

ESTRUTURAS DE REPRESENTAÇÃO PARA REUSO DE DADOS NO CONTEXTO DA ECOLOGIA DE PESQUISA: CRIS INSTITUCIONAL

REPRESENTATION STRUCTURES FOR DATA REUSE IN THE CONTEXT OF RESEARCH ECOLOGY: INSTITUTIONAL CRIS

Emanuelle Torino^a

Caio Saraiva Coneglian^b

Silvana Aparecida Borsetti Gregorio Vidotti^c

RESUMO

Introdução: O *Current Research Information System* (CRIS) desponta como resposta à necessidade de otimização no processo de gestão do grande volume de dados e informações que permeiam o ciclo de vida da pesquisa. Tais informações estão dispersas em múltiplos ambientes informacionais, de características complementares, hoje isolados que podem ser integrados, e fornecer um panorama mais amplo da pesquisa. **Objetivo:** Nesse contexto, o presente estudo discute as possibilidades de integração de dados de diferentes sistemas que armazenam dados relacionados à ecologia de pesquisa, utilizando estruturas de representação e infraestrutura semântica para compor uma ecologia de pesquisa constituída como um CRIS institucional. **Metodologia:** Como procedimento metodológico, utiliza a revisão bibliográfica para o embasamento teórico-conceitual do estudo, visando contextualizar a infraestrutura da ecologia de pesquisa, em especial as estruturas de representação, as ontologias e os princípios FAIR. **Resultados:** Como resultado é apresentado um modelo conceitual de integração dos dados de uma ecologia de pesquisa, capaz de subsidiar a composição de um CRIS institucional. **Conclusões:** A compatibilização de estruturas de representação disponíveis em múltiplos sistemas de informação e a utilização de tecnologias já existentes visando constituir um CRIS institucional exigem que a atuação do profissional de informação esteja embasada em conhecimentos de arquitetura da informação, metadados, padrões e protocolos de comunicação, interoperabilidade e tecnologias da web semântica, conforme modelo proposto no presente estudo.

Descritores: Representação da Informação. Reuso de Dados. CRIS. Ontologia.

^a Doutoranda em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista (PPGCI/UNESP). Bibliotecária da Universidade Tecnológica Federal do Paraná (UTFPR). E-mail: emanuelle@utfpr.edu.br

^b Doutor em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista (Unesp). Docente do Centro Universitário Eurípedes de Marília (UNIVEM). E-mail: caio.coneglian@gmail.com

^c Doutora em Educação pelo Programa de Pós-Graduação em Educação da Universidade Estadual Paulista (Unesp). Docente do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista (Unesp). E-mail: silvana.vidotti@unesp.br

Princípios FAIR.

1 INTRODUÇÃO

Ao longo dos anos, as discussões concernentes às atividades de pesquisa estiveram relacionadas principalmente à gestão dos objetos que resultam do seu processo. Nesse sentido, evidenciam-se estudos que abordam ambientes de publicação científica, como periódicos, e de disponibilização e ampliação de visibilidade, como repositórios digitais. Mais recentemente, evidenciou-se que durante do processo investigativo são produzidos dados de pesquisa (simulações, equações, áudios, vídeos, softwares, planilhas, esquemas, experimentos), cuja necessidade de gestão e disseminação assume o centro das atuais discussões no bojo da Ciência Aberta.

Tais dados, em geral, são resultados de pesquisas que receberam financiamento com recurso público. Por conseguinte, instâncias governamentais e agências de fomento, aos poucos, sinalizam a necessidade de gestão e disponibilização dos dados de pesquisa, em repositórios de dados de pesquisa, para uso em acesso aberto e contínuo. Para isso, é utilizado um repositório que requer um processo diferenciado de tratamento e disponibilização dos conteúdos integrais e cujo contexto é representado por metadados (SAYÃO; SALES, 2016). Dessa forma têm-se a relevância do processamento, distribuição e uso aos quais estão atrelados processos de organização e representação da informação.

Tais ambientes gerenciam uma parcela do ciclo de vida da pesquisa, seus resultados, expressos em publicações em texto completo e, mais recentemente, os dados de investigação, que permitem novos resultados a partir de seu reuso, alavancando processos de reprodutibilidade e produção científica que permitem confirmar ou refutar conhecimentos científicos. Esses objetos, resultantes das atividades de investigação são disponibilizados em ambientes digitais, cuja representação é realizada utilizando metadados.

Contudo, há um grande volume de informações geradas ao longo do ciclo de vida da pesquisa (de uso claramente administrativo e de gestão, relacionado a pesquisadores, projetos de pesquisa, financiamentos, recursos de fomento à pesquisa, instituições e estruturas), bem como os relacionamentos que

permeiam esse universo, cuja gestão ocorre de forma heterogênea entre instituições, agências de fomento e órgãos governamentais.

Nesse cenário, torna-se emergente gerenciar informações e relacionamentos existentes no âmbito da pesquisa, e estabelecer o contexto do processo de investigação de instituições de ensino e pesquisa. Com isso, tem-se uma visão macro da ecologia informacional institucional, que possibilita integrar informações de projetos de pesquisa, atores envolvidos, recursos empregados, resultados obtidos e demais relacionamentos possíveis.

Muitas dessas informações já são gerenciadas por sistemas de informação com funcionamento independente, como o Diretório de Grupos de Pesquisa, o Currículo Lattes, o ORCID iD, as Bibliotecas Virtuais de Agências de Fomento, os periódicos científicos, os repositórios digitais de publicações e de dados de pesquisa; cuja integração pode consistir em uma infraestrutura de informação completa sobre o processo da pesquisa, o que pode ocorrer no ambiente interno de uma instituição. Nesse sentido, destaca-se a necessidade de planejamento considerando uma única entrada de informações e objetos, e várias possibilidades de uso e reuso a partir da integração de sistemas de informação, por meio das estruturas de representação.

Para que isso ocorra, é necessário integrar os sistemas de informação, o que requer compatibilização dos dados existentes nos diferentes sistemas, e exige o envolvimento de todos os detentores dos ciclos de vida dos dados. Contudo, em nível micro, é possível que sejam realizadas integrações de dados existentes em sistemas institucionais visando o estabelecimento de uma ecologia informacional de pesquisa a ela relacionada, que pode incluir dados de fontes externas coletados por meio de protocolos de comunicação.

Essa integração pode se constituir em um sistema de informação de pesquisa corrente, do inglês, *Current Research Information System* (CRIS), também chamado de gerenciamento de informações de pesquisa, tradução de *Research Information Management* (RIM). É um ambiente de gestão do ciclo de vida da pesquisa que interliga o projeto de pesquisa, os envolvidos, sejam eles pessoas, instituições e/ou agências de fomento e seus resultados, utilizando-se de modelos de referência como o *Common European Research Information*

Format (CERIF) e de outras tecnologias, como as da web semântica.

Assim, a presente pesquisa parte da necessidade de gerenciar o ciclo de vida da pesquisa, utilizando-se da representação da informação e de infraestrutura semântica em ecossistema de pesquisa. Para tanto, objetiva discutir as possibilidades de integração de dados de diferentes sistemas que armazenam dados relacionados à ecologia de pesquisa, por meio de estruturas de representação e infraestrutura semântica para compor uma ecologia de pesquisa, constituída como um CRIS institucional.

Como procedimento metodológico, utiliza a revisão bibliográfica para o embasamento teórico-conceitual do estudo visando contextualizar a infraestrutura da ecologia de pesquisa, em especial as estruturas de representação, as ontologias e os princípios FAIR. Como resultados é apresentado um modelo conceitual de integração dos dados de uma ecologia de pesquisa, capaz de subsidiar a composição de um CRIS institucional.

2 ECOLOGIA DE PESQUISA: O CRIS INSTITUCIONAL

O termo “ecologia” foi utilizado na área da Ciência da Informação em dois contextos destacados, inicialmente por Davenport (1998), que abordou a ecologia da informação, no âmbito da Gestão da Informação e por Oliveira (2014), que tratou das ecologias informacionais complexas, no contexto da Arquitetura da Informação Pervasiva.

