



XXI ENANCIB

Encontro Nacional de Pesquisa em Ciência da Informação

50 anos de Ciência da Informação no Brasil:
diversidade, saberes e transformação social

Rio de Janeiro • 25 a 29 de outubro de 2021

XXI Encontro Nacional de Pesquisa em Ciência da Informação – XXI ENANCIB

GT-8 – Informação e Tecnologia

AGREGAÇÃO DE METADADOS DE REPOSITÓRIOS DIGITAIS CULTURAIS: O CASO DAS ENTIDADES VINCULADAS À SECRETARIA ESPECIAL DE CULTURA

METADATA AGGREGATION OF DIGITAL CULTURAL REPOSITORIES: THE CASE OF ENTITIES ATTACHED TO THE SPECIAL SECRETARY OF CULTURE

Luis Felipe Rosa de Oliveira – Universidade de Brasília (UnB)

Dalton Lopes Martins – Universidade de Brasília (UnB)

Modalidade: Trabalho Completo

Resumo: O contexto brasileiro dos repositórios digitais culturais é uma importante fonte de informação para o desenvolvimento de pesquisas na área de ciências humanas, bem como é um insumo indispensável ao consumo cultural da sociedade atual. Porém o contexto nacional da disponibilização de acervos na web é granular, com vários pontos de acessos diferentes, sem um local único e agregado para acesso e uso dos objetos culturais. É sob essa problemática que este estudo é proposto, em busca de entender como pensar uma estratégia de agregação dos metadados de acervos culturais das entidades vinculadas à Secretaria Especial de Cultura, com o objetivo agregar conceitualmente os repositórios digitais heterogêneos das instituições brasileiras de memória através de um único modelo conceitual de metadados. Como metodologia foi aplicado um estudo de caso de quatro etapas que envolveu a seleção, coleta e análise dos metadados de repositórios dessas entidades, bem como o mapeamento para o modelo conceitual Dublin Core, que foi escolhido devido aos resultados da análise dos metadados, que indicou que metadados mais simples são mais comuns entre os repositórios das entidades. Além disso o mapeamento para o Dublin Core conseguiu alcançar 41% dos repositórios analisados, e mesmo deixando de incluir mais da metade do contexto observado, se demonstra uma perspectiva válida para pensar estratégias de interoperabilidade entre esses acervos digitais culturais no âmbito nacional.

Palavras-Chave: Mapeamento de metadados. Repositórios culturais digitais. Interoperabilidade de acervos digitais. Dublin Core.

Abstract: *The Brazilian context of cultural digital repositories is an important source of information for the development of research in the area of human sciences, as well as an indispensable input for the cultural access and use of today's society. However, the national context of making collections available on the web is granular, with several different access points, without a single and aggregated place for consumption of cultural objects. It is under this problem that this study is proposed, seeking to understand how to think about a strategy for aggregating the metadata of cultural collections of entities linked to the Special Secretary of Culture, with the objective of conceptually aggregating the heterogeneous digital repositories of Brazilian memory institutions through a single conceptual metadata model. As a methodology, a four-step case study was applied that involved the selection, collection and*

analysis of metadata from these entities repositories, as well as the mapping to the Dublin Core conceptual model, which was chosen due to the metadata analysis results, which indicated that simpler metadata are more common among entity repositories. In addition, mapping to Dublin Core managed to reach 41% of the analyzed repositories, and even failing to include more than half of the observed context, it demonstrates a valid perspective for thinking about interoperability strategies between these cultural digital collections at the national level.

Keywords: *Metadata Mapping. Digital cultural repositories. Interoperability of digital collections. Dublin Core.*

1 INTRODUÇÃO

Os acervos culturais digitalizados são um insumo indispensável à pesquisa e acesso e uso social. Para pesquisadores das Ciências Humanas, os acervos digitalizados disponíveis através de repositórios digitais são fontes para o desenvolvimento dos estudos da área, já para a sociedade, a possibilidade de explorar objetos culturais através do meio digital potencializa as condições de acesso à cultura no Brasil.

No contexto da Ciência da Informação, por exemplo, estudos relacionados aos processos de organização da informação em repositórios digitais culturais podem ser evidenciados (OLIVEIRA; MARTINS, 2019; MELO; PINTO; TRÉZZE, 2017; MARTINS, 2018). São os repositórios digitais de instituições, como por exemplo a Biblioteca Nacional (BN), a Fundação Nacional de Artes (FUNARTE), ou a Fundação Casa de Rui Barbosa (FCRB) que se tornam fonte de informação para a produção de pesquisa.

A publicação na web dos dados de acervos digitalizados através de repositórios digitais ainda tem grande potencial de apropriação através da aplicação de métodos digitais de análise, comuns no âmbito da área de pesquisa denominada Humanidades Digitais, onde os pesquisadores utilizam de técnicas para a construção de indicadores, análise de métricas, geração automática de metadados, ou até o uso de aprendizagem de máquina para reconhecimento de imagens (XUE; LI; REN, 2020; JANKOVIĆ, 2020).

