

WEB DOCUMENTAÇÃO: UMA EXPERIÊNCIA DA BASE DE DADOS NA CONSTRUÇÃO DE UM CONJUNTO DE DOCUMENTOS INTERLIGADOS

WEB DOCUMENTATION: AN ARCHIVAL SCIENCE DATABASE EXPERIENCE IN BUILDING A SET OF INTERCONNECTED DOCUMENTS

Katia Isabelli Melo

Professora do curso de Arquivologia - Universidade de Brasília. Doutora em Documentación: Archivos y Bibliotecas pela Universidad Carlos III. Líder do Grupo de Pesquisa Estudos prospectivos: formação e atuação profissional do arquivista/CNPq. Orcid: <https://orcid.org/0000-0002-0559-3891>. E-mail: isabelli@unb.br

Douglas Paiva

Arquivista. Universidade de Brasília. Membro do Grupo de Pesquisa Estudos prospectivos: formação e atuação profissional do arquivista/CNPq. Orcid: <http://orcid.org/0000-0003-1181-4109>. E-mail: dglspai-va2@gmail.com

RESUMO

O presente artigo resgata o processo de construção da Base de Dados em Arquivística, BDA, ocorrida em 2019, e apresenta as principais alterações e adaptações implantadas, visando maior funcionalidade e fluidez na recuperação das informações. A ferramenta surge com uma proposta inédita e original, considerando a inexistência de uma base de dados abordando, especificamente, a produção científica e técnica, em arquivística, oriunda dos autores brasileiros. Aborda, também, a implantação da taxonomia visando agilidade na busca da informação, interrelacionar e contextualizar as informações. O estudo inclui uma revisão bibliográfica de conceitos que resultou nos destaques: taxonomia, relacionamento, web documentação, raspagem de dados e contexto documental. Partido disso, houve um mapeamento dos metadados a serem transformados em taxonomia e definição de sua estrutura. Posteriormente, ocorreu a migração dos itens, antes registrados na base, para a nova estrutura de dados. Em seguida, aconteceu a padronização e correção dos termos taxonômicos migrados. Por fim, a atualização dos manuais e o cadastramento de novos itens seguindo a nova estrutura. A partir da inclusão de novas funcionalidades, espera-se que a BDA apresente maior fluidez no atendimento às necessidades de informação dos profissionais da informação, sobretudo dos arquivistas, discentes dos cursos de Arquivologia e demais profissionais da informação.

Palavras-chave: arquivística; base de dados; produção científica; taxonomia; *web* documentação.

ABSTRACT

This article discusses the construction process of the Archival Science Database (BDA) that took place in 2019, highlighting the main changes and adaptations implemented to enhance functionality and streamline information retrieval. The tool represents an innovative and original proposal, as there is currently no database specifically focused on Brazilian-authored scientific and technical production in archiving. The article also addresses the implementation of taxonomy to expedite information search, interrelates and contextualizes data, and consolidates a research tool. The study included a bibliographical review of concepts that resulted in the identification of taxonomy, relationship, web documentation, data scraping, and document context as critical features. This was followed by mapping metadata to be transformed into taxonomy and defining its structure. Subsequently, migration of items previously registered in the database was performed into the new data structure, followed by normalization and correction of migrated taxonomic terms. Finally, manuals were updated, and new items were registered following the new structure. With the inclusion of new features, the BDA is expected to provide greater fluidity in meeting the information needs of information professionals, particularly archivists, students of archival courses, and other information professionals.

Keywords: archival Science; database; scientific production; taxonomy; web documentation.

INTRODUÇÃO

No Brasil, algumas ações nos levam a crer que a década de 1970 é um marco significativo para a Arquivologia, com o surgimento do movimento associativo, em 1971, a realização do primeiro congresso científico da área, em 1972, a institucionalização da graduação em Arquivologia na Universidade Federal de Santa Maria, em 1977, e o reconhecimento da profissão de arquivista no ano seguinte, 1978, referendadas por Fonseca (2008).

Inicialmente, a base teórica em arquivística constituía-se, basicamente, da produção científica europeia, sobretudo de autores holandeses, franceses, ingleses e espanhóis, com contribuições norte-americanas dos profissionais canadenses e americanos. No caso brasileiro, identificou-se que uma das obras mais antiga, de autoria nacional, data de 1931, ainda que estejamos no processo de pesquisa para identificar o primeiro livro publicado no país. Identificar e consolidar em um repositório toda a produção científica, até então completamente dispersa, revela-se como o principal objetivo da Base de Dados em Arquivística, BDA. Segundo Cunha, as bases de dados “são fontes de informação computadorizadas que podem ser pesquisadas num modo interativo ou conversacional através de um terminal de computador, telex ou mesmo um microcomputador” (CUNHA, 1989, p. 45).

Nas últimas décadas percebeu-se uma amplitude na produção e uso dos recursos tecnológicos nas mais variadas atividades cotidianas, ou seja, no ambiente de trabalho, na prática do ensino, na aquisição de bens e serviços, na interação entre as pessoas, dentre outras. O mesmo ocorreu em relação à pesquisa em arquivística, principalmente os resultados de aplicações teóricas e práticas, observações, estudos de caso, análise científica, produzidos pelos profissionais da área e áreas afins. Nesse sentido, a BDA, surge com a proposta de reunir numa ferramenta toda a produção científica e técnica de autores brasileiros, abarcando a pluralidade das temáticas arquivísticas, bem como acompanhando o crescimento das reflexões na área. A Base utiliza a ferramenta Tainacan, tendo como desenvolvedores, docentes da Universidade Federal de Goiás e da Universidade de Brasília. O Tainacan é vinculado à WordPress, sistema livre e aberto de gestão de conteúdo para a internet. Desde que foi desenvolvida, a BDA recebe suporte operacional do Núcleo de Tecnologia da Informação, da Faculdade de Ciência da Informação, da Universidade de Brasília. Na implantação e atualização das novas funcionalidades, agregou-se a participação do arquivista Douglas Paiva.