O denominador entre as duas abordagens centra-se em um olhar “ecológico” para a informação. Para Davenport (1998), a ecologia da informação possui quatro atributos: integração de diferentes tipos de informação; reconhecimento de mudanças evolutivas; ênfase na observação e na descrição; e ênfase no comportamento pessoal e informacional. Enquanto para Oliveira (2014, p. 134) as ecologias informacionais complexas se constituem como “[...] conjunto de espaços, ambientes, canais, mídias, tecnologias e sujeitos com seus comportamentos, todos interligados e conectados de maneira holística pela a informação.” Segundo Vechiato, Oliveira e Vidotti (2016, p. 54) “Na ecologia persiste o sistemismo de ambientes e relações complexas que ocorrem de forma intra e extraecológica”.

Partindo da definição de Oliveira (2014), no contexto do presente trabalho as ecologias de pesquisa estão diretamente vinculadas à Arquitetura da Informação Pervasiva (OLIVEIRA, 2014), uma vez que são constituídas de um conjunto de ambientes informacionais digitais (CAMARGO; VIDOTTI, 2011; OLIVEIRA, 2014), nos quais os objetos de pesquisa são armazenados e, de sistemas de informação baseados na Arquitetura da Informação Clássica, de abordagem sistêmica (MORVILLE; ROSENFELD, 2006).

Desta forma, têm-se, no contexto do ciclo de vida da pesquisa, em ecologias de pesquisa, a pervasividade da informação, que pode estar armazenada em diferentes sistemas de informação e refere-se a sujeitos, contextos e conteúdos. Assim, a ecologia de pesquisa, como as ecologias informacionais complexas, é constituída por um conjunto de ambientes, tecnologias e sujeitos interligados e entrelaçados pela informação (OLIVEIRA, 2014).

No contexto institucional, a ecologia de pesquisa ocorre no relacionamento entre informações dispersas em diferentes sistemas, que gerenciam dados de organização, infraestrutura para pesquisa, projetos de pesquisa, grupos de pesquisa, pesquisadores, fomentos, resultados de pesquisa. Os quais podem estar armazenados em sistemas internos e externos à instituição.

Ao que neste estudo chamamos de informações do ciclo de vida da pesquisa, para Lopatenko (2001, tradução nossa) são dados da pesquisa:

[...] informações sobre resultados de pesquisas, projetos, publicações, organizações, pesquisadores, publicados na web desempenham um papel cada vez mais penetrante na pesquisa moderna. A alta dependência da pesquisa moderna sobre os resultados de pesquisa já alcançados produz a condição para que o pesquisador tenha capacidade para recuperar informações de pesquisa de maneira eficiente.

Pode-se inferir que tais informações podem ser encontradas nas instituições de pesquisa, contudo de forma dispersa, em diferentes sistemas de informação, com tratamento heterogêneo, sem padronização ou integração e, quiçá, em alguns contextos, ainda sem gerenciamento digital. Ao ampliar tal inferência para o relacionamento externo à instituição, por exemplo com agências de fomento ou mesmo na web, vislumbramos o quão caótica pode ser

a gestão, interoperabilidade e recuperação.

Por outro lado, para a atividade de gestão reunir essas informações possibilita retratar o cenário da pesquisa no ambiente organizacional, de forma rápida e precisa. Nesse contexto, o CRIS pode se constituir em uma plataforma única ou na integração de dados armazenados em múltiplas plataformas que armazenam dados concernentes ao ciclo de vida da pesquisa que, neste estudo, é considerado uma ecologia de pesquisa.

Os sistemas de informação de pesquisa corrente apresentam-se como uma solução para a realidade apontada (LOPATENKO, 2001; ASSERSON; JEFFERY, 2009; JEFFERY; ASSERSON, 2010; SHEPPARD, 2010). Embora ainda não sejam realidade no Brasil, já se tornaram objeto de estudo de Sales e Sayão (2015, p. 167) que definem CRIS como “um banco de dados que armazena metadados sobre atividades de pesquisa de interesse corrente”. Os autores apontam como resultado de sua implantação o gerenciamento integrado das informações que permite, dentre outras coisas, análises, avaliações e relatórios precisos, íntegros e organizados, bem como a ampliação da visibilidade da pesquisa realizada por uma instituição para pesquisadores, gestores de pesquisa, agências de fomento, mídias relacionadas à ciência e à sociedade.

Para Joint (2008, p. 571, tradução nossa), CRIS é definido como:

[...] sistema de informação que pode gerenciar toda a informação relevante da pesquisa, começando com oportunidades de financiamento, passando pelo estágio de redação e submissão de propostas, seguindo com as propostas bem-sucedidas, que se tornam projetos ativos que serão gerenciados até a conclusão – estágio no qual são gerados resultados, muitos dos quais são publicações ou algum outro artefato da atividade de pesquisa.

Para a constituição de um CRIS institucional a partir da integração das estruturas de representação, será necessário conhecer os sistemas de informação que as armazenam, as formas de exibição dos dados, os protocolos de comunicação disponíveis, para definir a estrutura de conversão.

Nesse sentido, recomenda-se a manutenção de um repositório de metadados que possibilite a gestão dos dados e metadados disponíveis em cada sistema de informação institucional, visando utilizá-los sempre que necessário. Brandt (2020, p. 81) ao abordar o repositório de metadados estabelece uma

comparação com um catálogo e o define como “um sistema que armazena representações dos metadados de negócio, juntamente com sua descrição e demais atributos pertinentes ao negócio, além de fornecer sua localização e elementos para sua gestão: quais os processos e sistemas utilizam determinado metadado e quem é seu gestor na empresa”.

A autora aborda metadados de negócios e afirma que a junção desses em um repositório de metadados permite um olhar macro da arquitetura de informação da instituição, o que possibilita melhora na gestão da informação, uma vez que o tratamento passa a ser realizado de forma normalizada em todos os usos. Vale destacar que este repositório pode ser construído utilizando um software ou, de forma mais simples, ser estruturado em planilhas, considerando-se as vantagens e limitações. Brandt (2020) apresenta um modelo conceitual para a constituição de um repositório de metadados e sua aplicação.

O tratamento dos metadados é imprescindível para realizar o mapeamento dos metadados heterogêneos advindos dos diferentes sistemas e a conversão, visando compatibilizá-los com o modelo de dados utilizado, necessários ao estabelecimento de estruturas que possibilitem a constituição e o funcionamento do CRIS institucional.

Tal funcionamento pode estar embasado em diferentes modelos de dados, utilizados na camada de representação da informação. Neste estudo, abordamos o CERIF por seu vínculo como modelo de referência para CRIS e ontologias, especialmente as Ontologias VIVO e CERIF, por suas abordagens vinculadas ao contexto das informações acadêmicas.

3 COMMON EUROPEAN RESEARCH INFORMATION FORMAT (CERIF)

Considerando que o desenvolvimento de CRIS se fortaleceu na Europa, foi estabelecido um modelo de referência para o desenvolvimento de CRIS tido como recomendação oficial, o CERIF¹ (mantido pelo euroCRIS²) que consiste em um padrão internacional de dados relacionais para o armazenamento e

¹ Disponível em: <http://www.eurocris.org/cerif/main-features-cerif>. Acesso em: 07 jul. 2020.

² Disponível em: <http://www.eurocris.org/>. Acesso em: 07 jul. 2020.

interoperabilidade de informação de pesquisa, no modelo entidade-relacionamento.

