

**ANÁLISE TEMÁTICA DOS METADADOS ASSUNTO NO REPOSITÓRIO INSTITUCIONAL DA
UNIVERSIDADE FEDERAL DE ALAGOAS**

***THEMATIC ANALYSIS OF SUBJECT METADATA IN THE INSTITUTIONAL REPOSITORY OF THE
FEDERAL UNIVERSITY OF ALAGOAS***

Valter dos Santos Andrade - Universidade Federal de Alagoas (UFAL) -
valter.andrade@sibi.ufal.br

Robéria de Lourdes de Vasconcelos Andrade – Universidade Federal de Alagoas (UFAL) -
roberia.andrade@ichca.ufal.br

Resumo: Analisa a representação temática dos metadados assuntos presentes nas monografia depositadas através do autoarquivamento pelos alunos do curso de Enfermagem na base de dados do Repositório Institucional da Universidade Federal de Alagoas. O percurso metodológico dar-se-á por meio de pesquisa básica, exploratório de características descrita com abordagem qualitativa e quantitativa. Foram mapeadas 20 monografias totalizando 78 metadados assuntos coletados. Destes 68 apresentaram conformidade com o do catálogo de assunto do Sistema de Biblioteca e apenas 9 não. O alto índice de compatibilidade deve-se pelo fato dos alunos já utilizam a Base de Dados de Descritores de Saúde e também incluem os cabeçalhos de assuntos constantes na ficha catalográfica de seus trabalhos, demonstrando com isso a importância do uso de termos autorizados.

Palavras-chave: repositório; autoarquivamento; representação da informação.

Abstract: *It analyzes the thematic representation of subject metadata present in the monographs deposited through self-archiving by Nursing course students in the Institutional Repository database of the Federal University of Alagoas. The methodological route will take place through basic research, exploratory of characteristics described with a qualitative and quantitative approach. 20 monographs were mapped, totaling 78 collected subject metadata. Of these 68 were compliant with the Library System subject catalog and only 9 were not. The high compatibility rate is due to the fact that students already use the Health Descriptors Database and also include the headings of constant subjects in the catalog of their works, thus demonstrating the importance of using authorized terms.*

Keywords: *repository; self-archiving; information representation.*

1 INTRODUÇÃO

O desenvolvimento das tecnologias de comunicação e informação (TICs), possibilitaram a grande produção e disponibilização de conteúdos informacionais em uma escala de proporções gigantescas nesse novo ambiente informacional, a Internet, sendo o

acesso a esses conteúdos de forma instantânea e ilimitado. Nesse novo ambiente, o acesso à informação tem-se apresentado como um grande desafio da Ciência da Informação (CI), principalmente quando se trata da informação científica, as quais são em sua maioria oriundas das universidades. Para as universidades, os desafios consistem em como organizar, preservar e principalmente disseminá-las.

Dentre as novidades desse novo ambiente informacional surgem os repositórios institucionais (RIs). Oriundo do movimento da Ciência Aberta, os RIs foram concebidos com o intuito de disponibilizarem as produções científicas em acesso aberto, sem alguma barreira ou entraves. Leite (2009, p. 21) define RI em “Um serviço de informação científica - em ambiente digital e interoperável - dedicado ao gerenciamento da produção intelectual de uma instituição”. O RI é um órgão de mediação da informação produzida pela instituição e seus usuários, é através dele que os usuários terão acessos a produção intelectual da instituição, suas descobertas científicas e inovações tecnológicas. Para cumprir verdadeiramente a sua função, o RI precisa prover a interoperabilidade e o autoarquivamento. Segundo Villalobos e Gomes,(2015, p. 4), a interoperabilidade consiste na “comunicação entre diferentes sistemas de informação”, disponibilizando os metadados os quais identificam, descrevem e localizam os recursos informacionais armazenados no RI. Sendo assim, ela possibilita a interação do sistema com o usuário, permite que tenha acesso aos dados, podendo compartilhá-los com outros sistemas de diferentes hardwares e softwares. Autoarquivamento vem se tornando uma das etapa operacional fundamental, ele possibilita o autor a realizar o depósito do seu documento diretamente na base de dados. Além disso, esse procedimento agiliza o processo de visibilidade e disponibilidade da publicação. No entanto, esse procedimento pode interferir na qualidade e principalmente na confiabilidade do sistema. Para Freitas (2019), sobre a questão do autoarquivamento, esse procedimento pode comprometer a qualidade dos descritores de assunto, dificultando no processo de busca da informação.