A partir da apresentação da Base de Dados em Arquivística no II Encontro de Arquivos, Bibliotecas e Museus, ocorrido em novembro de 2022, em Porto, Portugal, reunimos informações para compor o presente estudo, sobretudo acerca da implantação de novos recursos e funcionalidades para a BDA, como as taxonomias. Adotou-se como metodolo-

gia uma pesquisa bibliográfica tendo como aporte teórico os estudos de Dahlberg (1978), Campos (2004), Terra et al. (2005) e Araújo Júnior; Souza; Albuquerque (2015).

Lançada oficialmente em setembro de 2021, na Semana Universitária da Universidade de Brasília, foi obtendo reconhecimento e formando parcerias no meio acadêmico. Destaca-se que a BDA constitui um infoproduto do Grupo de Pesquisa “Estudos Prospectivos: formação e atuação profissional do arquivista”, a partir de desdobramento da linha de pesquisa Publicações científicas no âmbito arquivístico nacional: pesquisa e recuperação da informação. O referido Grupo integra o Diretório dos Grupos de Pesquisa no Brasil, do Centro Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

CONSTRUÇÃO E EVOLUÇÃO DA BDA

A Universidade de Brasília atua em três segmentos, sendo: ensino, pesquisa e extensão. No segmento de extensão, o discente participa de eventos e projetos por meio do Decanato de Extensão, DEX, propiciando engajamento com a sociedade em geral mediante as ações propostas. Uma outra modalidade, vinculada ao segmento da pesquisa, é o Projetos de Iniciação Científica, ProIC, sob a coordenação da Diretoria de Iniciação Científica, do Decanato de Pós-Graduação. O ProIC tem a finalidade de introduzir os discentes na prática da pesquisa. Em ambos os projetos, os discentes são orientados por um docente.

O processo de construção da BDA teve a participação exclusiva de discentes dos cursos de Arquivologia e Biblioteconomia atuando de forma voluntária e outros, vinculados como bolsistas. Desde a fase inicial, a equipe foi constituída com recursos de bolsas de iniciação científica, ProIC¹, e dos projetos de extensão, vinculados ao DEX². Os integrantes da equipe mostraram-se engajados com as frentes de trabalho, ainda que em determinados momentos tenham ocorrido remanejamentos internos e desligamentos voluntários.

Destacaremos a pesquisa e a contribuição significativa desse viés para a arquivística, considerando as suas diversas representações nos formatos livros, capítulos de livros, artigos científicos trazendo visibilidade para a área. A pesquisa científica permite novas reflexões sobre o pensar e o fazer arquivístico, e mesmo acerca de outras temáticas que, anteriormente, não constituíam objeto de estudo e que possibilitam outros olhares investigativos, contribuindo para a geração e a ampliação do conhecimento arquivístico.

Constatou-se que a produção científica em arquivística revela-se dispersa. Uma parcela está registrada em diversas bases de dados, outra apresenta-se somente na for-

¹ Em 2019 teve início o primeiro Projeto de Iniciação Científica, com uma equipe composta por três alunos e uma aluna, ambos do curso de Arquivologia da UnB.

² O primeiro projeto de extensão teve início em 2020 tendo como integrantes da equipe duas alunas e um aluno do curso de Arquivologia da UnB.

ma impressa, e outra desconhecemos, sobretudo as mais antigas, registradas no formato impresso. A esse grupo, inclusive, inserem-se os primeiros estudos debatidos nos eventos científicos da área em que foram produzidos poucos exemplares, sendo que uma parcela está praticamente indisponível. Mesmo em bibliotecas específicas constataram-se algumas lacunas das publicações. Tais fatores impactaram diretamente no processo de pesquisa, que se revelou, e ainda se mantém, como a mais exaustiva do processo de cadastramento.

Paralelamente, outros problemas foram observados tais como links de acesso quebrados, que não direcionam para a informação solicitada. Em outras situações verificou-se que algumas páginas não estavam acessíveis e outras fora do ar.

Como forma de difusão para a comunidade arquivística das etapas e resultados obtidos na construção da BDA, a equipe apresentou alguns resultados nos eventos científicos em âmbito nacional e internacional. No âmbito acadêmico ocorreram participações em eventos de extensão, onde somam-se entrevistas por meio de lives e outras, reproduzidas em podcast. Inserem-se, ainda, os registros nos periódicos científicos, em formato digital, conforme indicados no Quadro 1.

Quadro 1 - Divulgação a BDA em eventos e publicações científicas

EVENTO/PROMOÇÃO	TEMÁTICA	ANO
Simpósio Internacional de Arquivos (Associação dos Arquivistas de São Paulo)	Base de Dados em Arquivística: a produção científica brasileira	2020
Projeto de extensão “Comunicação institucional da Faculdade de Ciência da Informação através das redes sociais Facebook e Instagram” (Universidade de Brasília)	Projeto Construção de um modelo de base de dados na área de Arquivística	2020
Semana Universitária (Universidade de Brasília)	Divulgando a Base de Dados em Arquivística (BDA): possibilidades de pesquisa	2021
Programa Arquivo ao Vivo (Associação de Arquivistas da Paraíba e Grupo de Estudos Arquivísticos)	Base de Dados em Arquivística e Periódicos Universitários	2021
Oficina promovida pelo Projeto SESA, Associação de Arquivistas da Paraíba e Curso de Arquivologia da Universidade Estadual da Paraíba	Base de Dados em Arquivística: uma experiência de pesquisa	2021
Podcast Eccoa (Curso de Arquivologia da Universidade Federal do Rio Grande do Sul)	Base de Dados em Arquivística (BDA) - Episódio 34	2021
IX Congresso Nacional de Arquivologia	Base de dados em Arquivística: os congressos de Arquivologia	2022
VII Reunião Brasileira de Ensino e Pesquisa em Arquivologia	Base de dados em Arquivística: preservando a produção científica brasileira	2022
Atividade de extensão O pensar e o fazer arquivístico: algumas reflexões (Curso de Arquivologia da Universidade de Brasília)	A pesquisa bibliográfica em Arquivologia e o papel do arquivista na preservação e difusão	2022