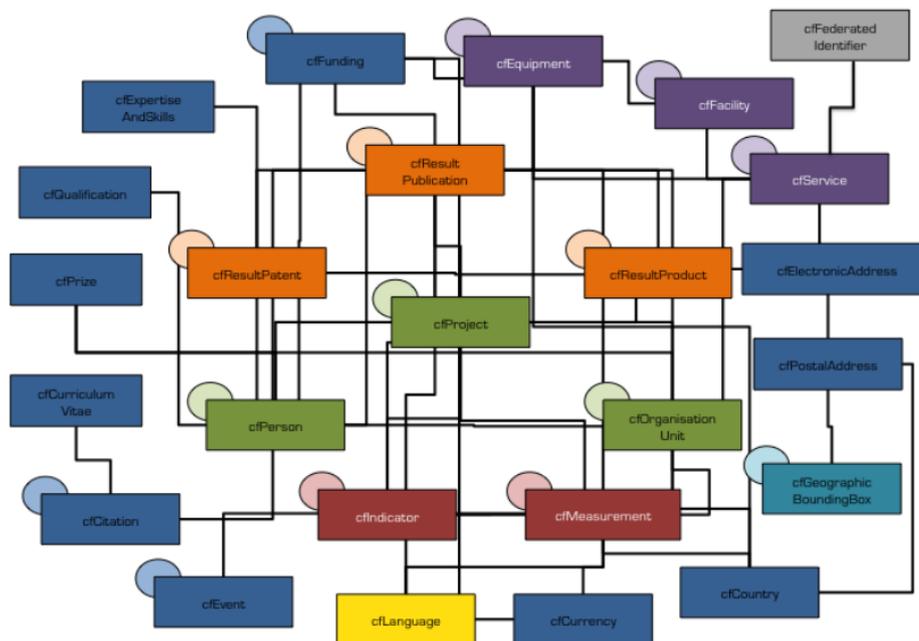
O CERIF consiste em “Um extenso modelo de dados, cobrindo todos os aspectos da informação de pesquisa, que pode ser usado como um modelo de referência para a criação de novos, bem como a extensão de sistemas de informação de pesquisa existentes”. (EUROCRIS, 2020d). Disponibiliza um padrão de troca de informação entre CRIS e outros sistemas de informação, como repositórios digitais e de dados de pesquisa, por meio do CERIF-XML.

O modelo CERIF está estruturado em três níveis: a) conceitual - consiste nas entidades e seus relacionamentos, materializada nas especificações; b) lógico - refere-se à descrição das entidades de pesquisa e seus relacionamentos, que consiste no próprio modelo; e c) físico - a formalização das entidades de pesquisa e seus relacionamentos, que consiste em scripts SQL para múltiplos bancos de dados.

O modelo CERIF (Figura 1) apresenta uma visão do domínio da pesquisa e estabelece relacionamentos entre as entidades, tendo como ponto central o projeto, e agrupa as entidades por contextos: atores (verde), resultados (laranja), indicadores/impacto (vermelho) e infraestrutura (roxo). No entorno dessas entidades há um segundo nível (azul e amarelo) no qual estão as entidades que capacitam o modelo relacional, por meio de ligações com as entidades básicas. Destaca-se que no modelo atual há um identificador federado (cinza) e de uma delimitação geográfica (azul petróleo), que permite a ligação com recursos externos, por meio do identificador.

Cumprе destacar que houve ainda a extensão da entidade `cfResultProduct` para incorporar ‘dados de pesquisa’ que pode estabelecer relacionamentos diversos como, por exemplo, com equipamento, instalação, serviço, financiamento, medição, indicador, patente, publicação, projeto, pessoa, organização e identificador federado (euroCRIS, 2020c).

Figura 1 - Entidades e Relacionamentos CERIF



Fonte: euroCRIS (2020a)

Dentre as entidades, há aquelas consideradas entidades de ligação (link) cujo rótulo formal é definido na camada semântica, sendo a entidade `cfClassification` (`cfClass`) uma das principais, uma vez que seu identificador interno `cfClassId` é utilizado para referências em todas as entidades de ligação no CERIF. As descrições estão no vocabulário CERIF - disponível em HTML, RDF e XML³.

Além disso, considerando o modelo entidade-relacionamento, cada entidade do CERIF conta com um identificador interno que atua como chave primária, utilizada para que os registros sejam reconhecidos e agregados a um objeto, atendendo às consultas ou regras (euroCRIS, 2020a).

A implementação de um CRIS pode ocorrer utilizando o modelo CERIF completo ou um subconjunto de entidades. Contudo, é necessário que haja um esquema estruturado que possibilite a interoperabilidade de dados, quer seja em estruturas CRIS ou a partir de dados provenientes de diferentes sistemas.

Para Sales e Sayão (2015, p. 170), o “CERIF desponta como a norma mais importante e conceitualmente mais bem elaborada [...] [para] o intercâmbio

³ Disponível em: <https://cerif.eurocris.org/vocab/>. Acesso em: 07 jul. 2020.

de informações entre as plataformas CRIS no âmbito da Comunidade Européia”.

O modelo se consolidou ao longo dos anos e, em decorrência disso, está em processo de refatoração, que pretende estender o modelo ou propor extensões. “Isso envolveria atualizar a especificação do CERIF para práticas de engenharia de software mais atuais que tenham modularidade e extensibilidade em seu núcleo, de modo que seja mais fácil descentralizar o trabalho no padrão e nas ferramentas.” (euroCRIS, 2020b, tradução nossa).

Essa alteração consiste na manutenção de um modelo central pelo euroCRIS que cobriria as principais entidades sobre o qual possam ser desenvolvidas extensões, em código aberto, de forma modular e independente por qualquer interessado. A representação desta alteração prevê inicialmente cinco estruturas, sendo: modelo conceitual expressões/codificação, acesso/consulta/transformação, ferramentas (*software*), e certificação/padrões. Vale ressaltar que há previsão de expansão das linguagens para uso de JSON e OWL, cujo uso é crescente na web semântica.

4 ONTOLOGIAS

As ontologias, no contexto da Ciência da Informação, estão sendo utilizados como importantes instrumentos de representação da informação. Tais instrumentos são capazes de representar um domínio, com uma alta expressividade, permitindo a definição de relacionamentos, hierarquia, instâncias, entre outros aspectos. Santarem Segundo (2010) define as ontologias sob uma visão mais relacionada com a área da Ciência da Informação. O autor expõe que as ontologias são construídas pela:

[...] necessidade de um vocabulário compartilhado em que um conjunto de informações possam ser trocadas e também reusadas pelos usuários de uma comunidade. Considere os usuários de uma comunidade seres humanos ou agentes inteligentes (SANTAREM SEGUNDO, 2010, p. 104).

Nessa definição, verifica-se a inserção de elementos fundamentais quando se trata de ontologias, como a questão dos vocabulários compartilhados. A ontologia é uma contextualização de um domínio, podendo ser utilizada como um vocabulário que poderá ser reusado para diversas situações e usuários.

Uma definição que insere ainda mais profundamente elementos da Ciência da Informação é dada por Campos e Campos (2014, p. 3827-3828), que afirmam que as ontologias fornecem “[...] um modelo para representar os pressupostos epistemológicos e ontológicos, relevantes para o entendimento de pesquisas e seu tratamento computacional através das iniciativas de dados interligados abertos, mas sua elaboração é um processo custoso”. As autoras ao inserirem questões como os pressupostos epistemológicos e ontológicos, incluem que as ontologias devem ser uma representação bastante fiel que possibilita a expressão de uma área do conhecimento e de um domínio, por meio do uso de ontologias. Por tais motivos, as pesquisadoras dizem que o processo de elaboração de ontologias é bastante custoso.

Por meio de todas essas definições, verifica-se características fundamentais dentro do conceito de ontologias, em que as ontologias apresentam um modelo computacional que representa determinados contextos, sendo que tais modelos devem buscar ser consensuais entre as comunidades relacionadas àquele determinado domínio. As definições de Santarem Segundo (2010) e Campos e Campos (2014) permitem inferir que as ontologias podem ser utilizadas como vocabulários compartilhados para diversos domínios, especialmente aqueles relacionados a iniciativas de dados abertos que devem necessariamente possuir vocabulários compartilhados.

Vale destacar ainda que a relação entre ontologias e *web* semântica começou a ser desenhada na proposta da *web* semântica no texto de Berners-Lee, Hendler e Lassila (2001). Nesse texto, os autores relatam a necessidade da *web* semântica possuir vocabulários que permitam ser realizadas inferências, em que estejam definidas regras, para conduzir os agentes computacionais a processarem a semântica do contexto de cada página.

Santarem Segundo e Coneglian (2015, p. 227) complementam o conceito de ontologias, no contexto da *web* semântica, dizendo que “[...] entende-se as ontologias como: artefatos computacionais que descrevem um domínio do conhecimento de forma estruturada, através de: classes, propriedades, relações, restrições, axiomas e instâncias”.