Na prática, o autoarquivamento é um processo que permite maior rapidez na divulgação da pesquisa científica, mas por outro lado, isso não garante a qualidade da representação temática, situação que pode repercutir na recuperação da informação.(FREITAS, 2019, p. 15).

Essa preocupação é pertinente visto que, o autor do trabalho depositado pode não ter utilizados padrões metodológicos para escolha dos metadados assunto, além de uso inadequado de alguns termos. É preciso que os RIs garantam a descrição de documentos com metadados padronizados. Segundo Sanchez, Silva e Vechiato (2018, p.48) “[...] a descrição dos objetos digitais e dos dados que compõem a pesquisa são determinantes para a recuperação e encontrabilidade de cada um deles”.

Neste aspecto, faz-se necessário uma análise criteriosa da forma da representação temática da informação nos ambientes digitais, principalmente nos RIs. A representação temática consiste na análise do conteúdo de um documento, ou seja, o tema ou assunto abordado por ele, objetivando a sua disseminação. Para Albuquerque, Gaudêncio e Santos (2019, p. 19), a representação da informação “Consiste em uma tarefa indispensável para a organização, recuperação, o acesso e uso dos registros do conhecimento. Para tanto, são criados processos, produtos e instrumentos que corroborem para essa finalidade”. Dentre os instrumentos estão os tesouros, listas de cabeçalhos de assuntos e glossários, os quais podem ser aperfeiçoados, atualizados e implantados nas estruturas dos sistemas de recuperação da informação na web. Esses instrumentos auxiliam na padronização dos metadados assuntos.

Diante desta concepção, o metadado assunto, apresenta-se como campo de busca fundamental para acesso e recuperação da informação, pois é através dele que na maioria das vezes, começa a pesquisa de um determinado documento.

O Repositório Institucional da Universidade Federal de Alagoas (RIUFAL), buscando intensificar a interoperabilidade entre os usuários do repositório, começou a testar de modo piloto a partir do ano de 2020, a utilização da prática do autoarquivamento nos cursos de Enfermagem e Serviço Social.

Neste sentido, a presente pesquisa levanta o seguinte questionamento: os metadados assuntos cadastrados pelos usuários no RIUFAL, no momento do depósito de seus trabalhos foram qualificados e validados tendo como referencial algum cabeçalho autorizado por tesouros especializados e se alguns desses metadados assuntos cadastrados apresentam algum tipo de ruído que possam prejudicar ao acesso aos documentos na base de dados do RIUFAL?

Nesse contexto, busca analisar a representação temática nos metadados assuntos contidos nos recursos informacionais depositados no (RIUFAL).

Diante do exposto, o objetivo geral da presente pesquisa é analisar a representação da informação no RIUFAL, no que concerne ao metadado assunto.

Tendo com objetivos específicos: a) Mapear os trabalhos de conclusão de curso; b) Avaliar os metadado assunto cadastrados através do autoarquivamento dos trabalhos de conclusão de curso no RIUFAL; c) Comparar os metadados assuntos utilizados com cabeçalhos autorizados do Catálogo de Assuntos do Sistema de Bibliotecas da Universidade Federal de Alagoas (SIBI/UFAL); d) Identificar se dentre os metadados assuntos cadastrados se existem algum que apresente ruídos que possa interferir no processo da busca e recuperação da informação no sistema.

Promover estudos que viabilizem a melhoria do acesso as informações presentes no RI é de suma importância para contribuir para a visibilidade e reconhecimento local, nacional e internacional da instituição.

2 REPOSITÓRIO INSTITUCIONAL

Tradicionalmente, as universidades têm sido reconhecidas como espaços de produção e transferência de conhecimento científico. O acesso à informação tem-se apresentado como um grande desafio, principalmente quando se trata da informação científica. O alto preço imposto as publicações de artigos científicos em periódicos renomados passou a constituir-se como uma grande barreira no processo de divulgação e acesso à informação científica.

As implicações dessa prática forçada pelos editores geraram a denominada crise dos periódicos científicos. A crise inicia-se nos Estados Unidos e afetou todo mundo, no Brasil ela fica evidente na década de 90, seus reflexos perdurou ainda por várias décadas.

Conforme Alves (2008, p. 128), buscando superar essa crise, pesquisadores de diversas partes do mundo se reuniram e deram início a um grandioso movimento em direção ao acesso aberto à informação científica a nível global. O movimento foi denominado de Movimento de Acesso Aberto à Informação Científica. *Open Access*(OA) - Acesso Aberto.

