISF (para representar ambiente acadêmico/pesquisa), FOAF (para pessoas), EVENT (para eventos), SKOS (para conceitos), BibO (para documentos) (ROCHA, 2016).

A estrutura de mais alto nível da ontologia é baseada na ontologia BFO, pois esta, oferece uma ampla base conceitual para representar a grande diversidade de elementos que compõem o ecossistema. Em BFO, entidades podem ser continuantes ou ocorrentes, representando, respectivamente, entidades que persistem através do tempo (como uma pessoa, um livro) e entidades que acontecem e se desenvolvem no tempo e no espaço (como curso, um projeto). BFO permite a representação de continuantes independentes, que não dependem de outra entidade para existirem (como uma pessoa); de continuantes especificamente dependentes, que dependem especificamente de uma entidade para existirem (como um professor, que depende de uma pessoa para existir); ou de continuantes genericamente dependentes, que têm sua existência dependente de uma entidade que pode mudar (como um livro).

A ontologia reutiliza seletivamente elementos da ontologia VIVO-ISF. Essa ontologia abrange uma grande parte dos recursos do ecossistema identificado e estende BFO. É usada em redes de comunidades de pesquisa que adotam o ambiente VIVO.

A estratégia para o desenvolvimento do ambiente também envolveu a busca de um modelo de colaboração e a busca de uma ferramenta de colaboração que atenda aos modelos de colaboração e de representação do conhecimento especificados.

Para definir a colaboração no ambiente, o Modelo 3C (FUCKS et al., 2007) foi escolhido, pois articula Coordenação, Cooperação e Comunicação da seguinte forma: para colaborar, os indivíduos têm que trocar informações (se comunicar), organizar-se (se coordenar) e operar em conjunto num espaço compartilhado (cooperar) (FUCKS, et al., 2007). As trocas ocorridas durante a comunicação geram compromissos que são gerenciados pela coordenação, que por sua vez organiza e dispõe as tarefas que são executadas na cooperação (FUCKS, et al., 2007).

O ambiente utiliza a ferramenta *Semantic Media* (SMW) (VÖLKE, 2006) para dar apoio à construção colaborativa do corpo de conhecimento e à descrição dos elementos do ecossistema e dos recursos do CEDAP. SMW foi escolhida por registrar o conhecimento em forma textual e através de descrições semânticas baseadas em ontologias, e por usar conceitos do modelo de colaboração (ROCHA, 2016). A cooperação é realizada através da construção coletiva de páginas *wiki*, e essas páginas são consideradas instâncias de classes uma ontologia. Cada página representa um elemento do corpo de conhecimento (como área de conhecimento, processo, serviço), um recurso do CEDAP (como equipamento, *software*) ou um elemento do ecossistema (como publicação, organização, projeto).

## 7 RESULTADOS

A equipe do CEDAP vem acompanhando o desenvolvimento de pesquisas realizadas na UFRGS, para identificar como as equipes destas pesquisas criam, processam e analisam dados; e quais são as expectativas e estratégias dessa equipes para armazenar e para promover o acesso e o reuso dos dados. A equipe do CEDAP também vem buscando conhecimentos na área, investigando pesquisas e práticas no estado da arte. Combinando os conhecimentos adquiridos com as experiências obtidas nas pesquisas acompanhadas, a equipe do CEDAP formata e promove serviços de dados da pesquisa. Tanto o CEDAP quanto as pesquisas acompanhadas têm obtido ganhos com essas interações. O CEDAP, na oportunidade em poder formatar seus serviços com inovação; as pesquisas, com os benefícios trazidos pelos serviços de dados.

O ambiente está operacional, possibilitando o desenvolvimento evolutivo de serviços, no contexto da inovação. Até o momento foram produzidas 148 páginas, que foram editadas 4900 vezes, abrangendo três corpos de conhecimento (digitalização, preservação digital, dados da pesquisa), com 18 áreas de conhecimento, 48 processos de negócio (classificados em planejamento e execução), entre outros. Resultados parciais do ambiente são apresentados por Pavão, Caregnato e Rocha (2016).

O corpo de conhecimento em digitalização está com grande parte de suas áreas de conhecimento e processos definidos, conforme resultados detalhados por Rocha (2016). A digitalização foi organizada nas áreas de conhecimento: seleção, conversão, controle de qualidade, produção das imagens, descrição e representação e, gerenciamento da digitalização. A organização do conhecimento nessas áreas trouxe um direcionamento para os estudos e para as práticas, e facilitou o processo de absorção do conhecimento e socialização das práticas adquiridas. Proporcionou agregação de valor aos serviços oferecidos, através das relações estabelecidas entre a pesquisa (áreas de conhecimento) e a prática (processos e serviços).

