



24° ENANCIB
Encontro Nacional de Pesquisa em Ciência da Informação
Perspectivas Contemporâneas na Ciência da Informação
• Vitória - ES • Ancib • PPGCI/UFES



XXIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXIV ENANCIB

ISSN 2177-3688

GT2 – Organização e Representação do Conhecimento

**CONSTRUÇÃO DE REPOSITÓRIOS DIGITAIS INSTITUCIONAIS ENDEREÇADOS A
PLATAFORMAS DE AGREGAÇÃO DE DADOS CULTURAIS: UMA EXPERIÊNCIA COM O ACERVO
DO CONSELHO ESTADUAL DE CULTURA DO ESPÍRITO SANTO**

***CONSTRUCTION OF INSTITUTIONAL DIGITAL REPOSITORIES FOR CULTURAL DATA
AGGREGATION PLATFORMS: AN EXPERIENCE WITH THE COLLECTION OF THE STATE
COUNCIL OF CULTURE OF ESPÍRITO SANTO***

Daniela Lucas da Silva Lemos – Universidade Federal do Espírito Santo (UFES)

Dirceu Flávio Macedo – Universidade Federal do Espírito Santo (UFES)

Luiz Carlos da Silva – Universidade Federal do Espírito Santo (UFES)

Abeil Coelho Júnior – Universidade Federal do Espírito Santo (UFES)

Modalidade: Trabalho Completo

Resumo: Instituições dedicadas ao patrimônio cultural utilizam a digitalização como ferramenta essencial para gerenciar, preservar e promover seus acervos online, visando democratizar o conhecimento cultural na rede. Os repositórios digitais institucionais desempenham um papel crucial nesse contexto, sendo adotados tanto por entidades públicas quanto privadas. Esta pesquisa tem como objetivo destacar as melhores práticas na organização da informação e do conhecimento aplicadas na construção de um repositório digital institucional, com ênfase no processo de modelagem da base de dados envolvida. Classificada como pesquisa aplicada, de abordagem qualitativa, exploratória e descritiva, o estudo foca no acervo cultural e arquivístico do Conselho Estadual de Cultura do Estado do Espírito Santo. Utilizou-se, para tal, um modelo metodológico composto por quatro fases integradas e interdependentes para construir um repositório digital institucional destinado a difundir seus dados com qualidade na plataforma de agregação de dados culturais MEDIATECA Capixaba. O resultado foi um repositório institucional digital com uma estruturação adequada frente a princípios de catalogação propostos por organismo internacional de referência, associados às tarefas dos usuários em encontrar, identificar, selecionar, acessar, navegar e explorar o conteúdo digital, tanto no repositório institucional quanto na plataforma MEDIATECA. Recomenda-se a produção de um plano de gestão de dados para as políticas de governança da Secretaria de Estado da Cultura, garantindo a qualidade e sustentabilidade das informações nos repositórios digitais institucionais e na própria MEDIATECA Capixaba.

Palavras-chave: Organização da informação e do conhecimento; Padrões de documentação; Repositórios digitais institucionais; Plataforma de agregação de dados culturais; MEDIATECA Capixaba.

Abstract: Institutions dedicated to cultural heritage leverage digitization as a vital tool for managing, preserving, and promoting their collections online, aiming to democratize cultural knowledge on the web. Institutional digital repositories play a crucial role in this context, being adopted by both public and private entities. This research aims to highlight best practices in information and knowledge organization applied to the construction of an institutional digital repository, with a focus on the database modeling process involved.

Classified as applied research with a qualitative, exploratory, and descriptive approach, the study focuses on the cultural and archival collection of the State Council of Culture of Espírito Santo. A methodological model comprising four integrated and interdependent phases was employed to construct an institutional digital repository designed to disseminate its data with quality on the MEDIATECA Capixaba cultural data aggregation platform. The result was an institutional digital repository with a structure aligned with cataloging principles proposed by an international reference body, facilitating users' tasks in finding, identifying, selecting, accessing, navigating, and exploring digital content, both in the institutional repository and on the MEDIATECA platform. It is recommended to produce a data management plan for the governance policies of the State Secretariat of Culture, ensuring the quality and sustainability of information in institutional digital repositories and in MEDIATECA Capixaba itself.

Keywords: Information and knowledge organization; Documentation standards; Institutional digital repositories; Cultural data aggregation platform; MEDIATECA Capixaba.

1 INTRODUÇÃO

Instituições do patrimônio cultural têm utilizado a digitalização como uma ferramenta crucial para gerir, conservar e divulgar o seu patrimônio online, de modo a democratizar o conhecimento cultural na rede (Soave; Lemos, 2022). No entanto, o processo de produção, organização e disseminação de informação em instituições relacionadas ao campo da cultura digital ainda perpassa por obstáculos consideráveis (Martins *et al.*, 2022; Lemos; Coelho Junior, 2023) acerca da busca por conteúdo de qualidade, incluindo objetos digitais e seus metadados, endereçado às bases de dados.

Nesse contexto, bibliotecas, museus, arquivos e outros espaços culturais consideram bases de dados como produtos ou fontes de informação cruciais para a sociedade quando adotados para realizar a mediação entre documentos e públicos (Lemos; Martins; Souza, 2023). Para tal, a partir do momento em que há necessidade de lidar com grandes volumes de dados oriundos dessas instituições, torna-se fundamental o uso de tecnologias de informação e comunicação (TICs) capazes de oferecer serviços sofisticados para organização, agregação, gestão, preservação, publicação, recuperação, acesso, uso e reúso de informações com qualidade (Sayão; Sales, 2022).