Ao observar perspectiva em âmbito nacional, como fonte de informação cultural existem 7 entidades culturais vinculadas à Secretaria Especial de Cultura (SEC)¹, estas entidades têm acervos digitalizados com o potencial de promover a apropriação dos novos estudos que utilizam dos métodos digitais para pesquisa no contexto do patrimônio cultural, são elas : a FUNARTE, a BN, a FCRB, o Instituto do Patrimônio Histórico e Artístico Nacional

¹ Entidades vinculadas à Secretaria Especial de Cultura - <http://cultura.gov.br/secretaria/entidades-vinculadas/>

(IPHAN), o Instituto Brasileiro de Museus (IBRAM), a Fundação Cultural Palmares (FCP), e a Agência Nacional do Cinema (ANCINE). Martins e Siqueira (2021) realizaram um primeiro estudo mapeando os repositórios digitais dessas instituições culturais e já apontam a ausência de repositórios na maioria delas e acervos ainda de pequeno volume nas instituições que possuem repositório, caso da FCRB, IPHAN e IBRAM.

Uma das características importantes sobre os acervos digitais destas entidades é que eles estão disponíveis de forma separada, nos sites das próprias instituições, e mesmo no próprio site muitas vezes o acervo está dividido em diferentes repositórios e páginas web pouco estruturadas. Essa condição desassociada de acesso aos objetos culturais brasileiros limita o acesso e o reuso para pesquisas que poderiam aumentar o potencial de difusão dos acervos. a

É preciso que as políticas informacionais dessas instituições levem em conta a usabilidade e o acesso às informações dispersas em seus sites, como por exemplo fomentando a implementação de repositórios digitais, com campo de busca avançada, facetadas, e capacidade de interoperabilidade, facilitando o processo de busca e recuperação da informação pelo usuário e por outros sistemas.

Martins, Silva e Carmo (2018) apontam que mundialmente existem alguns exemplos de integração de repositórios de acervos culturais. No contexto europeu por exemplo, a Europeana² é uma importante iniciativa de interoperabilidade de repositórios digitais de instituições do patrimônio cultural em uma única plataforma, através de um único ponto de acesso. Essas iniciativas geralmente agregam os dados dos objetos digitais culturais, os estruturam de forma semântica através de um esquema padronizado de metadados, além de seguir os princípios dos dados ligados³.

Essa problemática da fragilidade no acesso agregado aos acervos digitalizados da cultura brasileira é um dos fatores que motivaram a proposta do projeto apoiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) intitulado “Interoperabilidade entre os repositórios digitais do patrimônio cultural brasileiro: da web semântica e dados abertos ligados às ferramentas de busca e recuperação da informação”⁴.

² Europeana - <https://www.europeana.eu/pt>

³ Princípios dos dados ligados - <https://www.w3.org/DesignIssues/LinkedData.html>

⁴ Projeto FAPESP “Interoperabilidade entre os repositórios digitais do patrimônio cultural brasileiro: da Web Semântica e dados abertos ligados às ferramentas de busca e recuperação da informação” -

Com vigência entre dezembro de 2020 e novembro de 2021 e sediado na Faculdade de Ciência da Informação da Universidade de Brasília, o projeto tem como objetivo principal “estudar as diferentes formas de interoperabilidade entre os repositórios digitais das entidades vinculadas à Secretaria Especial de Cultura”.

E é sob o âmbito deste projeto que a presente pesquisa é desenvolvida, mais especificamente com referência ao terceiro resultado esperado do projeto, que é o “mapeamento semântico”, cujo objetivo é analisar os metadados utilizados pelas entidades vinculadas à SEC em seus repositórios digitais e mapeá-los para um modelo conceitual escolhido com base na análise dos próprios metadados.

Dessa forma esta pesquisa busca responder à questão geral de como pensar uma estratégia de agregação dos metadados de acervos culturais das entidades vinculadas à SEC, e tem como objetivo agregar conceitualmente os repositórios digitais heterogêneos das instituições brasileiras de memória através de um único modelo conceitual de metadados.

Como objetivos específicos espera-se, diagnosticar e analisar os metadados dos acervos digitais das instituições brasileiras de memória; propor um padrão de metadados com base nos resultados do diagnóstico e análise dos acervos digitais; e mapear os metadados dos acervos digitais para o padrão único de metadados escolhido.

A hipótese inicial é de que a análise dos metadados das instituições deve apontar uma falta de padronização no uso dos metadados para descrição dos objetos digitais dos acervos da maioria das instituições, e isso acarretará em uma alta complexidade de aplicação do modelo conceitual escolhido, que deverá apresentar como característica a representação mínima da maioria dos acervos, e permitir uma possível agregação dos metadados entre as entidades.