Esse movimento preconiza que a informação científica seja de livre acesso, disponibilizada na internet sem nenhuma restrição e principalmente se essa informação for oriunda de financiamento público. Conforme Santos Junior (2010, p. 50), Stevan Harnad, um dos principais pesquisadores responsáveis pelo movimento de livre acesso, junto com outros estudiosos, desenvolvem duas estratégias de ação para tornar-se um documento livre de barreiras de acesso, a qual ele denominou de vias: a via dourada (*goldenroad*) e a via verde (*greenroad*). A via dourada refere-se à produção e ampla disseminação de periódicos eletrônicos de acesso aberto em rede, já a via verde refere-se à criação de repositórios institucionais (RIs).

Os RIs são ferramentas adequadas para a gestão do conhecimento científico, potencializam a condução de processos que maximizam a criação, o comportamento, a disseminação e o uso do conhecimento científico, bem como, agilizam os processos de comunicação científica.

Os RIs têm-se tornado uma alternativa relevante para as instituições de pesquisa e ensino reunirem a produção científica em um único local, potencializando a disseminação e o acesso à informação científica. Um dos pré-requisitos para que os RI cumpram esse papel é que eles garantam a descrição de documentos com metadados padronizados (GOMES, 2016).

É nesse panorama que surgem os RIs, com a finalidade de agilizar o processo de comunicação científica quanto para a gestão do conhecimento científico.

Com a “Via verde”, o Movimento de Acesso Aberto, estabelece a estratégia do autoarquivamento, o qual tem como objetivo que os autores disponibilizem diretamente, as suas produções científicas em repositórios de acesso aberto.

Para Assis (2013, p. 212),

O autoarquivamento é o mecanismo de depósito em repositórios institucionais recomendado pelo movimento de acesso aberto à informação científica, seguindo as estratégias da Via Verde. O autoarquivamento é aquele realizado pelo próprio autor. A adoção do autoarquivamento por parte das instituições tem como objetivo garantir os depósitos nos repositórios institucionais, maximizando assim, a sua quantidade de documentos (ASSIS, 2013, p. 212).

O movimento de acesso aberto à informação científica preconiza que os resultados das pesquisas financiadas com dinheiro público devem ser de acesso aberto, justificando o

uso desses recursos, e que essas pesquisas estejam disponíveis gratuitamente na web.

Conforme enfatizam Freitas, Dal'Evedove e Tartarotti:

O avanço das tecnologias digitais culminou em uma rápida e crescente disponibilização de conteúdos na internet. Esta evolução tecnológica alterou o processo de comunicação científica, com destaque para o movimento mundial em favor do acesso livre e gratuito ao conhecimento gerado no ambiente das universidades e institutos de pesquisa, promovendo benefícios significativos ao desenvolvimento da ciência a partir do acesso livre de barreiras aos resultados de pesquisas conduzidas mediante a concessão de recursos públicos. (FREITAS; DAL'EVEDOVE; TARTAROTTI, 2021, p. 169).

Esse entendimento parte do princípio de que tudo que é financiado com recurso público, deverá voltar para a sociedade de forma gratuita, contribuindo para o desenvolvimento de todos.

O autoarquivamento nos RIs, apresenta-se como a alternativa viável para os pesquisadores confrontarem com a exclusividade impostas pelas grandes editoras e plataformas científicas em dar visibilidade as suas produções, to torna-se relevante também pelo fato de que ele pode agilizar o processo de visibilidade, pois possibilita a inclusão imediata no repositório, não sujeito à morosidade que as vezes ocorre devido ao acúmulo de trabalhos executados pela equipe do repositório, que geralmente é bastante reduzida. A prática do autoarquivamento também aproxima a interação entre o autor, a instituição e um universo de usuários que utilizam o ambiente da Web para produção e compartilhamento de pesquisas científicas.

3 REPRESENTAÇÃO DA INFORMAÇÃO E A INDEXAÇÃO

A representação da informação destacam-se como parte fundamental na tarefa de garantir a recuperação de um item informacional, principalmente quando se trata do meio digital. Os meios digitais proporcionam a facilidade para as pessoas publicarem seus conteúdos informacionais em uma mutiplicidades de formatos, ocasionando ambiguidades e heterogeneidade. Esses fatores dificultam o acesso e a recuperação das informação em ambientes digitais.