EVENTO/PROMOÇÃO	TEMÁTICA	ANO
Atividade de extensão Estudos prospectivos: formação e atuação do profissional arquivista (Curso de Arquivologia da Universidade de Brasília)	O processo de construção de ferramenta tecnológica visando a preservação e a difusão da produção científica em Arquivologia: estudo de caso	2022
II Encontro Internacional de Arquivos, Bibliotecas e Museus (Universidade do Porto)	Base de Dados em Arquivística: fonte de pesquisa referencial no Brasil	2022
Revista Participação (Decanato de Extensão / Universidade de Brasília)	Construção da Base de Dados em Arquivística: uma ferramenta de pesquisa	2022

Fonte: Melo e Paiva (2023)

2.1 ESTABELECIMENTO DAS CATEGORIAS DA BDA

Na composição da BDA foram estabelecidas três categorias: Periódicos, Eventos científicos e Monografias³. Um dos resultados de busca refere-se à autoria dos itens registrados nas respectivas categorias. Atualmente, a BDA registra cerca de 6.500 autores, sendo a maior parcela de autores nacionais. A participação de autores estrangeiros na Base revela-se quando da tradução de artigos para o idioma português e mesmo na participação de palestrantes convidados, nos eventos nacionais e internacionais realizados no Brasil.

Os Anais do I Congresso Brasileiro de Arquivologia, I CBA, realizado em 1972, registram um primeiro panorama das discussões científicas da área. Os eventos científicos constituem o registro fiel da produção científica, sobretudo o CBA que revelou-se como pioneiro.

Conforme mencionado anteriormente, no que se refere às monografias, a base teórica adotada nos cursos de formação em Arquivologia refletia os estudos de autores norte-americanos e europeus, sobretudo holandeses, franceses, ingleses e espanhóis. Uma contribuição significativa e que se mantém nas referências acadêmicas é a obra Arquivos modernos: princípios e técnicas, de Theodore Roosevelt Schellenberg, traduzida pela Profa. Nilza Teixeira Soares. Anterior à essas obras, a BDA registra algumas publicações, conforme o Quadro 2. Uma pesquisa acerca das primeiras publicações abordando arquivística, editadas no Brasil, constam da tese A patrimonialização cultural de arquivos no Brasil, de Francisco Alcides Cougo Junior, apresentada em 2021,⁴ na Universidade Federal de Pelotas.

³ Na BDA, a categoria Monografias é composta por livros e capítulos de livros, incluindo também cartilhas e manuais.

⁴ Disponível em: http://guaiaca.ufpel.edu.br:8080/bitstream/prefix/7423/1/Tese_Francisco_Alcides_Cougo_Junior.pdf. Acesso em: 5 maio 2023.

Quadro 2 - Primeiras monografias registradas na BDA⁵

TÍTULO	AUTOR(A)	ANO
Manual de classificação e arquivamento de papéis e documentos commerciaes e civis	Ernani Macedo de Carvalho	1931
Os arquivos particulares do Itamaraty	Aluizio Napoleão	1940
Arquivar e achar	Eric Watson White	1941
Arquivar em medicina pelo sistema decimal	Alvino Paula	1943

Fonte: Melo e Paiva (2023).

No que se refere aos periódicos científicos, a revista *Arquivo & Administração*, editada pela Associação dos Arquivistas Brasileiros, consagra-se como o primeiro periódico específico da área de Arquivologia, editada no período de 1972 a 2014. Ainda que revele algumas interrupções, a revista destaca-se como um dos mais importantes periódicos científicos brasileiros na área de Arquivologia. Com abordagem arquivística, identificamos artigos isolados publicados anteriores às edições de revistas dos arquivos estaduais, como a *Revista do Archivo Municipal de Minas Gerais*, que apresenta o primeiro artigo em 1896, conforme Quadro 3.

Quadro 3 - Primeiros artigos publicados com temática arquivística

TÍTULO	AUTOR	ANO
Archivo Publico Mineiro - Lei n. 126, 11 jul. 1895	-	1896
Archivo Publico Mineiro - Histórico; organização; fins; vantagens; sede e pessoal	Arquivo Público Mineiro	1927
Regulamento do Archivo Publico Mineiro - Decreto n. 860, de 19 de setembro de 1895.	Arquivo Público Mineiro	1929
As riquezas do Archivo	Escragnolle Doria	1929
Os papéis antigos do Arquivo Municipal de São Paulo	Nuto Santanna	1934

Fonte: Melo e Paiva (2023).

Em 2004, ocorre na cidade de Brasília, o I Congresso Nacional de Arquivologia, organizado pela Associação Brasiliense de Arquivologia, com o apoio do curso de Arquivologia da Universidade de Brasília. Outros eventos surgiram, posteriormente, sendo al-

⁵ Vários títulos de monografias registrados por Cougo Júnior estão em fase de cadastramento na BDA. Disponível em: http://guaiaca.ufpel.edu.br:8080/bitstream/prefix/7423/1/Tese_Francisco_Alcides_Cougo_Junior.pdf. Acesso em: 5 maio 2023.

guns promovidos pelos cursos de Arquivologia ou instituições arquivísticas públicas. A agenda dos eventos é diversificada, ora com discussões abrangentes, ora com temáticas mais específicas, como o Seminário Internacional de Arquivos Pessoais, organizado pela Fundação Getúlio Vargas, e o Seminário Arquivo e Memória Escolar, promovido pela Associação de Arquivistas de São Paulo. O Quadro 4, apresenta um recorte dos eventos da área realizados até o presente momento, alguns registrados na BDA e outros em processo de cadastramento.

Quadro 4 - Eventos da área

TÍTULO	INSTITUIÇÃO(ÕES) PROMOTORA(S)	PERIODICIDADE
Congresso Brasileiro de Arquivologia	Associação dos Arquivistas Brasileiros	Bienal
Congresso Nacional de Arquivologia	Fórum das Associações de Arquivistas – FNArq	Bienal
Jornada Arquivística	Universidade Federal do Estado do Rio de Janeiro	Anual
Reunião Brasileira de Ensino e Pesquisa em Arquivologia	Fórum Nacional de Ensino e Pesquisa em Arquivologia – FEPARQ	Bienal
Simpósio Internacional de Arquivos	Associação dos Arquivistas do Estado de São Paulo	Bienal
Seminário Nacional de Governança Arquivística	Grupos de pesquisa das universidades	Anual
Simpósio de História dos Arquivos e da Arquivologia	Coordenação do curso de Arquivologia e Programa de Pós-Graduação em Ciência da Informação da Universidade Federal Fluminense	-

Obs.: O Congresso Brasileiro de Arquivologia interrompeu permanentemente as edições em 2014.