O corpo de conhecimento em digitalização trouxe clareza e formalização aos processos de digitalização. Grande parte dos processos foi identificada, sendo que muitos já foram detalhados. Isso traz grandes benefícios para o gerenciamento do

CEDAP, pois facilita a implantação da gestão de processos do CEDAP, visto que é ponto de partida para o mapeamento dos processos.

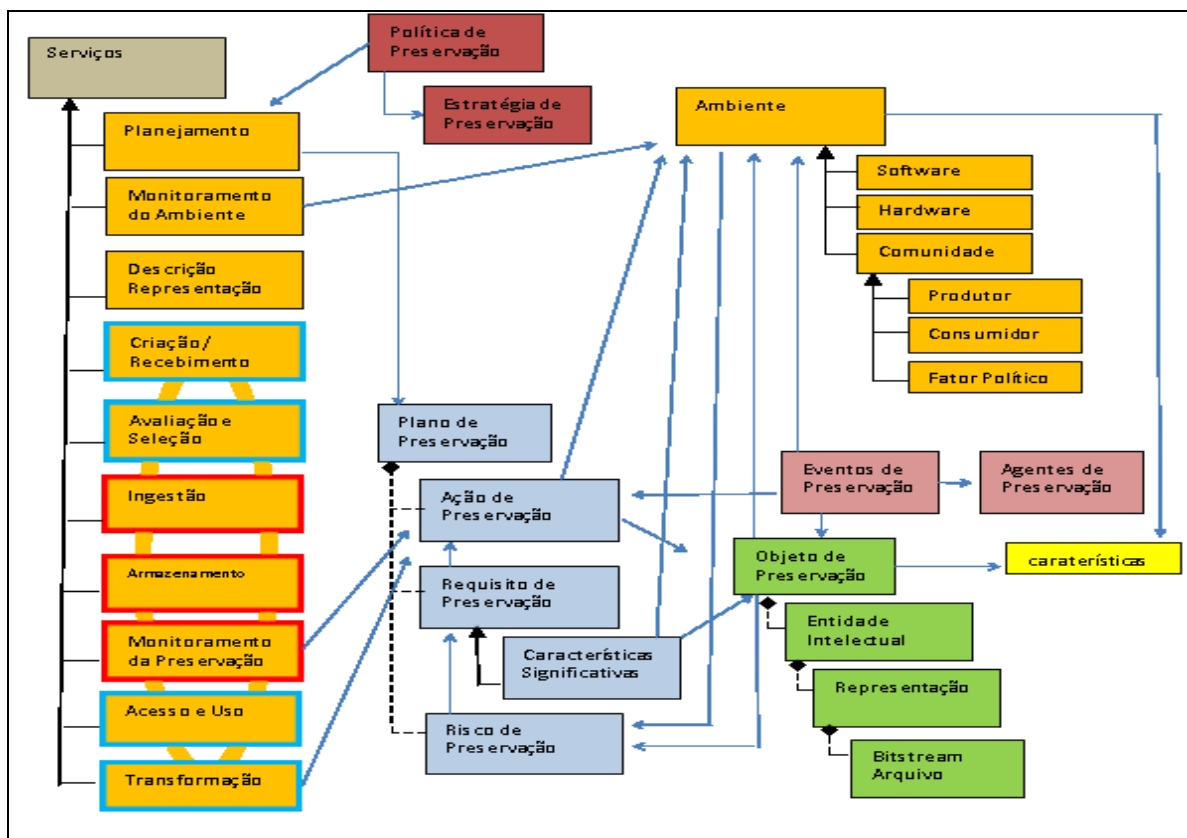
No corpo de conhecimento em digitalização foram desenvolvidos estudos e recomendações, como recomendações para definir critérios de seleção, recomendações para definir configurações de imagens digitais e estudos para representar os objetos digitalizados em pacotes. Essas recomendações e estudos trouxeram segurança e qualidade na oferta de serviços do CEDAP, visto que estão alinhadas com práticas e pesquisas atuais, desenvolvidas por instituições de renome mundial. A área de conhecimento, descrição e representação, trouxe a perspectiva de produzir as representações digitais dos objetos digitalizados em pacotes de acesso e de preservação, conforme recomendações de preservação digital.