Os repositórios digitais institucionais são considerados TICs que se adentram nessa perspectiva, sendo utilizados por instituições públicas e privadas para gestão, preservação e disseminação do conhecimento produzido pelas organizações. Marcondes e Sayão (2009, p. 10) destacam que os “Repositórios institucionais são entendidos como elementos de uma rede ou infraestrutura informacional [...] destinados a garantir a guarda, preservação a longo prazo e, fundamentalmente, o livre acesso à produção científica de uma dada instituição”.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

Entende-se também, que podemos reconhecer os repositórios institucionais como um tipo de repositório digital (Martins; Lemos; De Andrade, 2021), qual seja um sistema de armazenamento de conteúdo digital que facilita o gerenciamento, a preservação e o acesso a diversos tipos de materiais digitais com uma variedade de conteúdo.

Numa perspectiva de construção de repositórios digitais institucionais e tecnologias de agregação, surge a plataforma de agregação de dados culturais (Pereira; Lemos, 2024) MEDIATECA Capixaba¹, fruto de um projeto de pesquisa inovador e interdisciplinar envolvendo áreas como Arquivologia, Biblioteconomia, Museologia e Ciência da Informação (CI), com potencial de transformar o patrimônio cultural brasileiro em fonte de inovação e desenvolvimento da sociedade.

Assim, a MEDIATECA, sob a gestão da Secretaria de Estado da Cultura (SECULT), pretende tornar públicos e de forma integrada os acervos digitais das instituições culturais do Estado do Espírito Santo, com os quais, atualmente, participam como provedores de conteúdo histórico e cultural a Biblioteca Pública do Estado, o Arquivo Público do Estado, o Museu de Arte, a Galeria Homero Massena, o Museu do Colono, a Orquestra Sinfônica, o Palácio Anchieta, o Conselho Estadual de Cultura, o Funcultura e a Rádio TV.

Ademais, a MEDIATECA é constituída atualmente por um total de 4.847 itens que compõem as fontes de informação representadas por acervos bibliográficos, museológicos e arquivísticos da plataforma, evidenciando a sua característica heterogênea em formatos e tipos documentais, incluindo documentos, fotografias, vídeos, músicas, matérias de jornal, artes gráficas, livros, periódicos, obras de arte e filmes produzidos no Espírito Santo.

Contudo, de modo que as fontes de informação possam ser disponibilizadas na plataforma MEDIATECA, permitindo encontrabilidade, identificação, seleção, aquisição, reúso, além de possibilitar navegação e exploração do conteúdo digital (International Federation of Library Associations and Institutions, 2016), torna-se necessária a inserção dessas fontes em tecnologias de repositório digital. Logo, para cada instituição provedora de conteúdo, os objetos digitais e seus metadados precisam ser produzidos e/ou obtidos e organizados no TAINACAN² (*software* de repositório digital utilizado no projeto de pesquisa) a partir da documentação legada dessas instituições que gerenciam, digitalizam e catalogam seus documentos em seus ciclos de vida.

¹ <https://midiateca.es.gov.br/site/>. Home-page da plataforma digital MEDIATECA Capixaba

² <https://tainacan.org/>. Home-page do portal do TAINACAN

Para o contexto da presente pesquisa, dentre as instituições provedoras de conteúdo da plataforma MEDIATECA Capixaba, optou-se por relatar a experiência com o acervo do Conselho Estadual de Cultura³ (CEC) pela sua importância à SECULT em salvaguardar os documentos permanentes, garantindo sua longevidade e preservação como patrimônios culturais da sociedade do Espírito Santo por meio do acesso às informações consideradas ostensivas; e pelos desafios impostos na representação de um conjunto de documentos diversificados para fins de organização em repositório institucional para servir de fonte de informação qualificada à plataforma de agregação. Dentre os desafios, vale destacar a proposição de uma modelagem conceitual da informação, útil à organização e representação das descrições que precisarão ser identificadas e padronizadas para o processo de catalogação dos recursos de informação do acervo cultural e arquivístico envolvido.

Ressalta-se que a padronização é fruto de elementos preponderantes discutidos na CI, especialmente na área de Organização e Representação da Informação e do Conhecimento (Lemos; Martins; Souza, 2023), na qual nos traz contribuições significativas à ação de curadoria digital quando fundamenta que a informação para ser organizada, precisa, *a priori*, ser descrita com padrões de documentação, de modo a viabilizar preservação, acesso, recuperação, agregação e reúso de objetos digitais disponíveis na rede.

Padrões de documentação geralmente são usados para apoiar práticas de catalogação e são direcionados à crescente digitalização de coleções de objetos culturais, que necessitam padronizar aspectos únicos de coleções culturais e de seus desdobramentos digitais (Gilliland, 2016). Nesse sentido, os padrões de documentação com os quais se pretende usar no projeto MEDIATECA Capixaba são fundamentados em aportes teórico-metodológicos das catalogações descritivas (Joudrey; Taylor; Miller, 2015; Gilliland, 2016) e de assunto (Lancaster, 1986), se atendo à criação e ao uso de metadados (e seus padrões), de vocabulários controlados e de regras de catalogação pertinentes a cada perfil de aplicação institucional (Baker; Nilsson; Johnston, 2008). Essas últimas determinam como elaborar a descrição de um recurso de informação e os pontos de acesso, tornando-se práticas essenciais na padronização, na descrição e, portanto, na agregação semântica de recursos de informação.

³ <https://secult.es.gov.br/conselho-estadual-de-cultura-2>. Página do CEC no portal da SECULT/ES

Diante ao exposto, a presente pesquisa busca responder a seguinte questão: *como ampliar a capacidade de busca e recuperação de conteúdo digital, tanto no repositório institucional do Conselho Estadual de Cultura do Espírito Santo quanto na plataforma MEDIATECA Capixaba?* Logo, o objetivo desta pesquisa é apresentar as boas práticas em organização da informação e do conhecimento usadas na construção do repositório digital do CEC, dando destaque ao processo de modelagem da base de dados envolvida.

2 PROCEDIMENTOS METODOLÓGICOS

A presente pesquisa é classificada como sendo de natureza aplicada, de abordagem qualitativa, e de caráter exploratório e descritivo, envolvendo o acervo cultural e arquivístico do CEC, um dos provedores de conteúdo da plataforma de agregação de dados culturais MEDIATECA Capixaba.