Fonte: Melo e Paiva (2023).

Os resultados parciais aqui consolidados acerca dos eventos científicos, monografias e artigos de periódicos mostravam-se dispersos. A partir da construção da BDA tornou-se possível mapear e consolidar a produção científica brasileira num instrumento específico, segundo as três categorias propostas e disponibilizar todas as informações para os pesquisadores.

2.2 TAXONOMIAS E RELACIONAMENTOS NA BDA

A BDA surgiu para criar uma centralização, facilitar o acesso e encurtar o processo de mergulho na memória científica em Arquivística, dando base aos pesquisadores. Possibilitando que a comunidade científica dê continuidade ou estabeleça novos campos de pesquisas, principalmente para promover um melhor intercâmbio entre estudos.

Quando um pesquisador, diante de um microcomputador ligado a um banco de dados [...] à procura de informações que definam, completem ou estabeleçam as fronteiras do seu trabalho de pesquisa, ele repete o mesmo gesto de quem mergulha na memória de seu grupo para reconstruir as lembranças comuns e dessa forma manter íntegra a sua comunidade (SAYÃO, 1996, p. 314).

A estrutura da BDA permite a recuperação das informações considerando os metadados estabelecidos sendo o título do item cadastrado, autor, local, editora, ano de publicação, termos que designam as palavras-chave, dentre outros. Inicialmente, os itens cadastrados apresentavam uma estrutura com alguns critérios de formatação e de indicação de conteúdo. Objetivando ampliar o alcance de novos investigadores e obter maior projeção internacional da BDA, iniciamos a inclusão dos resumos em outro idioma, acompanhados das respectivas palavras-chave, quando registrado no artigo original. Destaca-se que os artigos científicos mais antigos estão isentos de resumos em outro idioma. Inclusive, os primeiros artigos deixam de apresentar resumo no idioma oficial do texto.

Considerando a necessidade em ampliar a funcionalidade da ferramenta e atender uma melhor forma de recuperação das informações, propõe-se a implantação de taxonomias. Segundo Araújo Júnior; Souza; Albuquerque, a taxonomia é um mecanismo de classificação que possibilita a identificação, localização e acesso. Salientam os autores,

A taxonomia é uma estrutura que possibilita classificar objetos, seres vivos, coleções de livros ou documentos em grupos ordenados hierarquicamente, a fim de possibilitar sua identificação, localização e acesso. Pode também ser definida como um sistema de classificação que apoia o acesso à informação, permitindo alocar, recuperar e comunicar informações em um sistema, de maneira lógica (ARAÚJO JUNIOR; SOUZA; ALBUQUERQUE, 2015, p. 63-64).

Na definição de Campos (2004, p. 24) a taxonomia consiste em “um mecanismo de representação do conhecimento que possibilita, para Ciência da Informação, a elaboração de linguagens documentárias verbais e notacionais, visando à recuperação de informações e a organização dos conteúdos informacionais de documentos”. Desse modo, é válido afirmar que a adoção da taxonomia, consiste numa excelente contribuição para a recuperação, compreensão e inter-relacionamento das informações.

Os sistemas de organização e representação do conhecimento são adotados na ordenação de conceitos de determinadas áreas com o objetivo de representar conteúdo para a recuperação. A taxonomia não adota uma metodologia única, que se aplica a todos os espaços, a sua contribuição e aplicação irá depender das particularidades e objetivos específicos de cada situação. É importante compreender que a composição mais mencionada, na literatura, quando se fala em taxonomia é a hierárquica, sendo uma forma de

caracterizá-la. A taxonomia organiza a informação da mais genérica até a mais específica, utilizando a hierarquia entre os termos. Essa relação é definida por Dahlberg como:

Se dois conceitos diferentes possuem características idênticas e um deles possui uma característica a mais do que o outro, então entre eles se estabelece a relação hierárquica ou relação de gênero e espécie. Pode-se então falar de conceitos mais amplos ou mais restritos. Pode-se também falar de conceito superior e inferior. O conceito superior é o mais genérico e o inferior é o mais específico. Se falamos de macieira temos como conceito mais amplo ou superior o conceito de árvore frutífera e mais genérico ainda o conceito de árvore.

A taxonomia, ademais de classificar e facilitar o acesso à informação, apresenta outras finalidades, conforme ressaltam Terra *et al.* (2005, p. 1):

representar conceitos através de termos, agilizar a comunicação entre especialistas e outros públicos; encontrar o consenso; propor formas de controle da diversidade de significação e oferecer um mapa de área que servirá como guia em processo de conhecimento. É, portanto, um vocabulário controlado de uma determinada área do conhecimento, e acima de tudo um instrumento ou elemento de estrutura que permite alocar, recuperar e comunicar informações dentro de um sistema, de maneira lógica.

A BDA é uma ferramenta desenvolvida com base no Tainacan, software aberto, identificado com um campo propício a melhorias. A partir de um levantamento considerando todas as funcionalidades disponibilizadas pelo Tainacan e em outras plataformas, buscou-se verificar a possibilidade de aplicação dessas funcionalidades e até mesmo na adaptação para alcançar os objetivos estabelecidos, assim como os mecanismos de migração.

Ao fim da avaliação foram identificados dois recursos do Tainacan que atendiam às novas necessidade da BDA sendo o relacionamento e as taxonomias. O relacionamento é uma importante ferramenta para a BDA por diversos motivos sendo o principal deles relacionado à busca na Web. Quando falamos em um documento em rede, não estamos falando de um documento individualizado e trabalhado apenas em seu conteúdo/texto, mas sim de um documento/item/objeto multifacetado que possui vínculos e que comunicam entre si. Este tratamento individualizado fez e faz diversos itens arquivísticos se perderem na Web, não sendo identificado com facilidade pelos buscadores. A partir desse entendimento, buscou-se integrar na BDA os dois recursos, relacionamento e taxonomia, propiciando mais visibilidade aos itens já registrados. As taxonomias constituem o pilar central dessa atualização visto que, além de auxiliar nesse processo de interrelação, apresenta outras características essenciais para a organização, localização e hierarquização dos itens.