O corpo de conhecimento em digitalização também trouxe ao CEDAP a perspectiva em atuar na digitalização de obras raras, hoje restrita a poucos centros de referência. Foram desenvolvidos estudos sobre critérios de seleção de obras raras, e sobre como representar e descrever estruturalmente os representantes digitais dessas obras (matrizes e derivadas). Esses estudos demonstraram que obras raras devem ser atendidas por serviços diferenciados.

O ambiente possibilitou a descrição e o registro de conhecimentos aprendidos no uso e no estudo dos equipamentos (*scanners*) e *softwares*. Cada equipamento e *software* é descrito através da ontologia, e é representado através de uma página. Essa página contém descrições e dicas de operações, na forma textual.

O corpo de conhecimento em preservação digital trouxe para o CEDAP a perspectiva de prover serviços em consonância com aqueles realizados por instituições de referência mundial. Nele, foi desenvolvido um modelo conceitual cujos resultados são detalhados por Pavão, Caregnato e Rocha (2016).

O modelo conceitual (Figura 3) representa uma compilação de conceitos presentes em padrões e modelos consagrados: os modelos conceituais do projeto PLANETS (DAPPERT; FARQUHAR, 2009) e de *California Digital Library* (2010); os padrões de metadados PREMIS (preservação digital) e METS (estrutural); o ciclo de vida da curadoria digital (HIGGINGS, 2008), o modelo de referência *Open Archival Information System* (CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS, 2012), o plano de preservação de Becker (2009), entre outros. O modelo conceitual permitiu ao CEDAP ter uma visão própria, atual, integrada e conceitualmente bem definida da preservação digital, possibilitando a oferta de produtos e serviços bem delimitados e em consonância com os avanços atuais.



**Figura 3:** Modelo conceitual da preservação digital

**Fonte:** Pavão, Caregnato e Rocha (2016).

O modelo conceitual (Figura 3) especifica a preservação digital através do **ambiente** de preservação, que restringe e delimita o objeto de preservação; do **plano de preservação**, que define ações de preservação a serem realizadas devido aos riscos identificados para objetos e ambiente de preservação; **serviços de preservação**, que são serviços nucleares que dão suporte a um objetivo da preservação; e **ações de preservação**, que são executadas devido aos riscos identificados. Alguns serviços formam o ciclo de vida da preservação digital, atuando em curadoria e preservação. Na Figura 3, os serviços de curadoria estão representados com bordas em azul e os serviços de preservação estão representados em bordas em vermelho. Os serviços que ocorrem em ciclo estão interligados por uma elipse.

O modelo conceitual subsidiou a definição das áreas de conhecimento em preservação digital e de seus processos. As áreas de conhecimento em preservação digital são: ambiente da preservação digital, planejamento da preservação digital, descrição e representação, acesso e uso da informação, produção da informação a ser preservada, transformação da informação preservada, preservação da informação, gerenciamento da preservação digital.

O corpo de conhecimento em dados da pesquisa está em fase de composição/consolidação das áreas de conhecimento. Há uma correlação muito grande entre as áreas de conhecimento de dados da pesquisa e da preservação digital, visto que seus ciclos de vida há um alinhamento entre seus ciclos de vida.

Identificou-se a definição do ciclo de vida do dado da pesquisa do CEDAP como chave para a produção de serviços com qualidade na área. Esse ciclo de vida está sendo construído partir da compilação e adaptação dos ciclos de vida existentes e consagrados.

## 8 CONSIDERAÇÕES FINAIS

CEDAP está na fase consolidação e experimentação dos seus produtos/serviços. A estratégia de buscar um ambiente de construção colaborativa e aberto, baseado em corpo de conhecimento, leva a uma evolução não linear do processo de P&D, em que a oferta dos produtos e serviços evolui em paralelo com a busca do conhecimento e do estado da arte, em um processo de inovação aberta e interativa. Essa estratégia vai ao encontro do ecossistema da pesquisa e do desenvolvimento em dados da pesquisa, que possui uma natureza aberta e colaborativa. Através dela, serviços e produtos do CEDAP passam a absorver resultados de pesquisas e práticas, assim como padrões, normas e recomendações atuais em nível internacional, com perspectiva de estarem em pé de igualdade com os melhores serviços oferecidos atualmente.