Marcondes (2001, p. 61) “De nada adianta a informação existir, se quem dela necessita não sabe da sua existência, ou se ela não puder ser encontrada”. A informação precisa está bem representada para ser recuperada pelos usuários que buscam para

satisfazerem as suas necessidades informacionais. A representação da informação apresenta-se de duas formas: representação descritiva e representação temática. A primeira referir-se às características específicas do documento, ou seja, descrição bibliográfica, e permiti a individualização do mesmo e a segunda, pela representação dos assuntos dos documentos, com o objetivo de tornar mais fácil a sua recuperação. Para Ferreira e Albuquerque (2013) a representação da informação busca atender prontamente a todos os usuários, disponibilizando-os de instrumentos que possam descrever as informações da forma mais clara possível, garantindo eficiência e eficácia na realização das pesquisas.

Nessa visão, a indexação torna-se ponto essencial para a representação da informação, parte determinante para o sucesso ou insucesso em uma busca. Daí a importância na escolha dos metadados assuntos que identificam um determinado item, pois eles possibilitam a sua mensuração quanto a pertinência e relevância para uma determinada pesquisa, principalmente nos ambientes digitais, onde encontra-se um grande volume de informação.

Lancaster (2004, p.9) considera a indexação em duas etapas: análise conceitual e tradução. A análise conceitual, refere-se à extração de conceitos que representem o conteúdo de um documento, “implica decidir do que trata um documento – isto é, qual o seu assunto,” enquanto na tradução consiste na representação dos documentos em termos oriundos de instrumentos de indexação, “envolve a conversão da análise conceitual de um documento num determinado conjunto de termos de indexação”. Segundo Dias e Naves (2007, p. 6), esses instrumentos de indexação são chamados de linguagens de indexação ou linguagens documentárias e “depois de traduzidos, para uma linguagem de indexação, passam a ser chamados de descritores de assunto, rótulos, cabeçalhos de assunto, termos de indexação ou enunciados”.

No que concerne sobre a tradução, Lancaster (2004), a tipifica em indexação por extração ou indexação derivada e indexação por atribuição.

No primeiro caso, as palavras ou expressões empregadas pelo autor no item documental são selecionadas pelo indexador para representar seu conteúdo. Ao utilizar-se desse recurso, o indexador faz uso do que é chamado de uma linguagem natural, ele seleciona palavras encontradas no título, nas palavras-chave ou no próprio conteúdo

informacional do documento. No segundo caso, indexação por atribuição, como o próprio nome diz, é atribuído descritores ao documento utilizando-se uma linguagem oriundas de fontes externa, não necessariamente o próprio documento, sendo assim, o indexador utilizar-se de vocabulários controlados, tesouros, entre outros, fazendo uso de uma linguagem controlada e/ou artificial, ou seja, buscando uma padronização.

Apesar que no ambiente web a linguagem natural ocorra com maior intensidade, devido a facilidade e liberdade das escolhas dos termos, faz-se necessário a busca de formas de padronização para melhor aproveitamento entre a linguagem dos usuários e a linguagem dos sistemas, e isso é possível através do uso dos vocabulários controlados.

4 PROCESSO METODOLÓGICO

Quanto aos objetivos a serem alcançados para este estudo, foi empreendida uma pesquisa básica do tipo exploratória de características de uma pesquisa descrita, com abordagem qualitativa referentes à representação da informação no metadado assunto das monografias depositadas através do autodepósito pelos do cursos de Enfermagem e Serviço Social no RIUFAL. Para tanto, a análise qualitativa do metadado em questão foi realizada mediante a verificação das palavras-chave empregadas pelos autores das monografias. Até o mês de julho de 2022 o RIUFAL apresentava 130 monografias depositadas no sistema. entretanto, devida a pandemia de COVID-19, e principalmente por problema estrutural na rede de fibra ótica da UFAL, afetando toda parte de computação, o repositório ficou sem acesso praticamente todo ano de 2021, voltando ao acesso a partir do segundo semestre de 2022. Por esse motivo, foi priorizado a trabalhar com os dados do ano de 2020, os quais totalizaram 40 monografias, sendo 20 do curso de Enfermagem e 20 do curso de Serviço. Tais trabalhos constituem o corpus de documentos analisados na presente pesquisa.