A partir do funcionamento da BDA constatou-se a possibilidade de agregar novas ferramentas e conhecimentos para auxiliar nesse objetivo, o que propiciou a reorganização dos metadados disponíveis na Base e o desenvolvimento de taxonomias.

O processo de implantação das taxonomias foi dividido em quatro etapas. A primeira consistiu no mapeamento dos metadados registrados na BDA para relacionamento e taxonomia sendo considerados os seguintes metadados: título da obra, autor(es), resumo, palavras-chave, nome do periódico, ano de publicação, categoria a que pertence, coleção a que pertence, edição, editora(s), abstract, keywords, local, natureza, organizador(es).

A segunda etapa considerou a migração dos itens anteriormente registrados na BDA para a nova estrutura de dados. Com a manutenção da BDA no Tainacan foi necessária uma estratégia de migração pois os metadados destacados para taxonomia apresentavam estruturas incompatíveis com o modelo de metadados no estilo de taxonomia. Nessa etapa foi feita a exportação, em formato CSV, dos itens para o tratamento dos metadados, ou seja, separação utilizando ponto e vírgula, vírgula ou ponto para as palavras-chave, autores e outros metadados que apresentavam multi-valores.

Posteriormente foi feita uma raspagem de dados, via Web scraping, utilizando a linguagem Python adaptada à estrutura Java do Tainacan, para recuperação das imagens que anteriormente eram registradas apenas para criação de miniaturas das capas das publicações cadastradas. Após a separação de cada item presente nos metadados foi identificado quais metadados eram essenciais aos itens individualmente e quais eram recuperáveis por via relacionamento. Esse procedimento possibilitou economizar tempo na etapa de registro de novos itens na Base considerando que serão registrados somente os metadados essenciais ao item sendo, os demais, recuperados automaticamente via relacionamentos, construindo uma teia de relação entre os itens registrados na BDA. A reinserção dos itens foi feita em lote por um arquivo CSV que, automaticamente, inseriram os termos novos nas taxonomias existentes.

Na terceira etapa ocorreu a normalização e correção dos termos taxonômicos migrados. Essa fase possibilitou a análise dos termos taxonômicos criados e, paralelamente, melhor qualidade para a apresentação dos itens cadastrados na BDA visto que contribuiu na identificação de possíveis erros ortográficos, duplicidade de termos, divergências causadas por espaçamentos e outros.

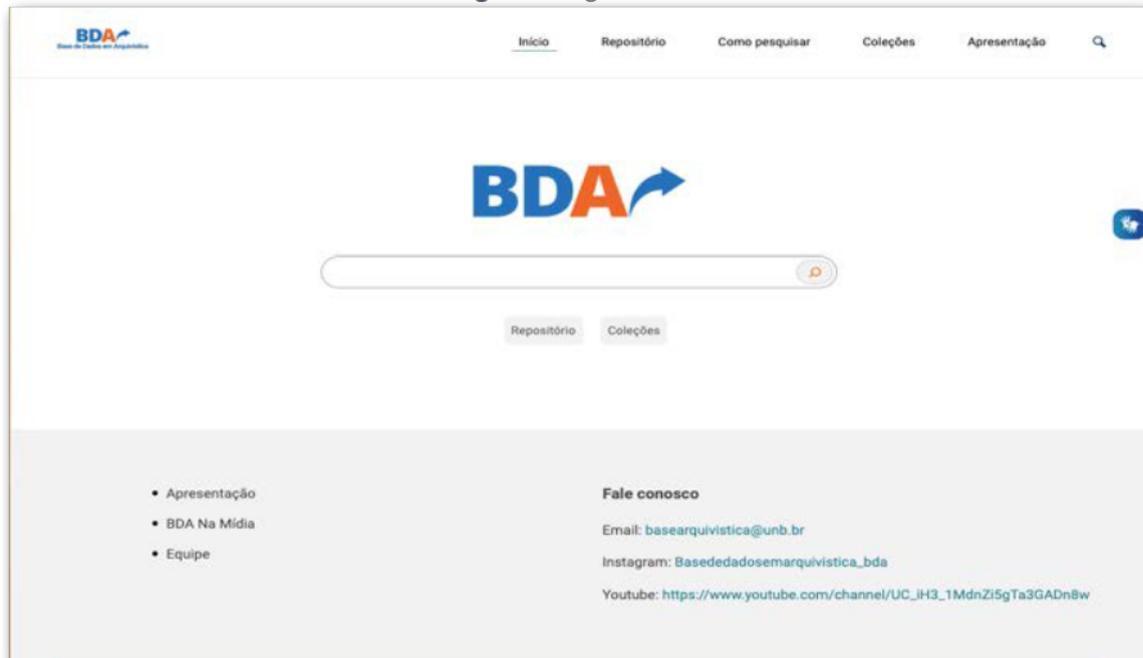
Na etapa subsequente foi criada a hierarquização dos termos e os devidos relacionamentos. Por fim, na quarta etapa procedeu-se à execução da atualização dos Manuais internos de cadastramento e a etapa de cadastramento propriamente dita para os novos itens, seguindo a nova estrutura proposta. Desde a criação da BDA sentiu-se a necessidade de elaborar um Manual, como um instrumento de apoio, que possibilitasse o registro da sequência das etapas para o cadastramento dos itens. Com a inclusão das taxonomias e relacionamento dos itens, fez-se necessária a Manual do colaborador bem como a atualização do Manual do pesquisador, com instruções do passo-a-passo para a realização das buscas na Base, disponível na página da BDA.