O CEC é um órgão consultivo vinculado à SECULT, responsável pela normatização, deliberação e atendimento às demandas da sociedade, integrando as ações de Política Cultural do Estado do Espírito Santo. Embora não seja uma instituição arquivística, o CEC possui um arquivo corrente, assim como outras instituições públicas e privadas, e desfruta de certa autonomia, produzindo documentos administrativos e coleções de documentos históricos e culturais.

A motivação para o desenvolvimento da preservação dos documentos e da informação do CEC está na composição do seu acervo, que inclui atas, fotografias, recortes jornalísticos e outras mídias, tanto digitais quanto físicas. Esses objetos foram acumulados como documentos arquivísticos e coleções durante os processos administrativos. Além disso, o acervo está enriquecido pelos dados dos tombamentos e pelos relatos em atas que narram a trajetória histórica do Conselho. Isso evidencia a necessidade de manter esses documentos em um ambiente digital preservado e redundante, alinhado com as técnicas, princípios e teorias arquivísticas.

Postas as considerações iniciais sobre o ambiente de informação da presente pesquisa, as etapas metodológicas para a sua condução podem ser visualizadas na Figura 1. Nela, encontram-se quatro etapas integradas e interdependentes voltadas à construção do repositório digital institucional do CEC.

Na etapa de Diagnóstico de Maturidade do Acervo (bloco verde), foram realizadas visitas à sede do CEC para compreender seu ambiente informacional junto aos usuários-chave e colaboradores do mesmo. Com base nas observações e dados coletados, foi feita uma avaliação diagnóstica da maturidade de tal ambiente à luz do manual para construção de acervos digitais do Instituto Brasileiro de Museus (Instituto Brasileiro de Museus, 2020), incluindo os indicadores política institucional (procedimentos e regras institucionais), pessoas envolvidas, tecnologias e sistemas adotados, e documentação do acervo (formas de gestão e organização da informação).

Figura 1 - Modelo metodológico para a construção do repositório digital institucional do CEC



Fonte: elaborado pelos autores (2024)

Em sequência, têm-se as etapas de Inventário e Digitalização (bloco azul). O inventário envolveu a identificação e a descrição das espécies documentais presentes no acervo, o que permitiu a criação de uma base de dados inicial, essencial para as etapas subsequentes. A forma de descrição e a determinação de alguns dados metadados para a composição do esquema de inventário foram subsidiadas por teorias e técnicas oriundas da Nobrade, a Norma Brasileira de Descrição Arquivística (Brasil, 2006). A justificativa para o uso da Nobrade estaria no fato de o CEC possuir um Serviço de Arquivo que necessita dos cuidados previstos pela área de Arquivologia, visando a sua organização, incluindo classificação, avaliação, descrição e preservação dos documentos, além do armazenamento e acondicionamento adequados.

Concomitante, tem-se a digitalização dos documentos, um processo que envolveu a conversão dos itens físicos para formatos digitais, utilizando equipamentos especializados para garantir a preservação da qualidade e a integridade dos materiais originais. Esta ação

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

foi realizada de acordo com as normas legais estabelecidas pelo Decreto nº 10.278/2020⁴ (Lei da Digitalização de Documentos), que regulamenta a Lei nº 12.682/2012⁵. Esta legislação garante que documentos digitalizados tenham os mesmos efeitos jurídicos que os documentos originais. Adicionalmente, a Lei nº 13.874/2019⁶, que institui a Declaração de Direitos de Liberdade Econômica, complementa a legislação de digitalização.

As técnicas específicas de digitalização foram pautadas pelo Conselho Nacional de Arquivos⁷ (CONARQ), que estabelece diretrizes e orientações para os procedimentos técnicos de digitalização de documentos públicos e privados para os órgãos e entidades do Sistema Nacional de Arquivos. Nesse sentido, a digitalização foi realizada com uma resolução mínima de 400 dpi, utilizando modos de cor variados entre os padrões RGB, tons de cinza e preto e branco (P&B), conforme necessário. O formato de saída para documentos textuais foi PDF/A, enquanto para as fotografias foi o JPG. Os cartazes foram salvos em uma mescla de PDF, JPG e PNG. Além disso, foram geradas versões em formato TIFF para todas as digitalizações realizadas pela equipe, visando a preservação a longo prazo. Ademais, aplicaram-se técnicas de Reconhecimento Óptico de Caracteres (*Optical Character Recognition* - OCR, no inglês) para tornar os textos pesquisáveis em arquivos PDF, facilitando a recuperação de informação em texto completo. Ao término da aplicação desses estágios, foi possível assegurar a integridade e a fidelidade dos documentos originais do Conselho Estadual de Cultura.

O bloco laranja pode ser considerado a espinha dorsal do modelo metodológico, pois projeta a base de dados do repositório digital institucional, por meio da qual se intenciona dados curados para transformação em fontes de informações qualificadas e destinadas à plataforma de agregação envolvida.

A etapa de Modelagem do Esquema de Metadados foi assistida por meio da técnica de criação de Matriz de Restrições (Baker; Nilsson; Johnston, 2008), em que se integra e prevê um conjunto de registros de metadados que são instâncias válidas para um dado perfil de aplicação, neste caso, o acervo cultural e arquivístico do CEC. Para tal, na ação de determinação dos elementos de metadados a compor a matriz de restrições, foram

⁴ https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/decreto/d10278.htm. Decreto nº 10.278/2020.

⁵ https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12682.htm. Lei nº 12.682/2012.

⁶ https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/l13874.htm. Lei nº 13.874/2019.

