

Tratamento da informação em acervos culturais: avaliação do uso de vocabulários controlados em coleções museológicas sob gestão do Instituto Brasileiro de Museus

Abeil Coelho Júnior

Universidade Federal do Espírito Santo, Programa de Pós-graduação em Ciência da Informação
Vitória, ES, Brasil

abeilc@hotmail.com

Daniela Lucas da Silva Lemos

Universidade Federal do Espírito Santo, Departamento de Biblioteconomia e Programa de Pós-graduação em Ciência da Informação Vitória, ES, Brasil

daniela.l.silva@ufes.br

DOI: <https://doi.org/10.26512/rici.v16.n1.2023.47531>

Recebido/Recibido/Received: 2022-12-20

Aceitado/Aceptado/Accepted: 2023-03-11

ARTIGOS

Resumo

Nos últimos anos, tem-se observado considerável adesão das instituições de patrimônio cultural ao processo de digitalização e disponibilização dos dados de seus acervos na internet. No entanto, a qualidade desses dados deve ser mensurada. Para isso, este trabalho busca avaliar a adequação das coleções museológicas disponibilizadas online pelo Instituto Brasileiro de Museus quanto ao uso de um vocabulário controlado, de acordo com as orientações do guia de catalogação de objetos culturais de referência, o *Cataloging of Cultural Objects*. Foi realizado o alinhamento dos metadados utilizados pelo Instituto Brasileiro de Museus e o guia de catalogação; em sequência, foi desenvolvido um script em Python para processamento dos dados das coleções e calculada a adequação à recomendação ao uso de vocabulário controlado. Os resultados apontam que apenas 5 dos 9 elementos recomendados utilizam vocabulário controlado. Além disso, apesar de essas coleções estarem sob gestão do Instituto Brasileiro de Museus, as práticas de catalogação não estão alinhadas entre as instituições. Recomenda-se, portanto, que sejam incorporadas práticas de catalogação baseadas em instrumentos de referência na modelagem de metadados das coleções museológicas sob gestão do Instituto Brasileiro de Museus, visando qualificar os padrões de documentação em uso atualmente.

Palavras-chave: Museus. Organização da informação. Representação da informação. Qualidade de dados. Vocabulário controlado.

Treatment of information in cultural collection: evaluation of the use of controlled vocabularies in museum collections managed by the Brazilian Institute of Museums

Abstract

In recent years, there has been a considerable trend of cultural heritage institutions joining the process of digitizing and making their collection data available on the internet. However, the quality of this data must be measured. To this end, this study seeks to evaluate the adequacy of the online museum collections made available by the Brazilian Institute of Museums regarding the use of a controlled vocabulary, according to the guidelines of the Cataloging of Cultural Objects reference guide. The metadata used by the Brazilian Institute of Museums was aligned with the cataloging guide; subsequently, a Python script was developed to process the collection data and calculate the adequacy to the recommendation of using a controlled vocabulary. The results indicate that only 5 of the recommended

9 elements use a controlled vocabulary. Additionally, despite these collections being managed by the Brazilian Institute of Museums, cataloging practices are not aligned between institutions. Therefore, it is recommended that cataloging practices based on reference instruments be incorporated into the metadata modeling of museum collections managed by the Brazilian Institute of Museums, with the aim of improving the documentation standards currently in use.

Keywords: Museums. Information organization. Information representation. Data quality. Controlled vocabulary.

Tratamiento de la información en las colecciones culturales: evaluación del uso de vocabularios controlados en las colecciones de los museos administrados por el Instituto Brasileño de Museos