Por meio dessas definições é possível verificar que o papel das ontologias

na *web* semântica é necessário quando se trata de explicitar as relações existentes entre os dados, embasado no domínio a qual pertence as informações. Resumindo, a ontologia, no contexto da *web* semântica, apresenta diversas funções como de evitar a ambiguação dos sentidos dos dados, realizar inferências, e permitir que diversos tipos de relações ocorram.

Catarino e Souza (2012) descrevem que a *web* semântica necessita ter vocabulários para representarem conceitos, relacionamentos, restrições e propriedades entre os dados. As autoras descrevem ainda que não existe uma recomendação da W3C de qual tecnologia utilizar mas, elas relatam que costuma-se utilizar vocabulários mais simples para realizar a descrição de coleções mais simples, e o uso de ontologias para coleções mais complexas.

4.1 ONTOLOGIA VIVO

A Plataforma VIVO permite o gerenciamento das informações acadêmicas e científicas de pesquisadores em que eles podem inserir informações e gerenciar pesquisas que estão ocorrendo, inserindo informações como discentes vinculados, publicações realizadas, disciplinas ministradas, entre outras. A consequência da inserção dessas informações é a geração automática dos dados em formatos RDF e semanticamente interligados.

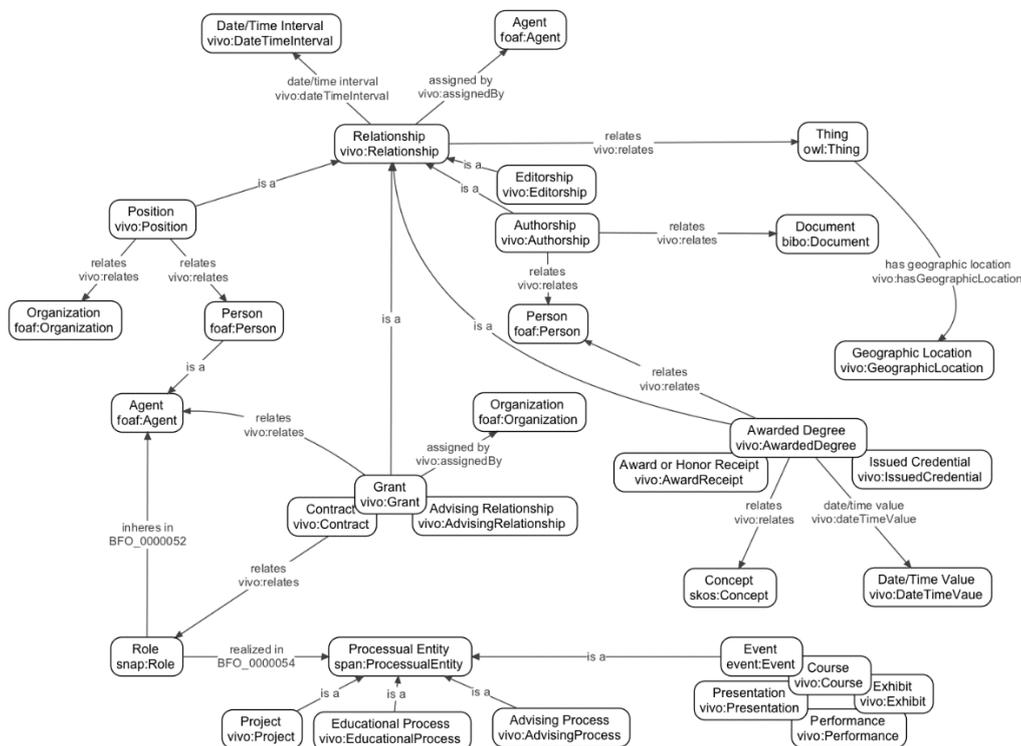
Efetivamente, a Plataforma VIVO armazena todos os seus dados seguindo os princípios da *web* semântica, em RDF, o que os torna sempre atualizados e publicados nos padrões do *linked data*. Nesse sentido, uma característica que é importante destacar é referente a estrutura que os dados serão disponibilizados, pois os dados de *linked data* necessitam seguir uma estrutura que insira um nível de semântica mais elevado, além de possibilitar que ligações com outros *datasets* sejam realizadas (LYRISIS, 2016).

No contexto da Plataforma VIVO, há uma ontologia que permite estruturar todos os dados armazenados e gerenciados por este sistema, sendo a estrutura que conecta tais dados com outras bases externas. Além disso, todo o sistema é baseado na estrutura desta ontologia, em que as opções dos tipos de informações tratadas pelo sistema estão todas contidas na ontologia.

A ontologia VIVO foi construída utilizando diversas outras ontologias e

vocabulários que a torna um ferramental compreensível por diversas outras ferramentas, e pode ser utilizada para outras aplicações, pois é uma representação do domínio de informações acadêmicas, contendo questões relativas a publicações, a docência, a orientação, entre outros. Vale destacar ainda que o uso de outras ontologias e vocabulários facilita a ligação com as bases de dados de *linked data*, por haver uma maior compatibilização entre os recursos com as demais bases (LYRISIS, 2016).

Figura 2 - Diagrama do Módulo Acadêmico da ontologia VIVO



Fonte: Triggs (2016)

A ontologia VIVO é apresentada na Figura 2, na qual é apresentado o módulo acadêmico, que é a parte da ontologia que mais está relacionada com o contexto deste estudo no qual são explorados os dados que constituem uma ecologia de pesquisa. Na figura são apresentadas apenas as classes da ontologia, sem apresentar as propriedades dos dados.

As informações apresentadas na Figura 2 indicam a variedade de classes, como os relacionamentos de editores, documentos e organizações. Além disso, é possível verificar alguns dos vocabulários que são parte desta ontologia, como

o FOAF⁴ e o BIBO⁵.

Como relatado, os dados gerados no âmbito de uma ontologia são estruturadas e com um nível de semântica formal adequado para a publicação nos princípios do *linked data*. Em suma, os dados estarão relacionados e utilizando as classes, as propriedades, os relacionamentos e as restrições que esta ontologia possui.

4.2 ONTOLOGIA CERIF

O CERIF, enquanto um modelo que representa as entidades e os relacionamentos existentes acerca da pesquisa científica, possui uma representação em formato de ontologia seguindo a linguagem OWL. A ontologia CERIF traduz o modelo de dados do CERIF, na semântica e na estrutura de uma ontologia OWL, permitindo a estruturação dos dados seguindo as recomendações da web semântica.

Ao tratar sobre a ontologia CERIF, o documento oficial deste instrumento aponta que o:

[...] objetivo [é] adaptar as informações de pesquisa do CERIF de acordo com o paradigma Linked Data [...]. As entidades e atributos do modelo relacional CERIF estendido foram transformados em axiomas de ontologia; os tipos e funções definidos na Semântica CERIF foram traduzidos em um vocabulário. As entidades de pesquisa do CERIF foram transformadas em classes RDF e seus atributos em propriedades. (euroCRIS, 2013).

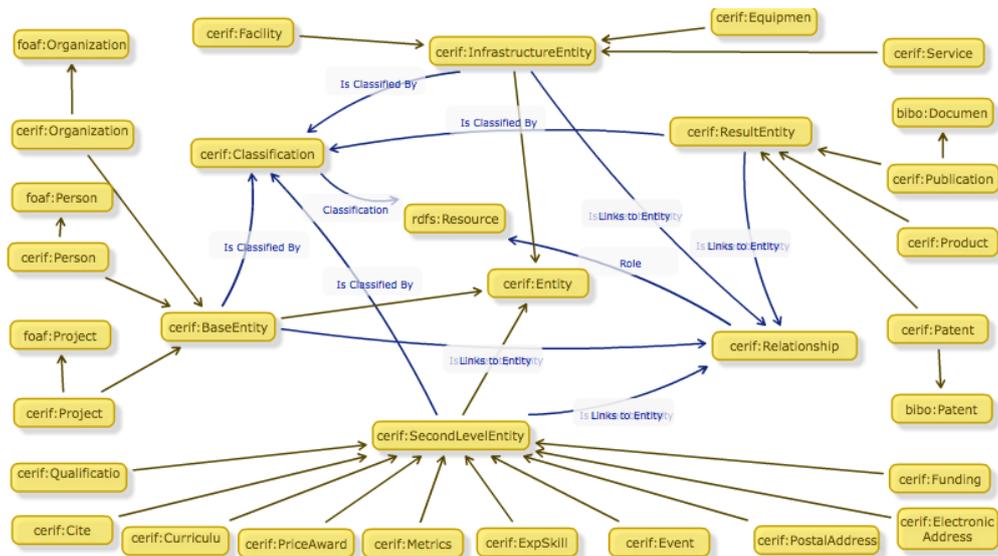
Nessa explanação é possível verificar a ligação entre a ontologia CERIF e os padrões do *linked data*, visando permitir a interoperabilidade com outros vocabulários e modelos. Adicionalmente, aponta-se o processo que levou à criação da ontologia que se deu a partir da adaptação e expansão do modelo CERIF.