⁷ https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/Diretrizes_digitalizacao__2021.pdf. Página contendo a documentação do Conarq, que versa sobre as diretrizes de digitalização.

registrados e destacados os metadados que pudessem ser alinhados (mapeados) com o padrão Dublin Core, modelo central da plataforma de agregação MEDIATECA Capixaba, conjuntamente com as regras de cardinalidade, as obrigatoriedades, os filtros de busca, os tipos de dados (se são numéricos, data, textuais, categóricos, entre outras), se faz uso de vocabulário controlado (se sim, indicar qual), se faz uso de regras de catalogação (se sim, indicar qual), orientações de uso, e outras características importantes à modelagem de informação em contexto digital. Do mesmo modo como foi feito para os inventários, para cada modelagem produzida, criou-se uma planilha de dados correspondente, a qual passou por uma validação de requisitos com os gestores responsáveis da SECULT.

A etapa Processo de Catalogação envolveu a produção de planilhas de dados, fundamentadas nos elementos de metadados registrados na matriz de restrições, com o propósito de catalogar inicialmente os itens inventariados. Torna-se importante destacar que, para cada item catalogado, foi registrado um valor de dado referente ao link de localização da sua respectiva mídia digital, produzida no processo de digitalização.

Já para a etapa associada ao Tratamento dos Dados, foram usadas ferramentas da área de Ciência de Dados para rotinas de pré-processamento de dados, incluindo Knime, Open Refine e scripts em python (Shintaku; Macêdo; Marin, 2023) para viabilizar a importação desses dados na coleção digital (base de dados) criada no Tainacan do CEC à luz da matriz de restrições modelada.

A última etapa Disponibilização dos Dados em Repositório Institucional (bloco roxo) dimensionou a continuidade do processo de catalogação no próprio repositório Tainacan. Uma vez os dados organizados e persistidos em repositório digital do CEC, incluindo arquivos de mídias e seus metadados, os mesmos podem ser coletados de forma automática, em períodos de tempo específicos, por meio de suas *Application Programming Interface* (APIs), e agregados em uma única base de dados da plataforma MEDIATECA Capixaba. Em suma, a catalogação sistemática e consistente a partir de uma matriz de restrições assegurou a organização e a eficiência na recuperação agregada dos documentos do CEC na plataforma.

3 RESULTADOS

Considerando o objetivo desta pesquisa em apresentar as boas práticas em organização da informação e do conhecimento empregadas na construção do repositório digital do CEC, dando destaque ao processo de modelagem da base de dados envolvida, a

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

apresentação dos resultados aqui presentes se concentrou na etapa em que se fez maior uso de padrões de documentação para fins de qualificar a base de dados implementada, qual seja o bloco laranja da Figura 1.

Uma vez realizadas as etapas do conhecimento atual do ambiente de informação para o levantamento de requisitos necessários à concepção do modelo de metadados da base de dados do CEC, bem como o inventário e a digitalização dos documentos identificados e selecionados como pertinentes ao procedimento, a matriz de restrições foi implementada a partir de 7 categorias de metadados, as quais representam as espécies documentais presentes no acervo, a saber: dados gerais; ata; cartaz; fotografia; recorte de jornal; resolução; termo; e ficha de bens culturais.

A categoria “dados gerais” traz elementos de metadados comuns a todas as espécies documentais do acervo. Para as demais categorias, foi elaborado um perfil de aplicação contendo elementos de metadados específicos a cada espécie documental, o que na tecnologia de repositório Tainacan torna-se possível condicionar este tipo de metadado à abertura de uma seção personalizada para catalogar o documento a partir da categoria selecionada em tal processo.

A Figura 2 mostra parte da matriz de restrições com a relação dos metadados e seus perfis de aplicação, evidenciando as decisões de modelagem para cada elemento, tornando, assim, uma representação da informação e do conhecimento diferenciada, visando, logo, melhorar os sistemas de catalogação, classificação, indexação, busca, navegação, agregação e recuperação, de forma a obter resultados mais significativos aos usuários no contexto de uma plataforma de agregação de dados culturais.