Resumen

En los últimos años, se ha observado una considerable adhesión de las instituciones de patrimonio cultural al proceso de digitalización y disponibilización de los datos de sus acervos en internet. Sin embargo, la calidad de estos datos debe ser medida. Para ello, este trabajo busca evaluar la adecuación de las colecciones museológicas disponibles en línea por el Instituto Brasileño de Museos en cuanto al uso de un vocabulario controlado, de acuerdo con las orientaciones de la guía de catalogación de objetos culturales de referencia, el *Cataloging of Cultural Objects*. Se realizó el alineamiento de los metadatos utilizados por el Instituto Brasileño de Museos y la guía de catalogación; a continuación, se desarrolló un script en Python para procesamiento de los datos de las colecciones y se calculó la adecuación a la recomendación del uso de vocabulario controlado. Los resultados indican que sólo 5 de los 9 elementos recomendados utilizan vocabulario controlado. Además, aunque estas colecciones están bajo la gestión del Instituto Brasileño de Museos, las prácticas de catalogación no están alineadas entre las instituciones. Se recomienda, por lo tanto, que se incorporen prácticas de catalogación basadas en instrumentos de referencia en la modelación de metadatos de las colecciones museológicas bajo gestión del Instituto Brasileño de Museos, con el objetivo de calificar los estándares de documentación en uso actualmente.

Palabras clave: Museos. Organización de la información. Representación de la Información. Calidad de los datos. Vocabulario controlado.

Introdução

Nos últimos anos, instituições de patrimônio cultural, a exemplo dos museus, têm aderido ao processo de digitalização e disponibilização de dados de seus acervos na internet, como é o caso do Instituto Brasileiro de Museus (Ibram) (IBRAM, 2022). Entretanto, investir no processo de digitalização e disponibilização de objetos culturais na rede não é suficiente (MARTINS *et al.*, 2022), pois não garante a qualidade dos dados bem como dos metadados envolvidos, considerando os diversos tipos de bancos de dados e sistemas de informação inclusos em processos de organização, modelagem e representação da informação. Diante deste cenário, considerações acerca da importância da qualidade de dados na internet surgem em diversos contextos (BIZER; HEATH; BERNERS-LEE, 2009; WILKINSON *et al.*, 2016; MACEDO; LEMOS, 2021; MARTINS *et al.*, 2022; LEMOS; COELHO JUNIOR, 2023).

Para o contexto da presente pesquisa, custodiadores e proprietários de dados, como, por exemplo, galerias, bibliotecas, arquivos e museus – GLAMs, acrônimo em inglês para tais termos, são os principais responsáveis pela qualidade de seus dados, com uma boa catalogação descritiva e de assunto (IFLA, 2016). Com o uso de padrões de documentação que orientam a estrutura de dados, valores de dados e conteúdo de dados (GILLILAND, 2016), as instituições contam com um conjunto de ferramentas que pode levá-las a uma boa prática de catalogação,

documentação consistente, e por consequência, maior acesso aos documentos pelo usuário final. No entanto, aqueles que fornecem os dados e aqueles que usam os dados também têm responsabilidades. Os coletores de dados e catalogadores têm o dever de rotular os dados e documentar metodologias de captação; os custodiadores têm o papel de fazer a manutenção e o controle de qualidade dos seus registros; e os usuários em reportar eventuais erros encontrados (CHAPMAN, 2005).

Com objetivo de viabilizar a interoperabilidade, acesso e reuso de seus dados (HARPRING, 2022), instituições de patrimônio cultural geralmente aderem a padrões de catalogação que produzem descrições de coleções de objetos culturais digitalizados ou natos digitais, os quais necessitam padronizar e normalizar aspectos únicos de coleções culturais tanto fisicamente quanto digitalmente, bem como fornecer dados administrativos para descrever a digitalização, os direitos autorais e as disposições de uso dos objetos.

Uma das formas de padronização e normalização para buscar a qualidade de dados de uma coleção cultural está na adoção de padrões de tratamento da informação atrelados ao uso de vocabulário controlado, que segundo Lancaster (2004) é definido como uma lista de termos autorizados, em que o catalogador ou indexador somente pode atribuir a um documento termos presentes na lista adotada pela unidade de informação envolvida. Geralmente, o uso de vocabulário controlado é recomendado a partir de orientações de guias de catalogação, que auxiliam em como elaborar a descrição de um recurso de informação e seus pontos de acesso (MARTINS *et al.*, 2021).