2.3 RESULTADOS ALCANÇADOS

O fator tempo é um elemento fundamental a ser considerado no processo de recuperação das informações na Internet. A partir da implantação das taxonomias, os resultados tornaram-se extremamente satisfatórios. Nesse quesito, o resultado positivo, claramente identificado, foi o aumento da velocidade do tempo de resposta que passou de 9 a 17 segundos para 0,41 a 1,9 segundos, sendo que em momentos de instabilidade, atualmente, chega a 8,8 segundos que é bem inferior dos anteriores 40 segundos. O resultado apresenta impacto também no processo de registro de itens na Base, favorecendo a localização das informações já existentes, o que garante redução de tempo a etapa de cadastramento.

A Figura 1 apresenta a página inicial da BDA, uma nova página de pesquisa intuitiva e geral, que possibilita uma análise de todos os itens da Base, verificando o termo pesquisado nos títulos e resumos. Essa é mais rápida devido ao fato de verificar apenas esses dois metadados, no entanto ela é suficiente para atender as necessidades da maioria das pesquisas.

Figura 1- Página inicial

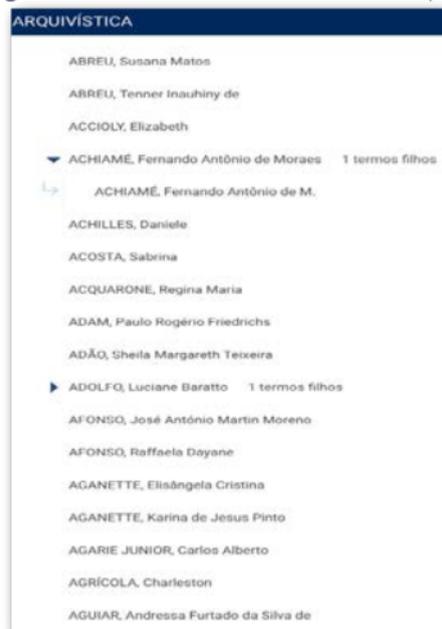


Fonte: Melo e Paiva (2023).

De maneira mais específica para os metadados “autor(es)” e “palavras-chave”, a estrutura taxonômica possibilitou que o registro de informações permanecesse fiel ao documento original, algo que desde o início foi definido como uma prática obrigatória no processo de registro. Observou-se, no entanto, que é comum um mesmo autor adotar

escritas diferentes para o registro do nome em publicações distintas, como um sobrenome a mais ou a menos, além da ocorrência de pseudônimos, e nesses casos o processo de localização da totalidade de obras produzidas por uma mesma pessoa ficava comprometido, visto que até mesmo um erro de espaçamento no registro poderia fazer com que a Base considerasse a publicação “X” como pertencente a um novo autor. Tais situações, apesar de não serem uma regra, ainda assim revelava-se como um empecilho no processo de filtragem, conforme representado na Figura 2.

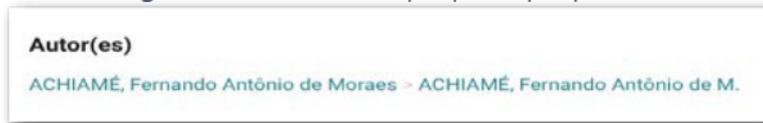
Figura 2 - Estrutura da taxonomia “Autor(es)”



Fonte: Melo e Paiva (2023).

Com a taxonomia foi possível manter a inscrição nominal original presente na obra, ao passo que também foi possível hierarquizar variações de nomes aplicados a uma mesma pessoa, sob um único termo pai. Isso gerou um metadado muito mais completo nos itens e com um *hyperlink* de filtragem, onde o usuário visualiza o nome do autor e pode clicar no mesmo e filtrar todas as obras publicadas com aquele nome. Mas caso o usuário pretenda ter acesso a todas as publicações daquela pessoa, independente do nome registrado na publicação, terá à sua disposição a trilha hierárquica daquele autor, podendo igualmente clicar no termo pai e filtrar pela nomenclatura mais abrangente, conforme registrado na Figura 3. A hierarquia está representada no termo mais abrangente, à esquerda, associado ao símbolo *Chevron* “>” seguido do termo à direita, utilizado no item original.

Figura 3 - Visual da hierarquia para o pesquisador



Fonte: Melo e Paiva (2023).

O mesmo processo foi aplicado nas palavras-chave, onde o processo de hierarquização potencializou a localização dos assuntos de interesse, reduzindo para aproximadamente 35% a quantidade de termos. As demais palavras-chave, por serem mais específicas, foram designadas como subitens dos termos mais genéricos. O procedimento possibilitou uma melhor filtragem dos itens visto que ao registrar um termo específico em um determinado item, toda a cadeia hierárquica passa a representá-lo, ao passo que anteriormente somente a palavra-chave específica remetia àquele item.

A título de exemplo, anteriormente, ao registrar um artigo tendo como palavra-chave, “Arquivo Público Mineiro” a busca dar-se-ia somente com o termo “Arquivo Público Mineiro”. Atualmente ao pesquisar o mesmo termo, a busca resultará nos termos “Arquivos Públicos” ou “arquivos” e as suas variações. Cabe ressaltar que a busca sempre é feita do termo mais específico ao mais genérico, assim sendo, a organização dos itens para o usuário revelará primeiramente o item que apresenta o termo mais específico, tal como pesquisado e, em seguida, os demais termos possibilitando uma maior exatidão no resultado, conforme representado na Figura 4. Caso nenhum item atenda a busca exata do usuário, a teia de inter-relacionamento retornará o item mais próximo pesquisado/cadastrado.