Figura 2 – Fragmento da matriz de restrições para a modelagem da coleção do CEC

1	Categoria	Metadado	Equivalente Dublin Core	Descrição metadado	Orientação usuário	Composto?	Composto Elemento	Condicional configuração	Público?	Obrigatório?	Tipo valor Tainacan	Filtro de busca?	Multivalorado?	Vocabulário controlado?	Termos controlados de catalogação	Equivalente guia de catalogação	Regras de catalogação
3	Dados gerais	Espécie Documental	Type	Espécie documental da i Divisão de gênero do		Não	-	Não	Sim	Sim	Lista de Selc	Sim	Não	Sim	Ata; Cartaz; f	Nobrade	1.5 Dimensão e suporte. 1.5.5. Reg
4	Dados gerais	Código de Referência	Identifier	Código de referência da Apresenta uma iden		Não	-	Não	Não	Sim	Textual	Não	Não	Não	Nobrade	1.1 Código de referência. Regra: 11-	
5	Dados gerais	Nível de Descrição	Description	Nível de descrição da Permite ao usuário id		Não	-	Não	Sim	Sim	Lista de Selc	Não	Não	Sim	Nível 5	Nobrade	1.4 Nível de descrição. 1.4.1. Regra:
6	Dados gerais	Título	Title	Título da unidade de des Identifica nominalme		Não	-	Não	Sim	Sim	Textual	Não	Não	Não	NA	Nobrade	1.2 Título. Regra: Deve-se registrar-
7	Dados gerais	Descrição	Description	Describe um ou mais as Descrição resumida d		Não	-	Não	Sim	Sim	Textual	Não	Não	Não	NA	Nobrade	3.1 Âmbito e conteúdo. Regra: Infor
8	Dados gerais	Dimensão Lógica (kb)	Dimension	Dimensão lógica da unid Identifica a dimensã		Não	-	Não	Sim	Sim	Númerico	Não	Não	NA	Nobrade	1.5 Dimensão e suporte. Regra: Rag	
9	Dados gerais	Produtor	Creator	Nome da entidade prod Identifica a entidade		Não	-	Não	Sim	Não	Textual	Sim	Não	Não	NA	Nobrade	2.1 Nome(s) do(s) produtor(es). Reg
10	Dados gerais	Ano	Date	Ano da unidade de desc Informa o ano em qu		Não	-	Não	Sim	Não	Númerico	Sim	Não	Não	NA	Nobrade	1.3 Data. 1.3.2.1 - Data Crônica. Reg
11	Dados gerais	Data	Date	Data da unidade de desc Informa a data em qu		Sim	-	Não	Sim	Não	Data	Não	Não	Não	NA	Nobrade	1.3 Data. 1.3.2.1 - Data Crônica. Reg
12	Dados gerais	Idioma	Language	Idioma da unidade de id Identifica o(s) idioma		Não	-	Não	Sim	Sim	Lista de Selc	Não	Não	Sim	Português	Nobrade	4.3 Idioma. Regra: Informe idioma!
13	Dados gerais	Localidade	Coverage	Local geográfico da unid Indica o Estado, Cida		Sim	-	Não	Sim	Não	Taxonomia	Sim	Sim	TX Localidad	Nobrade	3.1 Âmbito e conteúdo. 3.1.2. Regri	
14	Dados gerais	Formato	Format	Formato da unidade de i Indica o formato digit		Não	-	Não	Sim	Sim	Lista de Selc	Sim	Não	Sim	JPG; PDF; PN	Nobrade	1.5 Dimensão e suporte. Regra: Rag
15	Dados gerais	Condições de Acesso	Rights	Condições de acesso da Fornece informação i		Não	-	Não	Sim	Sim	Lista de Selc	Sim	Não	Sim	Domínio Púb	Nobrade	4.1 Condições de acesso -4.1.1. Reg
16	Dados gerais	Coleção	Source	Coleção da unidade de i Informa a coleção em		Não	-	Não	Sim	Sim	Lista de Selc	Não	Não	Sim	Acervo Digit	Nobrade	2.3 História arquivística. 2.3.2. Regr
17	Dados gerais	Instituição Custodiadora	Source	Instituição custodiadora Informa a instituiçã		Não	-	Não	Sim	Sim	Lista de Selc	Não	Não	Sim	Conselho Est	Nobrade	1.1 Código de referência. 1.1.1. Reg
18																	
19	Ata	Tipo de Sessão	Type	Tipo de sessão da ata. Auxilia o usuário a lo		Não	-	Não	Sim	Não	Lista de Selc	Sim	Não	Sim	Assembleia;	Nobrade	1.5 Dimensão e suporte. 1.5.5. Reg
20	Ata	Número da Ata	Identifier	Numeração oficial das at Permite ao usuário lo		Não	-	Não	Sim	Não	Númerico	Não	Não	NA	Nobrade	3.3 Incorporações. Regra: Registre i	
21	Ata	Número revisada da Ata	Identifier	Numeração revisada das Permite ao usuário lo		Não	-	Não	Sim	Não	Númerico	Não	Não	NA	Nobrade	2.3 Incorporações. Regra: Registre i	
22	Ata	Número de Páginas	PageCount	Número total de páginas indica ao usuário o n		Não	-	Não	Sim	Não	Númerico	Não	Não	NA	Nobrade	2.3 Incorporações. Regra: Registre i	
23	Ata	Presidente do CEC	Contributor	Nome do presidente em Permite ao usuário ef		Não	-	Não	Sim	Não	Taxonomia	Sim	Não	Sim	VC: Pessoas f	Nobrade	2.2 História administrativa / biograf
24	Ata	Vice-Presidente do CEC	Contributor	Nome do vice-president Permite ao usuário ef		Não	-	Não	Sim	Não	Taxonomia	Sim	Não	Sim	VC: Pessoas f	Nobrade	2.2 História administrativa / biograf
25	Ata	Presidente da Sessão	Contributor	Nome do presidente da Permite ao usuário ef		Não	-	Não	Sim	Não	Taxonomia	Sim	Não	Sim	VC: Pessoas f	Nobrade	2.2 História administrativa / biograf
26	Ata	Secretário(a) Executivo(a)	Contributor	Nome do responsável pe Permite ao usuário ef		Não	-	Não	Sim	Não	Taxonomia	Sim	Não	Sim	VC: Pessoas f	Nobrade	2.2 História administrativa / biograf
27	Ata	Quorum	Description	Existência ou não de qu Informa se uma dete		Não	-	Não	Sim	Não	Booleano	Sim	Não	Sim	Não	Nobrade	3.3 Incorporações. Regra: Registre i
28	Ata	Especificação	Type	Especifica se é original ou Ajuda a classificar e n		Não	-	Não	Sim	Não	Lista de Selc	Sim	Não	Sim	Original; Cóp	Nobrade	1.5 Dimensão e suporte. 1.5.5. Reg: 1
29	Ata	Exatidão	Description	Exatidão amostrada. Fornece uma lista dos		Não	-	Não	Sim	Não	Lista de Selc	Sim	Não	Sim	Manuscrito; i	Nobrade	3.3 Incorporações. Regra: Registre i

Fonte: elaborado pelos autores (2024)