A Figura 3 apresenta as principais entidades e relacionamentos que são parte da ontologia CERIF.

⁴ Disponível em: <http://xmlns.com/foaf/spec/>. Acesso em: 17 jul. 2020.

⁵ Disponível em: <https://bibliontology.com/>. Acesso em: 17 jul. 2020.

Figura 3 - Ontologia CERIF



Fonte: euroCRIS (2013)

A partir da Figura 3, é possível verificar que a ontologia CERIF foi concebida a partir de diversos vocabulários, como o FOAF e OWL, mas tendo como centro o CERIF, com as suas diversas entidades.

5 PRINCÍPIOS FAIR

Os princípios FAIR (Figura 4) buscam orientar a implantação e o desenvolvimento de ambientes de dados para sejam *findable* (F), *accessible* (A), *interoperable* (I) e *reusable* (R) - cuja tradução para o português é localizáveis, acessíveis, interoperáveis e reutilizáveis. De acordo com FORCE21 (2020, tradução nossa):

[...] por meio da definição e amplo apoio de um conjunto mínimo de princípios e práticas orientadores acordados pela comunidade, provedores de dados e consumidores de dados - tanto máquinas quanto humanos - poderiam descobrir, acessar, interoperar e reutilizar sensatamente, com a devida citação, as vastas quantidades de informações geradas pela ciência contemporânea com uso intensivo de dados.

Os princípios referem-se aos dados (objetos digitais), aos metadados (representação do objeto digital) e às infraestruturas que mantêm os dados, e, seu alcance favorece a exploração, o compartilhamento e o uso por humanos e aplicações computacionais. Destaca-se ainda que todas as facetas FAIR,

embora relacionadas, são independentes de forma que é possível implementá-las gradativamente.

Figura 4 - Princípios FAIR



Fonte: Australian National Data Service (2020, tradução nossa).

O primeiro princípio 'localizável' (F) tem como premissa o fato de que para (re)usar dados é necessário localizá-los, por isso, tanto os dados quanto os seus metadados devem ser facilmente localizáveis e legíveis por humanos e por aplicações computacionais possibilitando, inclusive, a descoberta e o processamento automático. Para tanto, é necessária a representação exhaustiva, a adoção de identificadores persistentes e a disponibilização em infraestruturas indexadas. Posteriormente, os dados precisam ser acessíveis (A) e por isso é imprescindível fornecer informações de como fazê-lo, o que inclui especificações sobre protocolos, níveis de acesso e autenticação. É possível que os dados necessitem interoperar (I) para fazer parte de fluxos que incluem o armazenamento, a análise e o processamento, assim, precisam ser facilmente legíveis e utilizar formatos de troca de dados, com vocabulário controlado e incluindo sempre que possível referências qualificadas. Para atingir ao objetivo de reuso (R) é imprescindível que os dados estejam descritos de forma exhaustiva, incluindo informações sobre os direitos de uso, a proveniência e o

contexto dos dados (GOFAIR, 2020).

Os princípios FAIR possibilitam identificar pontos de convergência no tratamento dos dados, metadados e infraestruturas visando maximizar sua localização, acesso e uso. Dessa forma, é necessário que todos os envolvidos – desde o planejamento da pesquisa, a coleta, o tratamento e o armazenamento dos dados – tenham em mente a adoção dos princípios, considerando as especificidades dos dados e do domínio.

Esses princípios e práticas simples devem permitir uma ampla gama de comportamentos integrativos e exploratórios e suportar uma ampla gama de escolhas e implementações de tecnologia, assim como o protocolo da Internet (IP) forneceu uma camada mínima [...] que permitiu a criação de uma vasta gama de ferramentas de fornecimento, consumo e visualização de dados na Internet. (FORCE21, 2020, tradução nossa).

Neste sentido, os princípios estão claramente relacionados visto que, por exemplo, o reuso de um conjunto de dados depende claramente de sua descoberta, acesso e possibilidade de processamento.

Assim, clarifica-se a relevância de alguns elementos: modelos de dados e estruturas de representação, atentando-se ao domínio dos dados, uso de vocabulário padronizado e reconhecido, por isso recomenda-se a adoção de padrões internacionalmente aceitos, além disso, a representação deve ser exaustiva e incluir metadados administrativos, descritivos, técnicos, estruturais, de uso, de preservação, de proveniência; uso de identificadores persistentes que favoreçam a descoberta, a citação e o relacionamento entre os dados por meio de referências qualificadas; adoção de protocolos de comunicação de dados abertos e de fácil implementação, visando favorecer o acesso aos dados; uso de formatos legíveis por máquinas e passíveis de preservação, que favoreçam a descoberta e o reuso; informar direitos de uso, visando favorecer a interoperabilidade técnica e as buscas automatizadas.

Considerando os aspectos apontados, o GoFAIR⁶ atua em diferentes países buscando divulgar e nortear a implantação dos princípios FAIR em um processo designado *FAIRification*⁷. O processo é estruturado em três pontos:

⁶ Disponível em: <https://www.go-fair.org/>. Acesso em: 12 jul. 2020.

⁷ Disponível em: <https://www.go-fair.org/how-to-go-fair/>. Acesso em: 12 jul. 2020.

metadados legíveis por máquinas, perfis de implementação FAIR e FAIR *Data Point*.

Nesse sentido, a relevância da adoção de vocabulários controlados legíveis por máquinas é reforçada pelo primeiro processo; o segundo reúne estratégias já utilizadas para a implantação de FAIR, no sentido de auxiliar a implantação; e, o terceiro, o FAIR *data point* se refere às infraestruturas.

FAIR *Data Point* consiste em um repositório que fornece acesso a metadados FAIR, acessíveis por meio de uma *Application Programming Interface* (API) para aplicações computacionais e a humanos, por meio de uma interface gráfica baseada em *web*. Dessa forma, possibilita que os dados e seus metadados, por um lado sejam expostos, recuperados e acessados, e, por outro, para que os dados sejam localizados e consumidos, quer seja por humanos ou aplicações computacionais (FAIR DATA TEAM, 2020). Para tanto, fornece especificações para que desenvolvedores possam tornar suas aplicações FDP e uma aplicação de referência para uso.

6 RESULTADOS E DISCUSSÃO

Visando mapear a abrangência de infraestruturas CRIS, o euroCRIS disponibilizou o *Directory of Research Information Systems* (DRIS)⁸, como parte do 'euroCRIS DSpace CRIS'⁹, um ambiente utilizado para armazenar e disseminar os resultados das atividades de discussão e divulgação da euroCRIS. O DRIS consiste em um catálogo constituído por informações de instituições que disponibilizam CRIS, por meio do preenchimento de um formulário, no qual informam um conjunto mínimo de metadados: *Type* (tipo de CRIS), *Organisation* (instituição responsável pelo CRIS), *Name* (nome do CRIS), *Platform* (*software* utilizado para sustentar o CRIS, quer seja plataforma comercial, livre ou desenvolvida na Instituição), *Status*, URL.

Em uma busca realizada¹⁰ no DRIS verifica-se o registro de 762 CRIS,

⁸ Disponível em: <https://dspacecris.eurocris.org/cris/explore/dris>. Acesso em: 07 jul. 2020.