O Ibram, objeto de estudo desta pesquisa, faz a gestão do patrimônio museológico a nível federal de 23 coleções digitais de 22 museus espalhados pelo país por meio do *software* Tainacan, uma ferramenta de organização e gestão de acervos digitais (MARTINS; LEMOS; ANDRADE, 2021). As bases de dados desses museus foram modeladas a partir do padrão de dados adotado internamente pela instituição, qual seja o modelo do Inventário Nacional de Bens Culturais Musealizados – INBCM (BRASIL, 2021). Contudo, o INBCM não pode ser considerado um guia de catalogação por não almejar, primariamente, requisitos descritivos únicos e singulares, vocabulários padronizados, indexação, localização, acesso e navegação em sistemas de recuperação da informação (SRIs) contemporâneos (IFLA, 2016).

No âmbito do domínio do patrimônio cultural, o *Cataloging of Cultural Objects* (CCO) tem sido recomendado como resultado do consenso de profissionais das comunidades de museus, bibliotecas, galerias e arquivos que pesquisam sobre práticas comuns de catalogação (BACA *et al.*, 2006; HARPRING, 2022). O guia fornece diretrizes para selecionar, ordenar e formatar dados usados para preencher registros de catálogo com base em categorias principais em CDWA e VRA Core. Porém, embora tenha sido inspirado no desenvolvimento dos elementos

VRA Core e do *Getty Vocabularies*, o CCO apresenta conceitos mais genéricos que podem ser utilizados com outros conjuntos de metadados, como, por exemplo, o *Machine-Readable Cataloging* (MARC21), o *Metadata Object Description Schema* (MODS), o Dublin Core, entre outros. Nesse sentido, o CCO aparece como um guia de referência no universo da catalogação, que pode auxiliar as práticas de catalogação atreladas ao uso de vocabulários controlados em instituições do patrimônio cultural como os museus. Assim, a questão de pesquisa está em responder *qual o nível de padronização e normalização dos metadados produzidos pelos museus sob gestão do Ibram por meio do emprego de vocabulários controlados à luz das recomendações do guia de referência CCO?*

Logo, o objetivo do artigo é avaliar a qualidade de dados na perspectiva do uso de vocabulário controlado nos elementos de metadados que descrevem as coleções dos museus sob gestão do Ibram.

O presente artigo está organizado como se segue. Esta introdução contextualiza a pesquisa, indicando a sua questão e o seu objetivo. A Seção 2 introduz um breve referencial teórico sobre linguagens documentárias, vocabulários controlados e seus empregos no âmbito das práticas de organização da informação no campo da Documentação. A Seção 3 elucida os procedimentos metodológicos empregados na condução da pesquisa. A Seção 4 apresenta os resultados a partir da execução de tais procedimentos. A Seção 5 elabora uma discussão a partir dos resultados alcançados. E, finalmente, a Seção 6 traz as considerações finais do artigo.

Vocabulário controlado como instrumento nas práticas de tratamento da informação

Linguagens documentárias compreendem a comunicação entre usuários e SRIs, sendo consideradas linguagens artificialmente construídas a partir da linguagem natural presente nos documentos, buscando-se obter um vocabulário controlado de um assunto específico (DODEBEI, 2002; LANCASTER, 2004). Já vocabulário controlado é definido por Lancaster (2004) como uma lista de termos autorizados, em que o indexador somente pode atribuir a um documento termos que existem na lista adotada pela unidade de informação em que trabalha. O vocabulário controlado tem por funções: i) o controle de sinônimos, através da definição de um termo padrão, com remissivas para os sinônimos; ii) a diferenciação entre os homógrafos¹; e iii) o agrupamento de termos em que os significados apresentem uma relação mais estreita entre si, como por exemplo, relações hierárquicas ou não hierárquicas.

¹ Vocabulário que tem a mesma grafia de outro, porém sentido diferente (ANSI, 2005).

A ANSI (2005) enfatiza que o principal propósito dos vocabulários controlados é fornecer um significado para a organização da informação, incluindo tradução, consistência, relacionamentos, visualização e recuperação perante a informação.