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

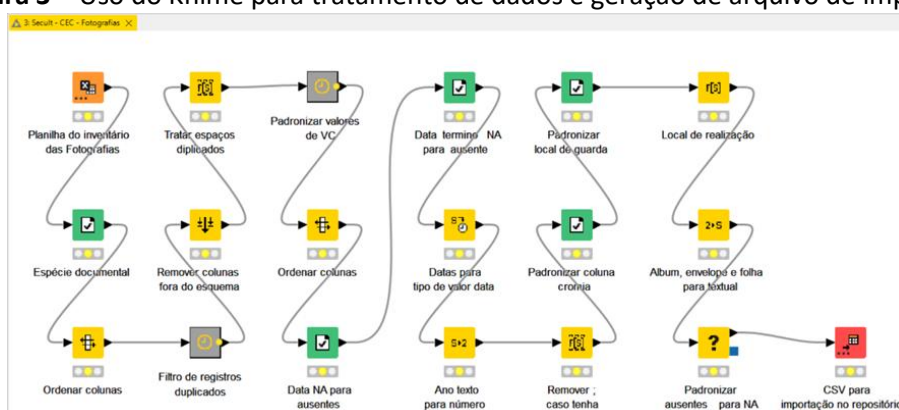
Também digno de nota na Figura 2 são os padrões de documentação usados na organização e representação dos documentos do CEC, vistos como estratégia de tratamento documental para fins de normalização, qualidade e intercâmbio de descrições entre sistemas de informação, pessoas e instituições (Gilliland, 2016; Zeng, 2019). Assim, as atividades de catalogação foram orientadas por regras de catalogação da Nobrade, por meio das quais são indicados elementos de metadados específicos para o contexto arquivístico, e determinado a obrigatoriedade ou não da entrada do dado nos elementos de metadados especificados, a exemplo do metadado “nível de descrição”, qual seja um descritor do tipo obrigatório que determina uma posição da unidade de descrição em uma estrutura hierarquizada de organização de um acervo. Assim, para todo o acervo do CEC, foi utilizado o “nível de descrição cinco”, seguindo a classificação da Nobrade. O nível cinco significa que a descrição foi realizada ao nível dos itens documentais individuais (BRASIL, 2006). Para além, as regras de catalogação adotadas na matriz de restrições permitem fazer uso conjugado com vocabulários controlados (listas de seleção, taxonomias e autoridades) e padrões de metadados (Dublin Core como um modelo de referência), visando obter informações normalizadas e consistentes (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2016) para busca, identificação, seleção, acesso, navegação e reúso eficientes por parte da comunidade usuária da plataforma. A seguir são apresentados alguns dos vocabulários controlados modelados na matriz:

- Autoridade de pessoas: organização dos nomes de presidentes, vice-presidentes, e secretários executivos do Conselho.
- Autoridade de locais: organização hierárquica de termos para valores de dados associados ao metadado localização do item, disponibilizando informação do estado, cidade e distrito, em um único metadado. Também utilizado nos recortes de jornal para criar uma relação hierárquica entre o nome do jornal e sua localidade.
- Listas de seleção: organização de pequenos grupos de termos em alguns metadados para maior controle sobre os valores de dados selecionados na catalogação.

A partir do ambiente de desenvolvimento da tecnologia de repositório, pôde-se criar o esquema de dados à luz da matriz de restrições, cuja documentação se torna excelente instrumento de gestão de projetos para a construção e manutenção de repositórios digitais institucionais.

Finalmente, tendo o esquema de dados criado no repositório, tornou-se possível tratar os dados das planilhas de catalogação, elaboradas e disponibilizadas pela equipe de catalogação alocada no projeto, e importar esses dados normalizados e padronizados na base de dados do Tainacan. A Figura 3 exemplifica o uso da tecnologia Knime para um fluxo de trabalho (*workflow*, em inglês) criado para tratamento de dados em planilhas, incluindo correção ortográfica, limpeza, normalização, criação de arquivos de autoridade, taxonomias, listas de seleção, e tratamento das mídias digitais.

Figura 3 – Uso do Knime para tratamento de dados e geração de arquivo de importação



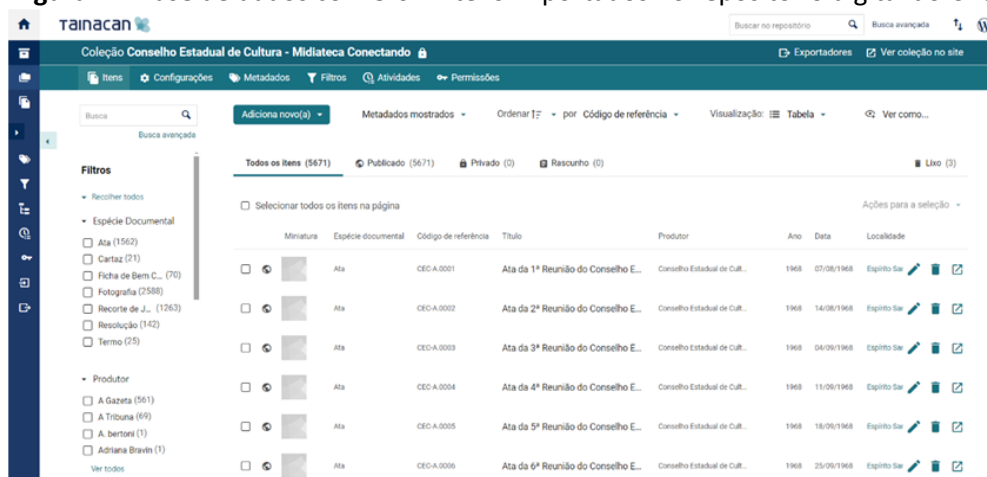
Fonte: elaborado pelos autores (2024)

Para a fase de importação, foi gerado um arquivo .CSV com os dados já tratados e devidamente mapeados com os elementos de metadados correspondentes ao esquema e dados criado no Tainacan. Isso garantiu que as informações como título, autor, data, descrição e outros fossem corretamente associadas a cada elemento do esquema.

Em sequência, o sistema do repositório é configurado, incluindo a criação de filtros para seleção de itens documentais, as páginas são personalizadas e os metadados, junto com as mídias digitais, são carregados para a implementação efetiva da base de dados no repositório digital (Figura 4).

Após a importação, ocorreram ações de validação e correção. Isso envolveu verificar se todos os dados foram carregados corretamente, resolver eventuais erros de importação e garantir que os metadados estivessem completos e precisos. Uma vez validados, os itens do acervo são indexados pelo sistema, tornando-os facilmente pesquisáveis e acessíveis aos usuários por meio da interface do Tainacan.