Essa hipótese é direcionada pelos resultados das primeiras etapas do projeto FAPESP já realizadas e documentadas no primeiro relatório, que descreve o “mapeamento sistemático das formas de organização da informação dos repositórios digitais de cultura”⁵. Em suma, os dados coletados e analisados desse mapeamento apontam para uma

<https://bv.fapesp.br/pt/auxilios/105582/interoperabilidade-entre-os-repositorios-digitais-do-patrimonio-cultural-brasileiro-da-web-semantica/>

⁵ Relatório do projeto FAPESP “Mapeamento sistemático das formas de organização da informação dos repositórios digitais de cultura” -

<https://pesquisa.tainacan.org/relatorios/produto-01-mapeamento-sistematico-das-formas-de-organizacao-da-informacao-dos-repositorios-digitais-de-cultura/>

dificuldade de identificar a documentação e os padrões utilizados nos repositórios digitais de maneira explícita.

Assim, o atual estudo se justifica sob o contexto de desenvolvimento da pesquisa científica proposta no projeto FAPESP mencionado nos parágrafos anteriores, com a premissa de complementar os resultados do projeto através da exploração de uma forma de agregação de metadados de repositórios digitais das entidades vinculadas à SEC. Para isso, os próximos tópicos apresentarão a metodologia utilizada e os resultados obtidos no processo de mapeamento.

2 METODOLOGIA

Como fruto do projeto de pesquisa atualmente em execução, o processo metodológico aqui proposto é análogo à metodologia aplicada no projeto e prevê um estudo de caso envolvendo as entidades vinculadas à Secretaria Especial de Cultura, que sob uma perspectiva exploratória, propõe mapear para um único esquema os metadados encontrados em repositórios nos sites das entidades.

O estudo apresentado neste artigo foi elaborado posteriormente à etapa de diagnóstico dos sites das entidades vinculadas à SEC executada, que previu analisar os possíveis repositórios digitais existentes nos sites das entidades, coletando seus links e informações sobre sua documentação, o que permitiu que parte desses links de possíveis repositórios coletados correspondam às fontes de informação utilizada nesta pesquisa.

A primeira etapa do estudo foi selecionar, a partir de características de representatividade, como quantidade de itens aproximada e sistemas de publicação utilizados, uma parcela dos links coletados de todas as entidades vinculadas à SEC. Como apresenta o quadro 1, foram selecionados 2 links de repositórios da Biblioteca Nacional, 3 links da FUNARTE, 3 links da Fundação Casa de Rui Barbosa, 2 links da Fundação Cultural Palmares, 2 links do IBRAM e 7 links do IPHAN (sendo 4 deles referentes às unidades de um repositório sob mesmo sistema, o Sistema Integrado de Conhecimento e Gestão/SICG). Não foram encontrados links de potenciais acervos no site da ANCINE e por isso referências à essa entidade não são encontradas neste estudo.

Quadro 1 - Links de repositórios selecionados para análise

ENTIDADE CULTURAL	LINK DO REPOSITÓRIO COLETADO
BIBLIOTECA NACIONAL	http://bndigital.bn.gov.br/acervodigital/
	http://acervo.bn.gov.br/sophia_web/
FUNARTE	http://sbrittod.funarte.gov.br/sophia_acervo/
	http://cedoc.funarte.gov.br/sophia_web/
	http://www.funarte.gov.br/colecoes-cedoc/
FUNDAÇÃO CASA DE RUI BARBOSA	http://rubi.casarui Barbosa.gov.br/
	http://iconografia.casarui Barbosa.gov.br/fotoweb/default.fwx
	http://acervos.casarui Barbosa.gov.br/index.html
FUNDAÇÃO CULTURAL PALMARES	https://www.flickr.com/photos/culturaneagra/
	http://www.palmars.gov.br/?page_id=50190
IBRAM	http://museudainconfidencia.acervos.museus.gov.br/
	http://museudearqueologiadeitaipu.museus.gov.br/
	http://acervodigital.iphan.gov.br/xmlui/
IPHAN	https://pergamum.iphan.gov.br/biblioteca/index.php
	http://portal.iphan.gov.br/videos
	https://sicg.iphan.gov.br/sicg/bemImaterial/acao/
	https://sicg.iphan.gov.br/sicg/bemImaterial/instituicao/
	https://sicg.iphan.gov.br/sicg/bem/visualizar/
	https://sicg.iphan.gov.br/sicg/bemImaterial/rel/

Fonte: Elaborado pelo autor.