⁹ Disponível em: <https://dspacecris.eurocris.org/>. Acesso em: 07 jul. 2020.

¹⁰ Realizada em 07 jul. 2020.

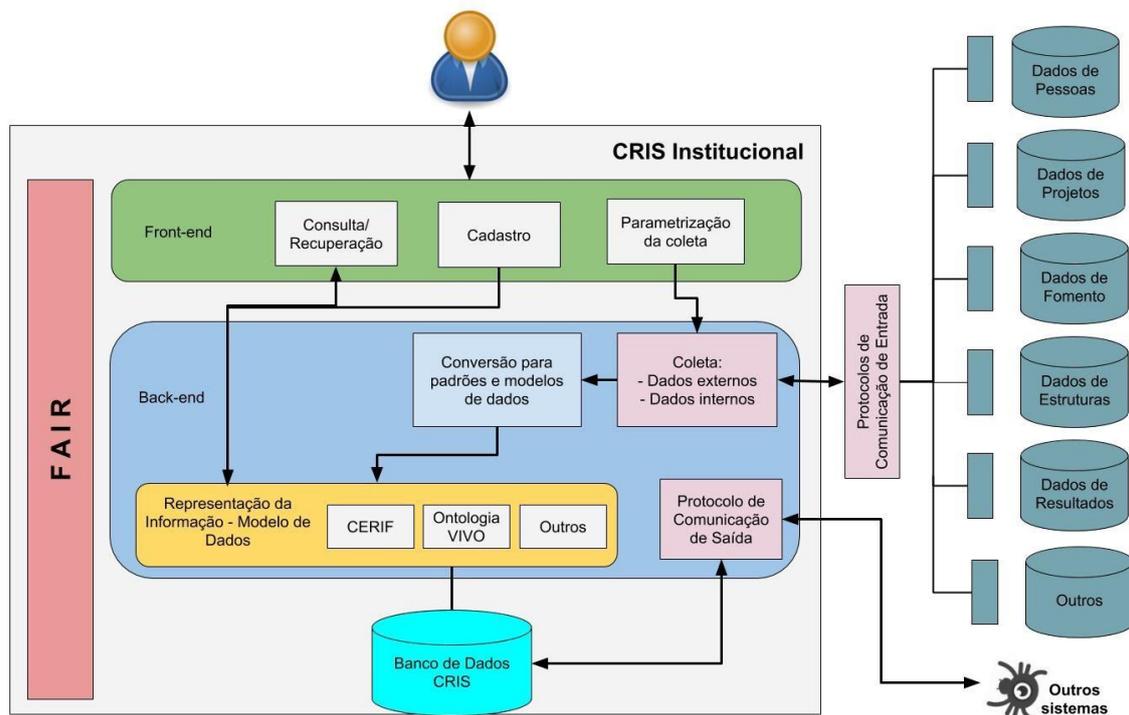
com uma concentração expressiva na Índia (207) e Noruega (132). No Brasil, há o registro de apenas 1 CRIS, pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), a Plataforma Sucupira.

Dessa forma, o presente estudo, a partir da discussão da representação da informação em ecossistema de pesquisa utilizando infraestrutura semântica apresenta um modelo conceitual que possibilita a integração de dados provenientes de sistemas que gerenciam dados e informações do ciclo de vida da pesquisa, notadamente os dados de organização, infraestrutura para pesquisa, projetos de pesquisa, grupos de pesquisa, pesquisador (pessoa), fomento, resultados de pesquisa, de forma a compatibilizá-los para constituir um CRIS, possibilitando o desenvolvimento de sistemas de informação de pesquisa corrente de forma mais massiva em cenário nacional e internacional.

O modelo (Figura 5) se constitui a partir da análise e uso das estruturas de representação da informação existentes em múltiplas plataformas, considerando que esses dados estão armazenados em diferentes sistemas internos e externos à instituição. Nesse sentido, será necessária a compatibilização das estruturas de representação, por meio da conversão para o modelo de dados utilizado, considerando o uso de dados estruturados, não estruturados e ainda padrões de metadados estabelecidos internacionalmente. Para tanto, utilizam-se ontologias e modelos informacionais da web semântica, possibilitando a extração dos dados necessários de múltiplas plataformas, a compatibilização da estrutura e da representação e a recuperação da informação pelos interessados em um ambiente informacional digital constituído como CRIS.

O modelo apresentado na Figura 5 é constituído *de front-end e back-end*, nos quais ocorrem diferentes processos e por meio dos quais o banco de dados é acionado, por humanos, por meio de uma interface gráfica e por agentes computacionais por meio de um 'Protocolo de Comunicação de Saída'.

Figura 5 - Modelo de CRIS Institucional



Fonte: Autoria própria (2020)

A alimentação do CRIS ocorre de forma automática, considerando o especificado na 'Parametrização da Coleta', realizada no *front-end*, que estabelece a forma, a periodicidade e as fontes coletadas, das quais será realizado o reuso de dados. O processamento ocorre no *back-end*, que aciona, por meio de 'Protocolos de Comunicação de Entrada' ou serviços, a obtenção de dados das mais variadas fontes informacionais disponíveis em bancos de dados internos e externos à instituição, dos quais são coletados dados de pessoas (como: nome, identificador persistente, titulação, contato, afiliação institucional, tipo de vínculo, projetos, tipo de contribuição no projeto), de projetos de pesquisas (como: tipo de atividade, *status*, fomento, resultado, função, infraestrutura), de fomento à pesquisa (como: tipo de financiador, identificador persistente, contato, tipo de financiamento, *status*, vínculo com documento), de estruturas para o desenvolvimento de pesquisa (como: tipo, instalações, equipamentos, *status*, políticas, custos) e de resultados de pesquisa (como: publicações, dados de pesquisa, patente, *software* e demais objetos relacionados; *status*, identificador persistente, fonte de publicação). É possível

ainda identificar outros dados necessários para reuso na ecologia de pesquisa e incluí-los à coleta. Os dados coletados são convertidos para o modelo de dados adotado no CRIS antes de serem inseridos no banco de dados, visando assegurar o atendimento aos princípios FAIR.

É possível ainda a alimentação manual, por meio do ‘Cadastro’ disponível no *front-end* que realiza o acesso diretamente ao modelo de dados do sistema, e tem acesso direto ao banco de dados do CRIS.

Na camada de representação da informação, na qual está contido o modelo de dados são apresentadas algumas ontologias que podem ser utilizadas, como a ontologia VIVO e o CERIF, mas não estando restrita a essas. Vale destacar a relevância da adoção de vocabulários controlados, visando assegurar a interoperabilidade. Nesse sentido, recomenda-se atenção às especificações do *Data Catalog Vocabulary* (DCAT)¹¹.

O modelo possibilita ainda a “Consulta/Recuperação” por humanos via interface gráfica, por meio do *front-end*, que passa pela camada de Representação da Informação e acessa o modelo de dados do sistema e, como consequência, o banco de dados do CRIS. E, por agentes computacionais por meio do ‘Protocolo de Comunicação de Saída’, que permite a coleta de dados por terceiros atendendo aos princípios FAIR.

Além disso, em uma infraestrutura CRIS institucional como a proposta neste estudo, considerando que o entorno da pesquisa esteja baseado em dados, informações e objetos digitais, recomenda-se que sejam armazenados apenas os metadados e que os objetos sejam sempre acessíveis a partir da sua fonte original de publicação, preferencialmente utilizando identificadores persistentes. Assim, seriam mantidos no CRIS institucional apenas os objetos que não estejam acessíveis em meio digital, ao qual deve ser atribuído um identificador persistente.

No modelo apresentado, a camada de ‘Representação da Informação’, na qual está contido o modelo de dados, que deve beneficiar os princípios FAIR, e assume papel relevante no seu atendimento, por conter vocabulário controlado, metadados e identificadores persistentes, presentes em F, A, I, R. Além dela, no

¹¹ Disponível em: <https://www.w3.org/TR/vocab-dcat/>. Acesso em: 16 jul. 2020.

back-end localiza-se o 'Protocolo de Comunicação de Saída' que correspondem igualmente à F, A, I, R. No *front-end*, a possibilidade de consulta e recuperação, por meio da camada de representação da informação e do modelo de dados, beneficia o acesso aos dados pelo atendimento aos princípios FAIR. Destaca-se a relevância de tornar a infraestrutura, os metadados e dados FAIR, o que pode ocorrer de forma gradativa.