Figura 4 – Base de dados com 5.671 itens importados no repositório digital do CEC



The screenshot shows the Tainacan digital repository interface. The top navigation bar includes 'Tainacan' and 'Coleção Conselho Estadual de Cultura - Midiateca Conectando'. Below the navigation bar, there are search and filter options. The main content area displays a list of items with the following columns: Minuturas, Espécie documental, Código de referência, Título, Produtor, Ano, Data, and Localidade. The table shows six items, all of which are 'Ata' documents from the 'Conselho Estadual de Cultura' produced between 1968 and 1969. The interface also includes a sidebar with filters for 'Espécie Documental' and 'Produtor'.

Minuturas	Espécie documental	Código de referência	Título	Produtor	Ano	Data	Localidade
<input type="checkbox"/>	Ata	CEC-A.0001	Ata da 1ª Reunião do Conselho E...	Conselho Estadual de Cult...	1968	07/08/1968	Espirito São
<input type="checkbox"/>	Ata	CEC-A.0002	Ata da 2ª Reunião do Conselho E...	Conselho Estadual de Cult...	1968	14/08/1968	Espirito São
<input type="checkbox"/>	Ata	CEC-A.0003	Ata da 3ª Reunião do Conselho E...	Conselho Estadual de Cult...	1968	04/09/1968	Espirito São
<input type="checkbox"/>	Ata	CEC-A.0004	Ata da 4ª Reunião do Conselho E...	Conselho Estadual de Cult...	1968	11/09/1968	Espirito São
<input type="checkbox"/>	Ata	CEC-A.0005	Ata da 5ª Reunião do Conselho E...	Conselho Estadual de Cult...	1968	18/09/1968	Espirito São
<input type="checkbox"/>	Ata	CEC-A.0006	Ata da 6ª Reunião do Conselho E...	Conselho Estadual de Cult...	1968	25/09/1968	Espirito São

Fonte: elaborado pelos autores (2024)

Por fim, torna-se válido ressaltar que os dados no repositório digital do CEC (atuais 5.671 itens) já se encontram aptos a serem colhidos e mapeados para a plataforma de agregação de dados culturais, podendo, assim, se somar aos 4.847 itens que compõem, atualmente, as fontes de informação da plataforma Midiateca Capixaba.

4 DISCUSSÃO

A presente pesquisa traz resultados pertinentes à comunidade de CI quanto aos princípios teórico-metodológicos da organização e representação da informação e do conhecimento (Martins *et al.*, 2022) empregados na produção de serviços de informação úteis à sociedade, especialmente aos profissionais da informação que estão lidando cada vez mais com a criação de repositórios digitais e plataformas de agregação de dados institucionais e governamentais como forma de preservar, difundir e viabilizar o reúso de conteúdo digital na rede (Pereira; Lemos, 2024).

Pode-se destacar a importância do uso de padrões de documentação (Lemos; Martins; Souza, 2023) nas práticas de modelagem de bases de dados, a partir da técnica de criação de perfis de aplicação (Baker; Nilsson; Johnston, 2008) bem como do uso de infraestrutura tecnológica adequada, tal como um repositório digital, na estruturação eficiente do acervo digital do CEC, facilitando a sua adequação aos princípios de catalogação, propostos por organismo internacional de referência (International Federation of Library Associations and Institutions, 2016), associados às tarefas dos usuários em encontrar, identificar, selecionar, acessar, navegar e explorar o conteúdo digital, tanto no repositório institucional quanto na plataforma Midiateca Capixaba. Nesse caminho, destaca-se a criação

e o uso de vocabulários controlados adequados à normalização terminológica dos arquivos de autoridade de nomes e locais pertinentes ao contexto do CEC, contribuindo com a disponibilização de bons filtros de seleção, com buscas e navegações abrangentes e específicas (a partir do uso de taxonomias), trazendo benefícios ao processo de recuperação da informação (Lancaster, 1986).

Adicionalmente, a aplicação de princípios, normas e padrões reconhecidos, como a Nobrade para regras de catalogação, viabilizou a recomendação de uso de elementos de metadados obrigatórios e pertinentes às práticas arquivísticas, bem como outros elementos complementares alinhados ao padrão de metadados Dublin Core, visando uma catalogação descritiva mais consistente dos objetos digitais (Joudrey; Taylor; Miller, 2015; Gilliland, 2016), tendo em vista que tal padrão de metadados é o modelo central usado na plataforma MEDIATECA para fins de agregação de dados comuns frente às suas 15 (quinze) categorias.

Contudo, embora a criação do repositório digital do CEC se manifeste como um marco para a preservação da memória institucional bem como a sua difusão em plataforma aberta de dados de pesquisa histórica e cultural, é essencial adotar modelos de governança responsáveis para assegurar a sua sustentabilidade e que, portanto, vão além de investimentos em infraestrutura tecnológica (Sayão; Sales, 2022). Logo, sem uma organização e gestão eficaz da informação, o estabelecimento de parcerias estratégicas e o comprometimento contínuo dos gestores, a plataforma corre o risco de sofrer com desatualizações e manutenção. Portanto, a implementação de políticas de governança torna-se relevante para assegurar a longevidade e a relevância do repositório institucional e da plataforma que agrega e difunde seus dados.

5 CONSIDERAÇÕES FINAIS

A experiência com a organização e representação da informação e do conhecimento na modelagem e criação da base de dados para a gestão do acervo do Conselho Estadual de Cultura representa um modelo metodológico genérico que pode ser utilizado em outros contextos informacionais, especialmente no campo da cultura. Esse modelo abrange diversas etapas fundamentais, incluindo o Diagnóstico de Maturidade do Acervo, Inventário e Digitalização, Modelagem do Esquema de Metadados, Processo de Catalogação e Tratamento dos Dados, visando fontes de informação qualificadas para seguirem com confiabilidade a plataformas de agregação de dados culturais. Esta metodologia, portanto,

pode ser aplicada a diferentes tipos de instituições culturais que precisam organizar e preservar seus acervos, visando torná-los acessíveis ao público.