Dentro do universo dos Sistemas de Organização do Conhecimento (SOCs) (HJØRLAND, 2007), consideram-se também as linguagens documentárias como resultantes da análise semântica de conceitos e relacionamentos de um domínio que busca, *a priori*, a organização de recursos de informação nos aspectos de: i) associação, gerando relacionamentos; ii) representação, gerando pontos de acesso e índices em processos de catalogação e indexação; iii) classificação, promovendo colocação e ordenação para os documentos; e iv) categorização, gerando esquemas de categorias. Logo, podem-se citar alguns SOCs alinhados com o conceito de linguagens documentárias ou vocabulários controlados usados em ambiência digital, porém numa perspectiva de caracterização por níveis de complexidade estrutural, isto é, pelo formalismo promovido pela adequabilidade no uso de tecnologias semânticas, especialmente no que se refere aos relacionamentos (HJØRLAND, 2007), a saber: taxonomias, tesouros e ontologias.

Taxonomias, por exemplo, são coleções de termos classificados em uma estrutura hierárquica, na qual se emprega relacionamentos de generalização e especialização. O Ibram utiliza o *software* Tainacan, que possui o recurso de produção de taxonomias para normalização de valores de seus tipos de metadados (MARTINS; LEMOS; ANDRADE, 2021). Já nos tesouros, as relações semânticas se estendem em equivalência, hierárquicas e associativas entre os termos, em que são claramente mostradas e identificadas através de indicadores de relação padrão (ANSI, 2005). Ontologias possuem os mesmos princípios dos vocabulários controlados (SILVA; SOUZA; ALMEIDA, 2008), isto é, trabalham com linguagem natural e fazem a delimitação de termos e de relações. Entretanto, a semântica envolvida na terminologia da ontologia se difere dos vocabulários controlados (dentro do contexto das linguagens documentárias) por incluir axiomas formais (através de declarações lógicas) que restringem a utilização do vocabulário. Outra distinção está no fato de as especificações de relações no contexto das ontologias serem em número superior às dos tesouros, por exemplo.

Os vocabulários controlados vistos como SOCs viabilizam, portanto, a recuperação de informação nos mais diversos ambientes de informação, a exemplo dos museus digitais, padronizando a entrada de dados, facilitando a estratégia de busca e, conseqüentemente, melhorando a interação do usuário com o SRI.

Para tal, Svenonius (2000) acrescenta que para a informação ser organizada no âmbito de um SRI, precisa ser descrita e que, o produto desse processo descritivo é a representação da informação. Ressalta que alguns tipos de representação da informação são construídos através

do uso de linguagens, as quais são subdivididas em linguagens que descrevem a informação (o conteúdo) e linguagens que descrevem o documento (o suporte), no todo ou em partes. A linguagem usada para descrever o documento está relacionada à representação descritiva, também tratada como processo de catalogação, que busca retratar aspectos específicos do documento que permitam a sua individualização e determinação dos pontos de acesso para proporcionar aos usuários a condição de encontrar, identificar, selecionar e obter o item por meio de um catálogo (IFLA, 2016). Já a linguagem destinada à descrição do conteúdo do documento está relacionada à representação temática que vislumbra aspectos intelectuais e semânticos, portanto, subjetivos, como a compreensão do assunto do documento para fins de tradução para uma linguagem documentária (LANCASTER, 2004; ANSI, 2005). Para os propósitos desta pesquisa, ambas as representações são consideradas na avaliação da qualidade dos dados das coleções ora envolvidas, sendo o emprego de vocabulário controlado avaliado tanto na descrição da Obra em seus aspectos de suporte quanto na descrição de seu conteúdo.

Finalmente, o cenário atual dos SRIs em ambiência digital evidencia um panorama de uso e reúso de vocabulários semânticos na organização de uma quantidade significativa de dados que tem sido gerada, conectada e distribuída na web, os quais necessitam de metadados eficientes para gestão, organização e recuperação de conteúdos associados a acervos culturais distribuídos na rede (MARTINS et al., 2022).