É perceptível a necessidade de otimizar a infraestrutura necessária para a gestão de informações de pesquisa visando torná-las acessíveis e reutilizáveis, assegurando ganhos para todos os envolvidos. Nesse sentido, o modelo de CRIS Institucional proposto pode torná-lo um FAIR *Data Point*.

O desafio de integrar os sistemas de informação relacionados à atividade de pesquisa compete a diferentes áreas, dentre elas a Ciência da Informação, para a qual o desafio apontado por Sales e Sayão (2015, p. 166) consiste em “como explorar as diferenças e sobreposições desses sistemas, sincronizando os seus processos em configurações e arquiteturas que resultem em serviços mais avançados e mais próximos aos pressupostos de uma ciência mais aberta e colaborativa”. E completam indicando a necessidade de estabelecer níveis satisfatórios de interoperabilidade diante de uma diversidade de repositórios e modelos de dados (SAYÃO; SALES, 2016, p. 112).

Em um relatório elaborado pela *Online Computer Library Center (OCLC)* que buscou discutir o gerenciamento de informações de pesquisa e o papel das bibliotecas afirma que:

Tecnologias avançadas, padrões e informações em rede oferecem maneiras novas e aprimoradas de coletar e gerenciar informações de pesquisa, e a interoperabilidade está se tornando um imperativo para instituições, bibliotecas, editoras e financiadores que buscam coletar e trocar pesquisas informações, reduzir a carga administrativa, apoiar a tomada de decisão institucional e promover marca institucional e prestígio (BRYANT; CLEMENTS; FELTES *et al.*, 2017).

Tal afirmação corrobora com a proposta deste estudo, de compatibilizar as estruturas de representação disponíveis em múltiplos sistemas de informação, e utilizar tecnologias já existentes visando constituir um CRIS institucional. Para tanto, é necessário que a atuação esteja embasada em conhecimentos de arquitetura da informação, metadados, padrões e protocolos

de comunicação, interoperabilidade e tecnologias da *web* semântica, conforme modelo proposto no presente estudo.

7 CONSIDERAÇÕES FINAIS

No contexto da Ciência da Informação, nas últimas décadas há atenção especial aos ambientes informacionais que gerenciam resultados de pesquisa, notadamente os periódicos científicos, os repositórios digitais e, mais recentemente, os repositórios de dados de pesquisa.

No entanto, para compreendermos a pesquisa a partir de uma ecologia é necessário considerar que o ciclo de vida da pesquisa é constituído por um conjunto de dados e informações que estão dispersas em diferentes ambientes que gerenciam dados e informações de parte de ciclo. A exemplo dos dados de pessoas, projetos, fomento, estruturas de pesquisa, organizações e os resultados de pesquisa.

Adicionalmente, a criação de novas tecnologias de informação e comunicação está beneficiando a ampliação das tipologias documentárias que favorecem a divulgação científica. Há *softwares*, patentes, blogs, *podcasts*, entre outros artefatos, que podem se favorecer de dados vinculados à ecologia de pesquisa, e que, a partir de uma representação e gestão adequada dos dados, podem facilitar a sua recuperação e o seu uso. Nesse contexto, toda a ecologia pode ser beneficiada ao realizar uma representação da informação adequada para a qual o profissional da informação está capacitado.

No contexto das ecologias de pesquisa, o gerenciamento dos dados que a compõem é complexo, considerando sobretudo a diversidade e dispersão das fontes de informação disponíveis, as estruturas de dados utilizadas e a dificuldade de localização, acesso, interoperabilidade e reuso.

Dessa forma, o presente estudo apresentou um modelo conceitual para a constituição de CRIS institucional a partir do reuso de dados de múltiplas fontes, da conversão de dados por meio de infraestrutura semântica realizada na camada de representação da informação, cujas bases são as ontologias, em especial a ontologia VIVO e o CERIF, que permitem um enriquecimento semântico ao modelo proposto. O referido modelo possibilita que as instituições

nacionais e internacionais que estejam atentas à necessidade de gestão das informações de pesquisa tenham a possibilidade de utilizá-lo como referência para a implantação de um CRIS, apoiado em princípios FAIR e com a possibilidade de se tornar um FAIR *Data Point*.

Destaca-se que a implantação de um CRIS, seguindo o modelo proposto, deve ocorrer a partir da definição de um modelo de dados pautados nas ontologias e padrões de metadados existentes, protocolos utilizados para a comunicação entre diferentes sistemas e em tecnologias que permitam a construção de uma arquitetura baseada no modelo criado.

Há, portanto, uma interdisciplinaridade entre as áreas da Ciência da Informação e da Ciência da Computação nesta pesquisa, ao utilizar elementos e teorias que se vinculam a primeira, bem como a aplicação de práticas e técnicas que estão, nativamente, presentes na segunda.

Enquanto trabalhos futuros, busca-se explorar a relação entre o modelo proposto e um FAIR *Data Point* visando tornar o modelo uma referência de utilização deste tipo de estrutura. Adicionalmente, busca-se criar uma implementação desse modelo, definindo um modelo de dados, bases vinculadas e protocolos de saída, permitindo que as instituições possam implantar essa solução.

REFERÊNCIAS

ASSERSON, .; JEFFERY, K. G. Current research information systems (CRIS): past, present and future. **Wissenschaftsmanagement**, Bergen, n. 1, p. 41-45, jan./feb. 2009. Disponível em: <http://bora.uib.no/handle/1956/6929>. Acesso em: 10 ago. 2017.

AUSTRALIAN NATIONAL DATA SERVICE. **FAIR data training**. Disponível em: <https://www.ands.org.au/working-with-data/fairdata/training>. Acesso em: 09 jul. 2020.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific American**, New York, v. 284, n. 5, p. 28-37, 2001.

BRANDT, M. B. **Modelagem da informação legislativa**: arquitetura da informação para o processo legislativo brasileiro. 2020. 266 f. Tese. (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Marília, 2020. Disponível em: <http://hdl.handle.net/11449/191740>. Acesso em: 09 jul. 2020.

BRYANT, R.; CLEMENTS, A.; FELTES, C.; GROENEWEGEN, D.; HUGGARD, S.; MERCER, H.; MISSINGHAM, R.; OXNAM, M.; RAUH, A.; WRIGHT, J. **Research Information Management: defining RIM and the library's role.** Dublin, OH: OCLC Research, 2017. Disponível em: <https://doi.org/10.25333/C3NK88>. Acesso em 07 jul. 2020.

CAMARGO, L. S. A.; VIDOTTI, S. A. B. G. **Arquitetura da informação: uma abordagem prática para o tratamento de conteúdo e interface em ambientes informacionais digitais.** Rio de Janeiro: LTC, 2011.

CAMPOS, L. M.; CAMPOS, M. L. A. Aplicação de dados interligados abertos apoiada por ontologia. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 15. 2014, Belo Horizonte. **Anais [...]** Belo Horizonte, MG: ANCIB, 2014. Disponível em: <http://enancib2014.eci.ufmg.br/documentos/anais/anais-gt8>. Acesso em: 2 fev. 2016.

CATARINO, M. E.; SOUZA, T. B. A representação descritiva no contexto da web semântica. **Transinformação**, Campinas, v. 24, n. 2, p. 77-90, 2012. Disponível em: https://www.scielo.br/scielo.php?pid=S0103-37862012000200001&script=sci_abstract&tlng=pt. Acesso em: 17 jul. 2020.