Figura 4 - Visual da hierarquia nas palavras-chave para o pesquisador

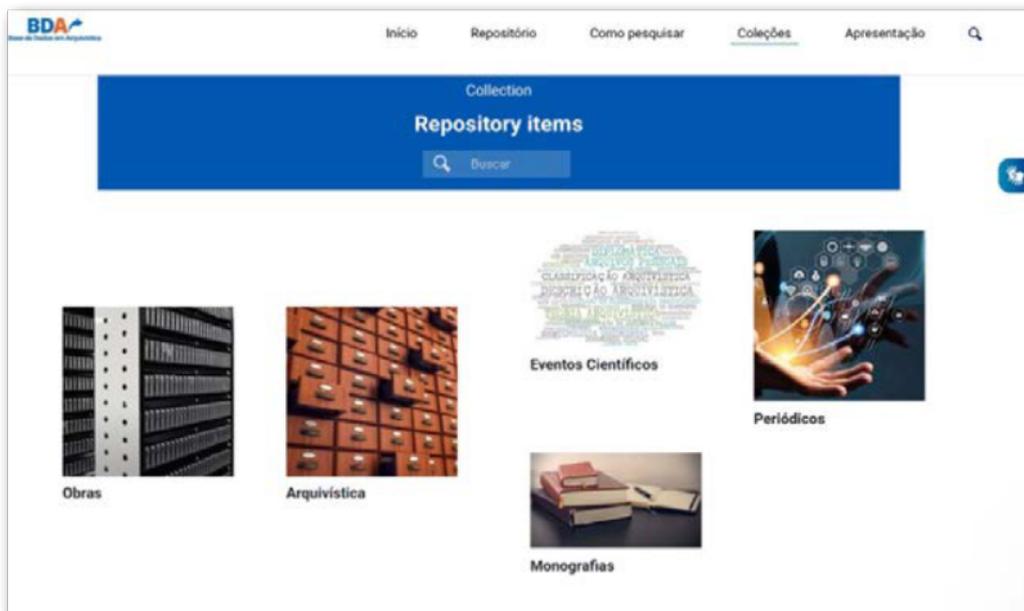


Fonte: Melo e Paiva (2023).

Conforme destacado acima, o termo à esquerda apresenta maior abrangência e o termo à direita, mais específico, é utilizado no item original. Outros benefícios do relacionamento e das taxonomias foram identificados em todos os metadados, como a possibilidade de busca por meio da navegação contínua entre os itens, a possibilidade de apresentar descritores para os metadados relacionados aos itens, o que gera maior riqueza da informação disponível. Inserem-se como benefícios adicionais a apresentação de significado de termos e nomenclaturas, a disponibilização de biografia, *link* para a Plataforma Lattes ou mesmo o *Open Researcher and Contributor ID, Orcid*, do(s) autor(es).

Na figura 5 é apresentada a aba de “Coleções”, onde o usuário tem a possibilidade de fazer uma busca mais detalhada entre os itens, que demanda um pouco mais de tempo devido a completude das análises na barra de pesquisa superior (o termo pesquisado é comparado com todos os metadados de todos os itens). Na página são apresentadas cinco possibilidades de filtros primários sendo que três deles constam desde o início da BDA: “Eventos científicos”, que registra todos os eventos e os respectivos itens, “Periódicos”, que apresenta todos os periódicos cadastrados e seus itens, e “Monografias”, que apresenta os livros, capítulos de livros, cartilhas e manuais. Dois filtros foram inseridos para complementar a etapa de busca na BDA sendo “Obras”, que apresenta os títulos dos livros, título dos periódicos e eventos científicos registrados na Base, e “Arquivística”, que registra os capítulos dos livros, artigos e as comunicações apresentadas nos eventos. O uso desses filtros permite que o usuário organize diversos filtros para atender às suas demandas de pesquisa, assim como possibilita uma navegação contínua entre os itens por meio de uma teia de relacionamento.

Figura 5 - Página de coleções da BDA



Fonte: Melo e Paiva (2023).

Outra aplicação da taxonomia deu-se no metadado “local” onde registra-se o espaço geográfico da publicação cadastrada. Atualmente, ao clicar um termo “local”, o pesquisador é direcionado para uma página com o ícone da bandeira desse local e uma descrição acompanhada de uma listagem com todos os registros cadastrados publicados.

No relacionamento com o periódico, monografia ou evento foi possível apresentar uma série de novas informações para contextualizar o item, desde o resumo existente, a

editora, a vinculação acadêmica, ano de publicação, seu foco e diversas outras informações que possam auxiliar na localização, contextualização e vinculação dos itens.

CONSIDERAÇÕES FINAIS

Os ajustes aplicados na BDA possibilitaram uma ferramenta dotada de uma estrutura mais enxuta de registro de dados, com um vocabulário mais controlado e preditivo, reduzindo o tempo de registro de novos itens assim como o estabelecimento de campos de controle interno contendo informações úteis à gestão dos trabalhos progressivos de registro de novos itens, por exemplo status da revista, se mostra-se ativa ou inativa, bem como o período de suas publicações.

A BDA é uma ferramenta de construção contínua considerando a necessidade de manutenção das etapas de pesquisa e registro de novos itens. A implantação de novas funcionalidades objetiva o atendimento das demandas dos pesquisadores.

Do ponto de vista estratégico, a BDA é pioneira na disseminação de conhecimento bem como contribui para a visibilidade da área ao revelar a evolução da produção científica e das discussões pertinentes à área.

REFERÊNCIAS

ARAÚJO JÚNIOR, R.; SOUSA, R. T. B. de; ALBUQUERQUE, S. F. de. Métodos, técnicas e instrumentos de organização e gestão da informação nas organizações. *In*: BAPTISTA, D.; ARAÚJO JÚNIOR, R. H. **Organização da Informação**: abordagens e práticas. Brasília. The-saurus, 2015, p. 44-68.

ARRUDA, W. R.; FELIPE, C. B. M.; SANTOS, R. F. Avaliação da qualidade das bases de dados BRAPCI e PERI da área de Ciência da Informação. **Ciência da Informação em Revista**, Maceió, v. 7, n. 1, p. 121-137, 2020. Disponível em: <https://www.seer.ufal.br/index.php/cir/article/view/8376/7410>. Acesso em: 5 maio 2023.

CABRAL, J. R., SANTOS, S. M. O saber-fazer arquivístico nas páginas da revista Arquivo & Administração. *In*: Encontro Internacional e Encontro de História da Anpuh-Rio: Histórias e Parcerias, 18., Rio de Janeiro. **Anais eletrônicos** [...], 2018, 11 p. Disponível em https://www.encontro2018.rj.anpuh.org/resources/anais/8/1529753786_ARQUIVO_CABRAL&-SANTOS-Saber-fazerarquivisticoemA&AINANPUH.pdf. Acesso em: 5 maio 2023.