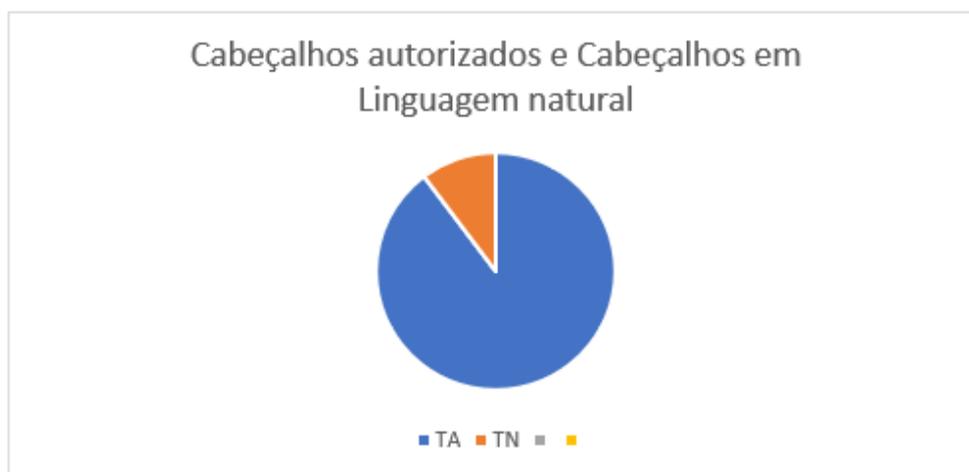
5 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Para a coleta dos dados foi realizado uma busca diretamente na base de dados do RIUFAL, onde foi feita a seleção de 40 monografias autoarquivadas pelos discentes concluintes dos curso de Enfermagem e Serviço Social, os quais serão coletados e analisados os cabeçalhos de assuntos. A partir dos dados levantados será elaborada uma planilha no

Excel com os dados de: títulos e cabeçalhos e posteriormente apresentados em quadros. Das 20 monografias do curso de Enfermagem foram coletados 78 cabeçalhos, enquanto das 20 de Serviço social foram coletados 71 cabeçalhos, totalizando 149 cabeçalhos. Os cabeçalhos coletados foram comparados com os cabeçalhos de assuntos autorizados pelos catálogos de assuntos do SIBI/UFAL. A princípio foi tabulado apenas os dados do curso de Enfermagem.

O curso de Enfermagem apresentou 69 (sessenta e nove) cabeçalhos autorizados e 9 (nove) de linguagem natural incluídos pelos autores. Percebe-se que há muitos cabeçalhos autorizados, correspondendo a aproximadamente a 89,75% de concordância, enquanto 10,25% cabeçalhos provenientes da linguagem natural, conforme Figura 1.

Figura 1- Cabeçalhos autorizados e cabeçalhos em linguagem natural



Fonte: Dados da pesquisa (2022).

Algo relevante para destacar pelo bom grau de concordância é o fato dos alunos de Enfermagem utilizarem com frequência a Base de dados DeCS para coletar termos da área da saúde, bem como o uso dos cabeçalhos de assuntos constante na ficha catalográfica, os quais são elaboradas pelos bibliotecários do SIBI/UFAL sendo item obrigatório em suas monografias. Os alunos do curso de Enfermagem ao fazerem o autodepósito foram orientados a incluírem também no momento do preenchimento dos metadados assuntos, além das palavras chaves do seu resumo, os cabeçalhos de assuntos da ficha.

Quanto a cabeçalhos com erros de grafia ou digitação, foram detectados apenas 4 (quatro) ocorrências, duas com erro de grafia e duas com erro de digitação.

6 CONCLUSÃO

Após análise, percebe-se que os metadados assuntos cadastrados pelos alunos de Enfermagem através do autodepósito, apresentam um bom grau compatibilidade com os termos autorizados utilizados no catálogo de assuntos do Sistema de Biblioteca da Universidade Federal de Alagoas (UFAL). Isso fica evidente pelo fato dos alunos da área de saúde utilizarem com frequência a base Decs para elaboração de suas pesquisas, assim, já tem uma familiaridade com utilização de cabeçalhos autorizados por fonte de referência autorizada. O fato da utilização da ficha catalográfica e da inclusão de seus cabeçalhos de assuntos tam bem corrobora para um alto índice de concordância, visto que é elaborado por profissionais bibliotecários, qualificados abituados em utilizar as fontes de referências específicas de cada área. Outro ponto a destacar é quer, apesar de apresentar um bom índice de concordância com o catalogo de assuntos do SIBI, percebe-se que alguns trabalhos pecam na questão de quantidade de termos utilizados, apresentado uma quantidade limitada, não explorando com mais afinco o conteúdo do documento.