## REFERÊNCIAS

- BECKER, C. et al. Systematic planning for digital preservation: evaluating potential strategies and building preservation plans. **International Journal on Digital Libraries**, v. 10, n. 4, p. 133-157, 2009.
- BLAKE, J. Ten Quick Tips for Using the Gene Ontology. **PLoS Computational Biology**, v. 11, 2013.
- BRASIL. Lei nº 13.243, de 11 de janeiro de 2016, dispõe sobre estímulos ao desenvolvimento científico, à pesquisa, à capacitação científica e tecnológica e à inovação e altera outras leis. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, janeiro 2016.
- BRASIL. **Lei n.º 10.973**, de 2 de dezembro de 2004. Dispõe sobre incentivos à inovação e à pesquisa científica e tecnológica no ambiente produtivo e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 3 dez. 2014.
- CALIFORNIA DIGITAL LIBRARY. **UC3 Curation Foundations**. 2010.
- CARLSON, J. The use of Lifecycle Models in developing and supporting data services. In RAY, Joyce M. (Ed.). **Research data management: Practical strategies for information professionals**. Purdue: University Press, 2014.
- CHESBROUGH, H. Open innovation: a new paradigm for understanding industrial innovation. In CHESBROUGH, H.; VANHAVERBEKE, W.; WEST, J. (Eds). **Open Innovation: Researching a New Paradigm**. Oxford: Oxford University Press, 2006.
- CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. **Reference model for an Open Archival Information System (OAIS): recommended practice**. Washington, DC: CCSDS, 2012.
- DAPPERT, A; FARQUHAR, A. Modelling organizational preservation goals to guide digital preservation. **International Journal of Digital Curation**, v. 4, n. 2, p. 119-134, 2009.
- FUKS, H.; et al. The 3C Collaboration Model. In: KOCK, N et ali. **The Encyclopedia of E-Collaboration**. Hershey:Information Science, 2007.
- GANZER, et al. Modelo de processo tecnológico: uma evolução histórica de modelo linear para modelo interativo. **Gestão Contemporânea**, Porto Alegre, n. 16, 2014.
- HIGGINS, S. The DCC curation lifecycle model. **International Journal of Digital Curation**, v. 3, n. 1, p. 134-140, 2008.
- LORD, Philip; MACDONALD, Alison. **e-Science Curation Report: Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision**.

2003. Disponível em: <[http://www.jisc.ac.uk/uploaded\\_documents/e-ScienceReportFinal.pdf](http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf)>. Acesso em: 17 jul. 2013.

NATIONAL SCIENCE FOUNDATION. **Proposal and Award Policies and Procedures Guide**. 2012. Disponível em: <[http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/nsf13\\_1.pdf](http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/nsf13_1.pdf)>. Acesso em: 17 jul. 2013.

ÖREN, T. Toward the body of knowledge of modeling and simulation. In: INTERSERVICE/INDUSTRY TRAINING, SIMULATION, AND EDUCATION CONFERENCE, 2005, Flórida. **Proceedings...** Flórida, 2005.

PAVÃO, C.; CAREGNATO, C.; ROCHA, R. Implementação da preservação digital em repositórios: conhecimento e práticas. **Revista Digital de Biblioteconomia e Ciência da Informação**, v. 14, n. 3, 2016.

ROCHA, R. Um sistema baseado na construção coletiva de em corpo de conhecimento para apoiar as atividades de um centro de digitalização. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Salvador. **Anais...** Salvador: UFBA, 2016.

ROTHWELL, R. Towards the fifth-generation innovation process. **International marketing review**, v. 11, n. 1, p. 7-31, 1994.

SALLANS, A; DONNELLY, M. DMP Online and DMPTool: different strategies towards a shared goal. **International Journal of Digital Curation**, v. 7, n. 2, p. 123-129, 2012.

SMITH, B. et al. **Basic Fomal Ontology 2.0**: draft specification and user's guide. Saarbrücken, Germany: IFOMIS, 2014. Disponível em: <<http://ifomis.uni-saarland.de/bfo/>>. Acesso em: 17 jul. 2013.

UK DATA SERVICE. **Research data lifecycle**. 2017. Disponível em: <<https://www.ukdataservice.ac.uk/manage-data/lifecycle>>. Acesso em: 15 nov. 2017.

VÖLKEL, M., et al. Semantic wikipedia. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 15., 2006, New York. **Proceedings...** New York: ACM, 2006.

WHYTE, A., TEDDS, J. Making the Case for Research Data Management. **DCC Briefing Papers**. Edinburgh: Digital Curation Centre. 2011. Disponível em: <<http://www.dcc.ac.uk/resources/briefing-papers>>. Acesso em: 20 fev. 2017.