O repositório digital do CEC, destacando a sua base de dados, foi criado a partir desse modelo metodológico, dando destaque às boas práticas em organização da informação e do conhecimento dimensionadas em tal modelo (bloco laranja da Figura 1). Logo, seus dados, tratados e normalizados a partir de padrões de documentação, podem ser colhidos e difundidos na plataforma de agregação MEDIATECA Capixaba, possibilitando interoperabilidade com dados de outras instituições de cultura vinculadas, buscas agregadas e abrangentes, e reutilização de metadados e seus objetos digitais para diversas finalidades.

Logo, os resultados alcançados conseguem responder à questão de pesquisa, para a qual se criou uma base de dados padronizada e normalizada para o Conselho Estadual de Cultura do Espírito Santo, provedor de conteúdo da plataforma MEDIATECA Capixaba. Usou-se, para tal, de boas práticas em organização da informação e do conhecimento, incluindo a criação de vocabulários controlados, o emprego de regras de catalogação, e a adoção de padrão de metadados, de modo a ampliar a capacidade de busca e recuperação de conteúdo digital, tanto no repositório institucional quanto na plataforma MEDIATECA Capixaba, alcançando, por conseguinte, o objetivo da pesquisa.

Como trabalhos futuros, sugere-se a proposição de um plano de gestão de dados para as políticas de governança da SECULT, de modo a manter e sustentar a qualidade das fontes de informação nos repositórios digitais institucionais e na própria plataforma de agregação de dados culturais MEDIATECA Capixaba.

REFERÊNCIAS

BAKER, Thomas; NILSSON, Mikael; JOHNSTON, Pete. **The Singapore Framework for Dublin Core Application Profiles**. DCMI, 2008. Disponível em: <http://dublincore.org/documents/singapore-framework/>. Acesso em: 10 jun. 2024.

BRASIL. Conselho Nacional de Arquivos. **NOBRADE: Norma Brasileira de Descrição Arquivística**. Rio de Janeiro: Arquivo Nacional, 2006. 124p. Disponível em: <https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/nobrade.pdf>. Acesso em: 01 mai. 2024.

GILLILAND, Anne Jervois. Setting the Stage. In: BACA, Murta. (ed.). **Introduction to metadata**. 3. ed. Los Angeles: Getty Research Institute, 2016.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

INSTITUTO BRASILEIRO DE MUSEUS. **Acervos digitais nos museus**: manual para realização de projetos. Universidade Federal de Goiás - Brasília, DF: Ibram, 2020.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS (IFLA). **Declaração dos Princípios Internacionais de Catalogação**. Haia, 2016. Disponível em: https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/icp/icp_2016-pt.pdf. Acesso em: 17 abr. 2023.

JOUDREY, Daniel N.; TAYLOR, Arlene G.; MILLER, David P. **Introduction to cataloging and classification**. 11 ed. Santa Barbara: ABC-CLIO, 2015.

LANCASTER, Frederick Wilfrid. **Vocabulary control for information retrieval**. Information Resources Press, 1986.

LEMOS, Daniela Lucas da Silva; COELHO JUNIOR, Abeil. Qualidade de dados em acervos do patrimônio cultural: uma avaliação diagnóstica semiautomática nos objetos culturais sob gestão do Instituto Brasileiro de Museus. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 28, p. 1-22, 2023.

LEMOS, Daniela Lucas da Silva; MARTINS, Dalton Lopes; SOUZA, Renato Rocha. Organização e representação da informação e do conhecimento em contextos informacionais: uma proposta de um modelo teórico-conceitual para a qualidade de objetos culturais digitais. **Fronteiras da Representação do Conhecimento**, Belo Horizonte v.3, n.2, ano III, Editorial, mar. 2023.

MARCONDES, Carlos Henrique; SAYÃO, Luis Fernando. À guisa de introdução: repositórios institucionais e livre acesso. IN: SAYÃO, Luis *et. al.* **Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação**. Salvador: EDUFBA, p. 9-21, 2009.

MARTINS, Dalton Lopes; LEMOS, Daniela Lucas da Silva; DE ANDRADE, Morgana Carneiro. Tainacan e Omeka: proposta de análise comparativa de softwares para gestão de coleções digitais a partir do esforço tecnológico para uso e implantação. **Informação & Informação**, [S. l.], v. 26, n. 2, p. 569–595, 2021.

MARTINS, Dalton Lopes; LEMOS, Daniela Lucas da Silva; DE OLIVEIRA, Luis Felipe Rosa; SIQUEIRA, Joyce; DO CARMO, Danielle; MEDEIROS, Vinicius Nunes. Information organization and representation in digital cultural heritage in Brazil: Systematic mapping of information infrastructure in digital collections for data science applications. **Journal of the Association for Information Science and Technology**, [S. l.], p. asi.24650, 2022.

PEREIRA, Joanicy Leandra; LEMOS, Daniela Lucas da Silva. Reuso da informação em plataformas de agregação: uma análise utilizando a Revisão Sistemática de Literatura no âmbito da cultura digital. **Ciência da Informação em Revista**, [S. l.], v. 11, p. e15829, 2024.

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

SAYÃO, Luis Fernando; SALES, Luana Farias. Plataformas de gestão de dados de pesquisa: expandindo o conceito de repositórios de dados. Plataformas de gestão de dados de pesquisa: expandindo o conceito de repositórios de dados. **Palavra clave**, v.12 n.1, oct. 2022.

SHINTAKU , Milton; MACÊDO, Diego José; MARIN, Luciano Heitor Gallegos (org.). **Tecnologias utilizadas em pesquisas acadêmicas em Ciências Sociais Aplicadas**. Brasília, DF: Ibict, 2023. 332p.

SOAVE, Maycon; LEMOS, Daniela Lucas da Silva. Curadoria digital em acervos do patrimônio cultural digital: aspectos teóricos e práticos no âmbito da Ciência da Informação. **Brazilian Journal of Information Science: research trends**, v. 16, 2022.

ZENG, Marcia Lei. Interoperability. **Knowledge Organization**, v.46, n.2, p. 122-146, jan. 2019.