Diante desta, fica evidente que a contribuição do bibliotecário na orientação é fundamental na busca de fontes qualificada, por isso é importante na elaboração de guias e manuais para orientação dos usuários do repositório em como depositar seus trabalhos, além disso, é fundamental que o bibliotecário do repositório faça a revisão dos metadados assuntos ates da validação final do documento na base de dados..

REFERÊNCIAS

ALBUQUERQUE, M. E. C. ; GAUDÊNCIO, S. M. ; SANTOS, R. F. Reflexões teóricas em representação da informação. In: ALBUQUERQUE, M. E. C. ; MARTINS, G. K. ; MOTA, D. A. R.(orgs.). **Organização e representação da informação e do conhecimento**: intersecções teórico-sociais. João Pesswoa: Editora UFPB, 2019.

ALVES, V. B. A. Open archives: via verde ou via dourada? **PontodeAcesso**, Salvador, v. 2, n. 2, p. 127-137, ago./set. 2008. Disponível em: <http://www.pontodeacesso.ici.ufba.br>. Acesso em: 06 abr. 2023

ASSIS, T. B. Análise das políticas de autoarquivamento nos repositórios institucionais brasileiros e portugueses. **InCID: Rev. Ci. Inf. Doc**, Ribeirão Preto, v. 4, n. 2, Ed. Esp., p. 212-227, jul./dez 2013.

FERREIRA, L. C. ; ALBUQUERQUE, M. E. C. A representação da informação para a organização do acervo em bibliotecas universitárias. In: ALBUQUERQUE, M. E. C. et al (orgs.).

Representação da informação: um universo multifacetado. João Pessoa: Editora UFPB, 2013.

FREITAS, M. P. **Autoarquivamento e representação de assunto:** estudo analítico de teses e dissertações do Repositório Institucional da UFSCar. 2019. 89 f. Dissertação (Mestrado em Ciência da informação) – Centro de Educação e Ciências Humanas, Programa de Pós-Graduação em Ciência da informação, Universidade Federal de São Carlos, São Carlos, 2019.

FREITAS, M. P.; DAL´EVEDOVE, P. R.; TARTAROTTI, R. C. D. Políticas de autoarquivamento em repositórios institucionais brasileiros: estudo analítico do metadado assunto. **Páginas A&B, Arquivos e Bibliotecas (Portugal)**, n. Especial, p. 169-175, 2021. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/157369>. Acesso em: 06 abr. 2023

GOMES, F. A. Padronização de metadados na representação da informação em repositórios institucionais de universidades federais brasileiras. **ponto de Acesso**, Salvador, v. 10, n. 2, p.93-94, ago. 2016. Disponível em: <http://www.pontodeacesso.ici.ufba.br>. Acesso em: 05 dez. 2023.

LANCASTER, F. W. **Indexação e resumos:** teoria e prática. 2.ed. Brasília: Brique de Lemos, 2004.

LEITE, F. C. L. **Como gerenciar e ampliar a visibilidade da informação científica brasileira:** repositórios institucionais de acesso aberto. Brasília: IBICT, 2009.

MARCONDES, C. H. Representação e economia da informação. **Ci. Inf.**, Brasília, v. 30, n. 1, p. 61-70, jan./abr. 2001. Disponível em: <http://www.scielo.br/pdf/ci/v30n1/a08v30n1>. Acesso em: 30 jun. 2019.

SANCHEZ, F. A. ; SILVA, N. B. P.; VECHIATO, F. L. Padrão de metadados para representação e organização da informação em repositórios de dados de pesquisas, **Informação & Tecnologia (ITEC)**, Marília/João Pessoa, v.5, n. 1, p.37-51, jan./jun. 2018.

SANTOS JUNIOR, E. R. **Repositórios institucionais de acesso livre no Brasil:** estudos Delfos. 2022. 89 f. Dissertação (Mestrado em Ciência da informação) – Faculdade de Ciência da Informação, Programa de Pós-Graduação em Ciência da informação, Universidade de Brasília, Brasília, 2010.

VILLALOBOS, A. P. O; GOMES, F. A. Padronização de matadados na representação da informação em repositórios institucionais de universidades federais brasileiras. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 16., 2015, João Pessoa. **Anais [...]** João Pessoa: ANCIB, 2015. Disponível em: <http://www.ufpb.br/evento/index.php/enancib2015/enancib2015/paper/viewFile/2664/998>. Acesso em: 05 abr. 2023.