A segunda etapa do estudo se deu a partir da coleta dos dados dos repositórios disponíveis pelos links selecionados utilizando scripts em *Python*, com o objetivo de coletar os metadados utilizados, e possibilitar uma posterior análise dos mesmos. A coleta desses links se deu de forma completa através daqueles que disponibilizaram um protocolo de coleta pública e automática como API (*Application Programming Interface*) ou OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*), como é o caso do Flickr da Fundação Cultural Palmares, dos dois acervos museológicos do IBRAM, e do acervo “Coleções CEDOC” da FUNARTE. Para os demais links não foram encontrados, até a data de desenvolvimento deste artigo, meios de coleta automática e por isso foi realizada a coleta de até 100 itens de forma aleatória através de raspagem de dados⁶. Os scripts utilizados nesta coleta estão disponíveis no GitHub⁷.

A terceira etapa teve como objetivo encontrar nos metadados coletados indícios que apontem para o uso de um único esquema de metadados que agregue os repositórios, de forma que esta etapa ocorreu através de uma análise quantitativa dos metadados coletados.

⁶ “Raspagem de dados é uma técnica para extrair dados de sites da web e salvá-los em um sistema de arquivos ou banco de dados para posterior recuperação ou análise.” (ZHAO, 2017, p. 1)

⁷ Scripts de coleta utilizados no estudo -

https://github.com/tainacan/data_science/tree/master/FAPESP/coleta_amostrai

Foi necessário reunir todos os metadados de cada repositório coletado em uma só planilha, e assim calcular a frequência da ocorrência de cada metadado, bem como sua representatividade (quantas entidades utilizaram determinado metadado). O processo realizado utilizou o *Microsoft Excel* e o *OpenRefine* para remover possíveis ambiguidades terminológicas.

A quarta etapa teve o objetivo de propor um mapeamento dos metadados para o esquema indicado na análise da etapa anterior, este mapeamento foi orientado pela pesquisa de Chan e Zeng (2006), que descreve tipos de mapeamento de metadados para interoperabilidade. O modelo que orientou o método aplicado neste estudo foi o modelo de alternância, que mapeia os metadados de diferentes esquemas para um único esquema. O esquema de metadados único escolhido foi o Dublin Core (DC), por ser o esquema mais simples para representação de objetos digitais na web, a amplamente utilizado em sistemas de repositórios digitais como DSpace⁸, Archivematica⁹ e Tainacan¹⁰ por exemplo.

O Dublin Core constitui um “conjunto de elementos de metadados planejado para facilitar a descrição de recursos eletrônicos.” (SOUZA; VENDRUSCULO; MELO, 2000, p. 1), este conjunto inicialmente agrega 15 elementos principais, com a possibilidade do uso de qualificadores. O DC também prevê simplicidade e facilidade na interpretação de seus metadados, além de já apresentar a implementação de recursos no contexto da web semântica.

Vale ressaltar que para a atual pesquisa não foi empregado o uso de qualificadores, que seriam uma alternativa para aumentar o escopo de metadados mapeados. Porém o objetivo do projeto não previu o nível de aprofundamento no conceito dos metadados que exige a implementação dos qualificadores, e também não foram obtidas informações suficientes sobre a documentação dos acervos encontrados para possibilitar esse aprofundamento.

Para realizar o mapeamento dos metadados obtidos para o esquema Dublin Core, foi criada uma planilha online com os metadados obtidos de todos os repositórios em uma coluna, e uma segunda coluna de validação com os 15 metadados do modelo simplificado do DC, e para cada metadado dos repositórios foi definido um dos 15 metadados possíveis, respeitando é claro as diretrizes de descrição de objetos digitais sob a qual o Dublin Core é

⁸ DSpace - <https://duraspace.org/dspace/>

⁹ Archivematica - <https://www.archivematica.org/pt-br/>

¹⁰ Tainacan - <https://tainacan.org/>

construído. Mais informações sobre a escolha e o esquema de metadados do Dublin Core serão apresentadas no tópico de resultados.

Vale ressaltar que o mapeamento de metadados realizado nesta pesquisa se limitou à sintaxe dos termos, não sendo considerados os eventuais vocabulários controlados utilizados nos repositórios ou os valores dos metadados. A limitação à terminologia dos metadados se deu pelo diagnóstico prévio feito, que indicou uma grande dificuldade de encontrar documentação sobre os repositórios, bem como padrões de metadados explícitos, o que dificulta o processo de mapeamento. Também vale ressaltar que o estudo não teve como objetivo efetuar um mapeamento extensivo dos metadados dos repositórios, e sim promover insumos para se pensar uma possível estratégia experimental de agregação entre os repositórios das entidades vinculadas à SEC.

3 RESULTADOS

Os resultados serão apresentados em dois contextos complementares, primeiramente serão exibidos os resultados da análise exploratória quantitativa dos metadados realizada em com o objetivo de indicar um esquema de metadados com potencial de agregação dos diferentes repositórios. Já para responder ao objetivo de mapear os metadados dos repositórios coletados, serão expostos os resultados do mapeamento realizado para o esquema Dublin Core, em busca de apresentar as características de uma possível agregação destes repositórios.