CAMPOS, M. L. de A. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. **Ciência da Informação**. Brasília, v. 33, n. 1, p. 22-32, jan./abril. 2019. Disponível em: <https://revista.ibict.br/ciinf/article/view/4618>. Acesso em: 5 mar. 2023.

COUGO JUNIOR, F. A. **A patrimonialização cultural de arquivos no Brasil**. 448 f. 2021. Tese (Programa de Pós-Graduação em Memória Social e Patrimônio Cultural) - Instituto de Ciências Humanas, Universidade Federal de Pelotas, Pelotas, 2021. Disponível em: http://guaiaca.ufpel.edu.br:8080/bitstream/prefix/7423/1/Tese_Francisco_Alcides_Cougo_Junior.pdf. Acesso em: 10 mar. 2023.

CUNHA, M. B. da. Bases de Dados no Brasil: Um potencial inexplorado. **Revista IBICT – Ciência da Informação**, Brasília, v. 18, n. 1, p. 45-57, 1989. Disponível em: <http://revista.ibict.br/ciinf/article/view/322/322>. Acesso em: 5 maio 2023.

DAHLBERG, I. Teoria do conceito. **Ciência da Informação**, Rio de Janeiro, v. 7, n. 2, p. 101-107, 1978. Disponível em: <https://revista.ibict.br/ciinf/article/view/115>. Acesso em: 14 mar. 2023.

FONSECA, M. O. K. As estruturas de produção de conhecimento arquivístico quadros em movimento. **Arquivo & Administração**, Rio de Janeiro, v. 7, n. 1, 2008.

FRAWLEY, W. J.; PIATETSKY-SHAPIO, G.; MATHEUS, C. J. Knowledge Discovery in Databases: an overview. **AI Magazine**, Palo Alto, v. 13, n. 3, p. 57-70, 1992. Disponível em <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1011>. Acesso em: 5 maio 2023.

MAIA, M. E.; FERREIRA, D. S.; BARRANCOS, J. E. Revista Analisando em Ciência da Informação: análise bibliométrica da produção científica em Arquivologia. *In*: Congresso Nacional de Arquivologia - CNA, 8., João Pessoa. **Anais eletrônicos [...]** Revista Analisando em Ciência da Informação - RACIn, João Pessoa, 6 (especial), p. 729-744, 2018.

MARÍN AGUDELO, S. A. Estado de la producción científica en Archivística y archivos en América Latina 2000-2009. Una aproximación. **Rev. Interam. Bibliot. Medellín**, Colombia, v. 34, n. 3, p. 257-269, 2011. Disponível em <http://www.scielo.org.co/pdf/rib/v34n3/v34n3a2.pdf>. Acesso em: 5 maio 2023.

MELO, K. I. Base de Dados em Arquivística: a produção científica brasileira. *In*: Simpósio Internacional de Arquivos. São Paulo (SP). **Anais eletrônicos [...]** São Paulo, Eventus, 2021. Disponível em: <https://www.even3.com.br/anais/simposiointernacionaldearquivos/292347-base-de-dados-em-arquivistica—a-producao-cientifica-brasileira/>. Acesso em: 5 maio 2023.

MELO, K. I., SOUZA, S. L. Base de Dados em Arquivística: preservando a produção científica brasileira. *In*: Reunião Brasileira de Ensino e Pesquisa em Arquivologia, VII. Rio de Janeiro. **Anais eletrônicos [...]** Rio de Janeiro, Even3, 2022. Disponível em: <https://www.even3.com.br/anais/simposiointernacionaldearquivos/292347-base-de-dados-em-arquivistica—a-producao-cientifica-brasileira/>. Acesso em: 5 maio 2023.

MELO, K. I.; PAIVA, D. Base de Dados em Arquivística (BDA): fonte de pesquisa referencial no Brasil. *In*: Encontro de Arquivos, Bibliotecas e Museus: do sincretismo à integração, II. Porto. **Anais [...]**, 2022.

MILLAR, L. On the crest of a wave: transforming the archival future. **Archives and Manuscripts: The Journal of the Australian Society of Archivists**, Crows Nest, v. 45, n. 2, p.

59-76, 2017. Disponível em: <https://doi.org/10.1080/01576895.2017.1328696>. Acesso em: 5 maio 2023.

MELO, K. I., SOUZA, S. L.; PAIVA, D. Construção da Base de Dados em Arquivística: uma ferramenta de pesquisa. **Revista Participação** - UnB, n° 38, p. 22-31, 2022. Disponível em <https://periodicos.unb.br/index.php/participacao/article/view/46805>. Acesso em: 5 maio 2023.

SAYÃO, L. F. Bases de dados: a metáfora científica. **Ciência da Informação**. Brasília, v. 25, n. 3, 1996. Disponível em: <http://revista.ibict.br/ciinf/article/view/629>. Acesso em: 5 maio 2023.

SILVA, A. P.; REGO-PIVA, L. M.; GUIMARÃES, J. A. C. Análise de domínio: um estudo nos anais da Reunião de Ensino e Pesquisa em Arquivologia (REPARQ). In: **A pesquisa e o ensino em arquivologia: perspectivas na era digital**. Belém: VI Reparq, 2019. p. 9-19. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/202102/001106584.pdf?sequence=1>. Acesso em: 5 maio 2023.

TARRÉ ALONSO, B.; MENA MUGICA, M. M. Análisis epistemológico en revistas del campo de la Archivística: Archivaria, Archival Science, Tábula, The American Archivist, Archives and Records, y Acervo durante el período 2009-2014. **Bibliotecas Anales de Investigación**. La Habana. v. 12, n. 1, p. 3-9, ene-jun. 2016. Disponível em: <http://revistas.bnjm.cu/index.php/BAI/article/view/156/166>. Acesso em: 5 maio 2023.

TERRA, J. C. C. *et al.* Taxonomia: elemento fundamental para a gestão do conhecimento. **Biblioteca Terra Forum Consultores**, 2005. Disponível em: <http://pessoal.utfpr.edu.br/mansano/arquivos/taxonomia.pdf>. Acesso em: 9 mar. 2023.

150

Recebido/ Received: 21/03/2023
Aceito/ Accepted: 27/03/2023 Pu-
blicado/ Published: 30/04/2023