DAVENPORT, T. H. **Ecologia da informação: por que só a tecnologia não basta para o sucesso na era da informação.** São Paulo: Futura, 1998.

euroCRIS. **CERIF in brief.** Disponível em: https://www.eurocris.org/eurocris_archive/cerifsupport.org/cerif-in-brief/index.html. Acesso em: 04 jul. 2020a.

euroCRIS. **CERIF refactoring project: an introduction.** Disponível em: <https://eurocris.org/cerif-refactoring-project-introduction>. Acesso em: 04 jul. 2020b.

euroCRIS. **CERIF support blog.** Disponível em: https://www.eurocris.org/eurocris_archive/cerifsupport.org/. Acesso em: 04 jul. 2020c.

euroCRIS. **The Common European Research Information Format Ontology: CERIF Ontology 0.2.** 2013. Disponível em: <https://www.eurocris.org/ontologies/cerif/1.3/index.html>. Acesso em: 12 jul. 2020.

euroCRIS. **Services.** Disponível em: <https://eurocris.org/services>. Acesso em: 04 jul. 2020d.

FAIRDataTeam. **FAIRDataPoint-Spec.** Disponível em: <https://github.com/FAIRDataTeam/FAIRDataPoint-Spec>. Acesso em: 12 jul. 2020.

FORCE21. Guiding principles for findable, accessible, interoperable and re-usable data publishing version b1.0. Disponível em:

<https://www.force11.org/fairprinciples#Annex1-1>. Acesso em: 07 jul. 2020.

GOFAIR. FAIR principles. Disponível em: <https://www.go-fair.org/fair-principles/>. Acesso em: 07 jul. 2020.

JEFFERY, K. G.; ASSERSON, A. CERIF-CRIS for the european e-infrastructure. **Data Science Journal**, v. 9, 2010. Disponível em: <https://datascience.codata.org/articles/abstract/10.2481/dsj.CRIS3/>. Acesso em: 04 jul. 2017.

JOINT, N. Current research information systems, open access repositories and libraries: ANTAEUS. **Library Review**, v. 57, n. 8, p. 570-575, 2008. Disponível em: <https://doi.org/10.1108/00242530810899559>. Acesso em: 19 set. 2017.

LYRASIS. **VIVO-ISF Ontology**. 2016. Disponível em: <https://wiki.lyrasis.org/display/VTDA/VIVO-ISF+Ontology>. Acesso em: 17 jul. 2020.

LOPATENKO, A. S. **Information retrieval in Current Research Information Systems**. 2001. Disponível em: <https://arxiv.org/abs/cs/0110026>. Acesso em: 22 jun. 2017.

MORVILLE, P.; ROSENFELD, L.. **Information architecture for the world wide web**. Sebastopol, CA: O'Reilly, 2006.

OLIVEIRA, H. P. C. **Arquitetura da informação pervasiva: contribuições conceituais**. 2014. 202 f. Tese (Doutorado em Ciência da Informação) - Universidade Estadual Paulista, Marília. Disponível em: <http://hdl.handle.net/11449/110387>. Acesso em: 01 set. 2017.

SALES, L. F.; SAYÃO, L. F. Ciberinfraestrutura de informação para a pesquisa: uma proposta de arquitetura para a integração de repositórios e sistemas CRIS. **Inf. & Soc.: Est.**, João Pessoa, v. 25, n. 3, p. 163-184, set./dez. 2015. Disponível em: <http://www.ies.ufpb.br/ojs/index.php/ies/article/view/23998/14535>. Acesso em 11 ago. 2017.

SANTAREM SEGUNDO, J. E. **Representação iterativa: um modelo para Repositórios Digitais**. 2010. 224 f. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Marília, 2010. Disponível em: <https://orcid.org/0000-0002-4216-0374>. Acesso em: 17 jul. 2020.

SANTAREM SEGUNDO, J. E.; CONEGLIAN, C. S. Tecnologias da web semântica aplicadas a organização do conhecimento: padrão SKOS para construção e uso de vocabulários controlados descentralizados. *In*: GUIMARÃES, J. A. C.; DODEBEI, V. (Org.). **Organização do conhecimento e diversidade cultural**. Marília: Fundepe, 2015. p. 224-233.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Inf. Inf.**, Londrina, v. 21, n. 2, p. 90-115, maio/ago., 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939/20122>. Acesso em: 10 set. 2017.

SHEPPARD, N. Learning how to play nicely: repositories and CRIS. **Ariadne**, n. 64, jul. 2010. Disponível em: <http://www.ariadne.ac.uk/issue64/wrn-repos-2010-05-rpt>. Acesso em: 01 ago. 2017.

TRIGGS, G. **VIVO-ISF Academic Module**. 2016. Disponível em: <https://wiki.lyrasis.org/display/VTDA/VIVO-ISF+Academic+Module>. Acesso em: 17 jul. 2020.

VECHIATO, L. F.; OLIVEIRA, H. P. C.; VIDOTTI, S. A. B. G. Arquitetura da informação pervasiva e encontrabilidade da informação: instrumento para a avaliação de ambientes informacionais híbridos. **Informação & Tecnologia**, v. 3, n. 1, p. 47-65. 2016. Disponível em: <https://periodicos.ufpb.br/index.php/itec/article/view/38365>. Acesso em: 29 jun. 2020.

REPRESENTATION STRUCTURES FOR DATA REUSE IN THE CONTEXT OF RESEARCH ECOLOGY: INSTITUTIONAL CRIS

ABSTRACT

Introduction: The Current Research Information System (CRIS), emerges as a response to the need for optimization in the management process of the large volume of data and information that permeate the life cycle of research. Such information is dispersed in multiple informational environments, of complementary characteristics, today isolated that can be integrated, and provide a broader panorama of research. **Objective:** In this context, the present study discusses the possibilities of data integration of different systems that store data related to research ecology, using representation structures and semantic infrastructure to compose a research ecology, constituted as an institutional CRIS. **Methodology:** As a methodological procedure, it uses the bibliographic review for the theoretical-conceptual basis of the study, aiming at contextualizing the research ecology infrastructure, especially the structures of representation, ontologies and FAIR principles. **Results:** The results are presented as a conceptual model of data integration of a research ecology, capable of supporting the composition of an institutional CRIS. **Conclusions:** The compatibility of representation structures available in multiple information systems and the use of existing technologies aiming to constitute an institutional CRIS, requires that the performance of the information professional is based on knowledge of information architecture, metadata, communication standards and protocols, interoperability and semantic web technologies, according to the model proposed in this study.

Descriptors: Representation of Information. Data Reuse. CRIS. Ontology. FAIR Principles.

ESTRUCTURAS DE REPRESENTACIÓN PARA LA REUTILIZACIÓN DE DATOS EN EL CONTEXTO DE LA ECOLOGÍA DE INVESTIGACIÓN: CRIS INSTITUCIONAL

RESUMEN

Introducción: El *Current Research Information System* (CRIS), surge como respuesta a la necesidad de optimizar el proceso de gestión del gran volumen de datos e informaciones que permean el ciclo de vida de las investigaciones. Esas informaciones están dispersas en múltiples ambientes informacionales, de características complementarias, hoy en día aisladas que pueden ser integradas, y proporcionar un panorama más amplio de la investigación. **Objetivo:** En este contexto, el presente estudio examina las posibilidades de integrar datos de diferentes sistemas que almacenan datos relacionados con la ecología de investigación, utilizando estructuras de representación e infraestructura semántica para componer una ecología de investigación, constituida como un CRIS institucional. **Metodología:** Como procedimiento metodológico, utiliza la revisión bibliográfica para la base teórico-conceptual del estudio, con el fin de contextualizar la infraestructura de la ecología de la investigación, especialmente las estructuras de representación, las ontologías y los principios FAIR. **Resultados:** Los resultados se presentan como un modelo conceptual de integración de datos de una ecología de investigación, capaz de apoyar la composición de un CRIS institucional. **Conclusiones:** La compatibilidad de las estructuras de representación disponibles en múltiples sistemas de información y el uso de las tecnologías existentes con el objetivo de constituir un CRIS institucional, requiere que la actuación del profesional de la información se base en el conocimiento de la arquitectura de la información, los metadatos, las normas y protocolos de comunicación, la interoperabilidad y las tecnologías de la web semántica, según el modelo propuesto en este estudio.

Descriptores: Representación de la Información. Reutilización de Datos. CRIS. Ontología. Principios FAIR.

Recebido em: 15/09/2020

Aceito em: 30/09/2020