Como mencionado no tópico de metodologia, foram coletados dados de repositórios nos sites de 6 entidades vinculadas à SEC, sendo elas IBRAM, IPHAN, FUNARTE, FCP, FCRB e FCP. Como resultado da coleta foram obtidos os metadados utilizadas em cada conjunto de repositórios selecionados para cada entidade, e como apresenta a tabela 1, as instituições com a maior quantidade de metadados diferentes são a FUNARTE e a BN, com mais de 90 metadados diferentes em seus repositórios. O baixo resultado de 8 metadados da FCP se deve à falta de repositórios mais complexos encontrados, já que foram selecionados para esta pesquisa o acervo fotográfico do *Flickr* e objetos exibidos em uma página simples do site da entidade.

Tabela 1 - Nº de metadados diferentes por entidade.

ENTIDADE	Nº DE METADADOS DIFERENTES	%
FUNARTE	106	38,6%
BN	93	33,9%
IPHAN	76	27,7%

IBRAM	68	24,8%
FCRB	48	17,5%
FCP	8	2,9%

Fonte: Elaborado pelo autor.

Vale apontar que os links de repositórios utilizados nesta pesquisa para a FUNARTE apresentam um contexto de repositórios de temáticas heterogêneas, além de um vasto acervo bibliográfico, também apresenta um acervo arquivístico rico contendo dossiês com imagens e registros de áudio e vídeo. Já os links de repositórios coletados para a BN têm um contexto bibliográfico mais evidenciado, porém uma grande quantidade de objetos catalogada (9 milhões de itens segundo a própria entidade). Estas características de ambas entidades podem contribuir para entender a ampla gama de metadados diferentes, porém além da quantidade de metadados é preciso entender quais são, e se expressam algum padrão.

Ao reunir os metadados de todos os repositórios das entidades em um único lugar, sem a remoção dos metadados com terminologia igual foram obtidos 399 metadados, ao remover os metadados terminologicamente iguais este número caiu para 274 metadados diferentes, demonstrando que apenas 31,32% dos metadados são compartilhados entre as entidades para a descrição de objetos em seus repositórios. Além disso, esse número aponta uma quantidade muito grande de metadados distintos, o que é uma barreira para a agregação destes repositórios.

Para tentar entender melhor quais metadados são mais compartilhados entre os repositórios das entidades e assim propor um modelo mais representativo para a agregação, foi realizado o cálculo da frequência de ocorrência dos metadados. A ocorrência dos metadados revelou que aproximadamente 76% dos metadados só ocorrem 1 vez indicando que a maioria dos metadados são heterogêneos, somente 5% dos metadados aparecerem mais de 4 vezes, confirmando que um pequeno conjunto de metadados indica uma representação comum entre os repositórios analisados.

Além da ocorrência de determinado metadado nos repositórios de uma mesma instituição, foi realizada a contagem de quantas entidades contém os metadados obtidos em algum de seus repositórios analisados, em busca de entender a representatividade institucional dos metadados.

A tabela 2 apresenta os metadados mais representativos de acordo com a presença nos repositórios analisados das entidades. É expressiva a presença de metadados mais simples como “Título”, “Autor”, “Assuntos”, “Descrição”, “Edição” e etc. Com exceção dos metadados “Descrição Física”, “Imprensa”, “Material” e “Série”, os demais metadados estão presentes em 7 dos 15 elementos do Dublin Core, por exemplo.

Tabela 2 - Representatividade dos metadados por entidade.

Metadado	Nº de Entidades	%
Título	6	100,0%
Autor	5	83,3%
Assuntos	4	66,7%
Descrição	4	66,7%
Descrição Física	3	50,0%
Edição	3	50,0%
Idioma	3	50,0%
Imprensa	3	50,0%
ISBN	3	50,0%
Material	3	50,0%
Série	3	50,0%
Título analítico fonte	3	50,0%
Título não controlado	3	50,0%

Fonte: Elaborado pelo autor.

Estes resultados levaram à escolha do esquema Dublin Core como proposta de agregação dos repositórios das entidades vinculadas à SEC neste estudo. Por mais que estes elementos estão presentes em outros esquemas de metadados, como o MARC por exemplo, o Dublin Core foi escolhido principalmente pela proximidade da quantidade menor de elementos descritivos, e pelas características mais simplificadas de representação do próprio modelo do DC.

Assim, a partir do potencial do Dublin Core, os 274 metadados diferentes foram mapeados para este conjunto simples de 15 elementos. Como primeiro resultado, aproximadamente 41% dos metadados foram mapeados para o Dublin Core, um percentual expressivo levando em consideração a heterogeneidade dos metadados, porém parte desse resultado se deve pela característica do mapeamento de muitos para um, que acabou unindo metadados diferentes em um só, como por exemplo “Título”, “Título alternativo” e “Título analítico fonte”, foram mapeados para o elemento “Título” do DC.

Ao observar a proporção dos metadados mapeados para o Dublin Core, o resultado apresentado no gráfico 1 indica que todos os 15 elementos encontraram alguma correspondência no conjunto de metadados dos repositórios coletados. O elemento “Título” foi o que mais demonstrou correspondência (10%) e o metadado Editor o que menos apresentou equivalência.

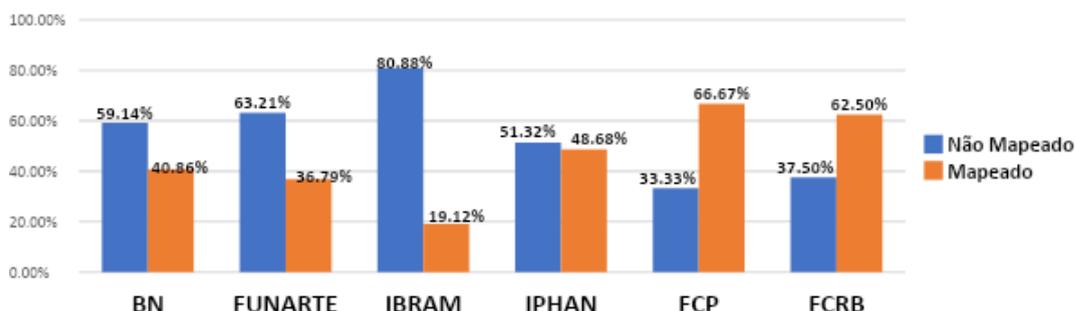
Gráfico 1 - Proporção de metadados do Dublin Core mapeados



Fonte: Elaborado pelo autor.

Por fim, ao observar a proporção de metadados mapeados por entidade (gráfico 2), quatro das seis entidades analisadas apresentam menos de 50% dos metadados mapeados para o Dublin Core, sendo o IBRAM a entidade com menos metadados mapeados (19,12%), e o IPHAN a entidade com quase a metade dos metadados mapeados (48,68%). Já as entidades, Fundação Cultural Palmares e a Fundação Casa de Rui Barbosa são as que mais apresentaram metadados mapeados.

Gráfico 2 - Proporção de metadados mapeados por entidade.



Fonte: Elaborado pelo autor.

Vale incluir nesta análise que os repositórios analisados do IBRAM são referentes à acervos museológicos, que provavelmente seguem o modelo de metadados especificado pelo próprio IBRAM denominado INBCM¹¹, mais direcionado para a gestão de objetos físicos dos museus.

Já quanto às entidades com proporção alta de mapeamento, no caso da FCP é esperado esse resultado pois a maior parte do acervo analisado está no *Flickr*, e os metadados dessa plataforma são voltados à descrição do objeto digital assim como no DC. No caso da FCRB, o acervo RUBI foi coletado via OAI-PMH e já estava todo padronizado sob o modelo do Dublin Core, o que contribui para sua grande proporção de metadados mapeados.

Por fim, os resultados da pesquisa apontam para uma confirmação da etapa de diagnóstico realizada no projeto que motivou este estudo e da hipótese levantada inicialmente, de forma geral não foi encontrada padronização explícita nos metadados obtidos da coleta dos repositórios das entidades vinculadas à SEC, o que por um lado dificulta o processo de mapeamento pois já sugere a barreira da falta de padronização na fonte dos dados, porém, por outro lado mostra a importância de pesquisas em direção à promoção da padronização e potencial agregação destes acervos culturais brasileiros.

4 CONSIDERAÇÕES FINAIS

O esforço de pensar estratégias de implementação de uma plataforma integrada de acesso aos objetos digitais culturais do Brasil é sem dúvidas significativo, são muitos fatores envolvidos, além das questões técnicas e conceituais, existe uma camada política que

¹¹ Modelo INBCM do IBRAM -

<https://www.museus.gov.br/resolucao-regulamenta-inventario-nacional-de-bens-culturais-musealizados/>

envolve diretamente os gestores das entidades culturais brasileiras e a própria gestão no nível dos órgãos federais responsáveis.

Nesse âmbito das políticas, podem ser citadas as pesquisas desenvolvidas por Martins e Dias (2019, 2020), que relatam a necessidade de se pensar políticas para acervos em rede de instituições de memória, apontando a falta de uma política pública nacional para a implementação dos acervos digitais dessas instituições. Sem essa orientação dos órgãos gestores, o processo de desenvolvimento e publicação dos acervos digitais fica restrito às diretrizes e recursos de cada instituição, levando a barreiras como a falta de um padrão mínimo comum entre os acervos de mesmo contexto por exemplo, que seria um passo importante na interoperabilidade dos repositórios entre as entidades culturais.

Mesmo assim, propor estudos no contexto técnico e conceitual da integração de acervos culturais é dar insumos para que as políticas tanto institucionais quanto nacionais sejam elaboradas. Pesquisas como a proposta neste estudo, auxiliam a manter evidente e atual a discussão sobre quais os desafios e oportunidades existem na interoperabilidade entre acervos digitais culturais, levando em consideração inclusive questões como a apropriação dos dados abertos ligados e da web semântica.

Pesquisadores brasileiros da ciência da informação têm colaborado para difundir estudos sob a temática de interoperabilidade de acervos culturais digitais, como por exemplo Marcondes (2012, 2016 A, 2016 B), Santarém Segundo, Silva e Martins (2019), Sayão (2007, 2008), além de Sacramento, Baião e Ferreira (2015). São estudos no âmbito das bibliotecas, museus, e arquivos, que vão de encontro a um objetivo geral comum, buscar explorar e entender melhor as possibilidades de estratégias de agregação de acervos digitais.

De forma complementar, o estudo aqui proposto, que é proveniente de um projeto em andamento com enfoque nos acervos disponíveis no site das entidades vinculadas à SEC, também propõe discutir estratégias de interoperabilidade dos acervos digitais dessas instituições, a proposta de análise dos metadados e mapeamento dos mesmo para um único modelo conceitual é um esforço de entender de maneira prática como se daria uma possível iniciativa de agregação desses acervos.

Mesmo sem o aprofundamento semântico, que é indicado no processo de mapeamento conceitual, os resultados apontam uma realidade de dificuldades na agregação dos metadados em um único modelo. A falta de documentação explícita no site das entidades analisadas é um agravante que limita à interpretação subjetiva dos metadados no

repositório digital publicado, outra barreira importante é a granularidade de repositórios digitais no site das entidades, o IPHAN por exemplo tem muito de seu acervo espalhado em páginas simples pelo site, e o acervo patrimonial e imaterial publicado pelo sistema SICG está dividido em outras quatro unidades que utilizam conjuntos de metadados diferentes.

Iniciativas como o repositório RUBI da FCRB¹² é um bom exemplo de como a integração dos acervos da instituição em uma só plataforma pode ser benéfica. Os acervos histórico, museológico e bibliográfico da entidade estão publicados sob um único modelo de metadados, o Dublin Core, e estão disponíveis para a coleta automática através do protocolo OAI-PMH. Em uma possível integração entre as entidades culturais a forma como esse repositório está publicado na web com certeza facilitaria o processo de interoperabilidade.

Outros acervos publicados através do Tainacan, como algumas coleções da FUNARTE e os acervos museológicos vinculados ao IBRAM também permitem a coleta automática dos dados seja através de API ou do protocolo OAI-PMH, mas não demonstram uma padronização dos metadados tão efetiva.

E no caso de acervos publicados através do sistema SophiA, principalmente bibliográficos, como o caso da BN e FUNARTE, apesar de apresentarem metadados relativamente padronizados sob o modelo MARC, não apontam uma forma explícita de coleta automática de metadados, o que dificulta a possibilidade de integração desses tipos de acervo com os demais.

Além dessas circunstâncias de infraestrutura tecnológicas enfrentadas na primeira etapa metodológica deste estudo, a perspectiva conceitual aponta que de forma geral que um conjunto de metadados mais simples são comuns entre os repositórios das entidades, o que leva a sugestão de um esquema descritivo mais restrito para a proposta de agregação desses metadados em um único modelo. Por isso o Dublin Core se demonstrou uma escolha interessante, além de sua camada representativa simples, possui respaldo conceitual reconhecido e utilizado internacionalmente e inclui iniciativas atuais como a representação semântica dos metadados.

Mesmo assim os resultados apontam que para quatro das seis entidades analisadas, menos da metade de seus metadados foram mapeados para o modelo do DC, ou seja, um mapeamento objetivo e sintático como foi realizado neste estudo reduziria para menos da metade os elementos descritivos dos objetos culturais digitais dos acervos dessas entidades.

¹² Sobre o repositório RUBI - <http://rubi.casaruibarbosa.gov.br/sobre.html>

Porém apesar desse resultado, é importante a constatação de que é possível realizar um processo de mapeamento conceitual entre os acervos das entidades culturais utilizando um modelo simples e efetivo como o DC. Entende-se que vale o incentivo a estudos mais aprofundados que apontem se existem também equivalência semântica entre os acervos, como uso de vocabulários controlados comuns, ou normas de catalogação comuns implícitas.

Dessa forma, espera-se que o esclarecimento dessa realidade conceitual dos metadados sirva de insumo para uma reflexão em direção às iniciativas que facilitem estratégias de interoperabilidade no âmbito dos acervos dessas instituições de memória, como por exemplo o desenvolvimento de interoperabilidade entre os repositórios de uma mesma instituição, e o uso de padrões de metadados e normas de catalogação que conservem um mínimo comum entre os repositórios.

REFERÊNCIAS

CHAN, Lois Mai; ZENG, Marcia Lei. Metadata interoperability and standardization—a study of methodology part I. **D-Lib magazine**, v. 12, n. 6, 2006.

DIAS, Calíope Victor Spíndola de Miranda; MARTINS, Dalton Lopes. Iniciativas brasileiras em torno da construção de uma Política Nacional para Acervos Digitais de Instituições de Memória: o desafio da memória em tempos de cultura digital. **Políticas Culturais em Revista**, v. 13, n. 1, p. 16-46, 2020.

DIAS, Calíope Victor Spíndola de Miranda; MARTINS, Dalton Lopes. Acervos Digitais: Perspectivas, Desafios e Oportunidades para as Instituições de Memória no Brasil. **Panorama Setorial da Internet**, n. 3, p. 1-13, 2019.

JANKOVIĆ, Radmila. Machine learning models for cultural heritage image classification: Comparison based on attribute selection. **Information**, v. 11, n. 1, p. 12, 2020.

MARCONDES, Carlos Henrique. “Linked data”—dados interligados-e interoperabilidade entre arquivos, bibliotecas e museus na web. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 17, n. 34, p. 171-192, 2012.

MARCONDES, Carlos Henrique. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. **Perspectivas em Ciência da Informação**, v. 21, n. 2, p. 61-83, 2016 A.

MARCONDES, Carlos Henrique. O papel dos modelos conceituais para interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus. **Ene**, v. 9, p. 16, 2016 B.

MARTINS, Dalton Lopes, SILVA, Marcel Ferrante, CARMO, Danielle do. Acervos em rede: perspectivas para as instituições culturais em tempos de cultura digital. **Em Questão**, Porto Alegre, v. 24, n. 1, p. 194-216, jan./abr. 2018.

MARTINS, Dalton Lopes, SIQUEIRA, Joyce. Análise e mapeamento da organização técnica dos acervos digitais das instituições federais vinculadas ao Ministério da Cultura: fontes de informação para o desenvolvimento das Humanidades Digitais. In: PIMENTA, Ricardo M., ALVES, Daniel. **Humanidades digitais e o mundo lusófono**. Rio de Janeiro: Editora FGV, 2021. 375 p.

MARTINS, Vinicius Pontes. BNDigital–10 anos: retrospectiva e perspectivas para os próximos 10 anos. **Seminário Serviços de Informação em Museus**, p. 159-169, 2018.

MELO, Elisete de Sousa; PINTO, Tiago Leite; TRÉZZE, Luziana Jordão Lessa. Peculiaridades na catalogação da coleção de cordel no Repositório Rui Barbosa de Informações Culturais (RUBI). **Revista Brasileira de Biblioteconomia e Documentação**, v. 13, p. 1183-1196, 2017.

OLIVEIRA, Luis Felipe Rosa de; MARTINS, Dalton Lopes. Fundação Nacional de Artes: estudo de caso da migração e publicação dos dados do acervo digital com o software livre Tainacan. **XX ENANCIB**, 2019.

SACRAMENTO, Carolina; BAIÃO, Fernanda de Araújo; FERREIRA, Simone Bacellar Leal. Um Estudo sobre Metadados para Preservação e Integração de Acervos Digitais Heterogêneos. **RelaTe-DIA**, v. 8, n. 1, 2015.

SAYÃO, Luís Fernando. Interoperabilidade das bibliotecas digitais: o papel dos sistemas de identificadores persistentes-URN, PURL, DOI, Handle System, CrossRef e OpenURL. **Transinformação**, v. 19, n. 1, p. 65-82, 2007.

SAYÃO, Luís Fernando; MARCONDES, Carlos Henrique. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **Transinformação**, v. 20, n. 2, p. 133-148, 2008.

SEGUNDO, Jose Eduardo Santarem; SILVA, Marcel Ferrante; MARTINS, Dalton Lopes. Revisitando a interoperabilidade no contexto dos acervos digitais. **Informação & Sociedade**, v. 29, n. 2, 2019.

SOUZA, Marcia Izabel Fugisawa; VENDRUSCULO, Laurimar Gonçalves; MELO, Geane Cristina. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ciência da Informação**, v. 29, n. 1, p. 93-102, 2000.

XUE, Song; LI, Yanlin; REN, Lei. Representing the Chinese Seal Stamping Catalogs Using IIIF & Serverless. In: **Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020**, p. 547-548, 2020.

ZHAO, Bo. Web scraping. **Encyclopedia of big data**, p. 1-3, 2017.