Procedimentos Metodológicos

Pela abrangência de museus, quantidade de material e recursos disponíveis na internet, foi escolhido o Ibram como objeto de avaliação. Como informado anteriormente, o Ibram utiliza o INBCM como modelo de dados para descrição dos acervos dos museus sob sua gestão. O modelo possui elementos discricionais para coleções museológicas, bibliográficas e acervos arquivísticos. Porém, para os objetivos deste trabalho foram considerados apenas os elementos de caráter museológico, dado o volume de itens museais pertencente à grande maioria dos acervos sob gestão do Ibram.

Desta forma, o primeiro passo foi realizar o alinhamento (mapeamento) entre os elementos descritivos de caráter museológico do INBCM com os elementos recomendados pelo guia de catalogação CCO, elencado como o modelo de regras de catalogação de referência aos propósitos da presente pesquisa, conforme Quadro 1. O alinhamento se deu a partir de um procedimento manual e intelectual baseado na aquisição de conhecimento sobre os dois instrumentos de pesquisa.

Quadro 1 – Alinhamento entre elementos descritivos INBCM e CCO

Capítulo CCO	Elemento CCO	Obrigatório CCO	Vocabulário Controlado CCO	Elemento INBCM	Obrigatório INBCM
I-Part 2	<i>Work Type</i>	Sim	Sim	Denominação	Sim
I-Part 2	<i>Title</i>	Sim	Não	Título	Não
II-Part 2	<i>Creator</i>	Sim	Sim	Autor	Sim
III-Part 2	<i>Measurements</i>	Sim	Sim	Dimensões	Sim
III-Part 2	<i>Materials and Techniques</i>	Sim	Sim	Material/Técnica	Sim
III-Part 2	<i>Physical Description</i>	Não	Sim	Estado de Conservação	Sim
III-Part 2	<i>Inscription</i>	Sim	Sim	Número de Registro	Sim
IV-Part 2	<i>Date</i>	Sim	Não	Data de Produção	Não
V-Part 2	<i>Creation Location</i>	Não	Sim	Local de Produção	Não
VII-Part 2	<i>Class</i>	Sim	Sim	Classificação	Não
VIII-Part 2	<i>Description</i>	Não	Não	Resumo Descritivo	Sim
VIII-Part 2	<i>Other Descriptive Notes</i>	Não	Não	Condições de Reprodução	Sim
VI-Part 1	<i>Related Works</i>	Não	NA	Mídias Relacionadas	Não
V-Part 2	<i>Location</i>	Sim	Sim	Situação	Sim
NA	NA	NA	NA	Outros Números	Não

Fonte: elaborado pelos autores (2022).

O segundo passo, a partir do alinhamento, foi captar os dados do Ibram a partir de *script*(COELHO, 2022) por meio da utilização da linguagem de programação *Python*, e com o uso das bibliotecas *Pandas*², *BeautifulSoup*³ e *Requests*⁴, para realizar a exportação em massa de todos os dados dos acervos dos museus no formato CSV⁵.

O terceiro passo incumbiu-se de avaliar, para cada elemento descritivo de cada acervo, a indicação de uso de vocabulário controlado (de acordo com a quarta coluna do Quadro 1), para o caso específico do Ibram, o emprego de taxonomia. A avaliação foi feita a partir dos dados disponibilizados pela API do próprio Tainacan, disponível no painel de exportação com nome "API do Tainacan em formato JSON". Essa API disponibiliza dados para além dos elementos de

² <https://pandas.pydata.org>

³ <https://beautiful-soup-4.readthedocs.io/>

⁴ <https://requests.readthedocs.io/en/master/>

⁵ https://pt.wikipedia.org/wiki/Comma-separated_values

metadados do INBCM, dentre eles a indicação da configuração do elemento de metadado, caso este seja do tipo taxonomia em uma determinada coleção.

Por fim, foram captados os dados de todas as 22 coleções de caráter museológicos disponíveis online pelo Ibram, e avaliados se os elementos de metadados nestas coleções possuíam a indicação de taxonomia. Desta forma, para cada coleção e elementos de metadados que possuísem a indicação de taxonomia e que tivessem valor preenchido foi atribuído: (i) o valor 1 (um), por estar de acordo com a regra de uso de vocabulário controlado, conforme regra do CCO; ou (ii) o valor 0 (zero), por não atender a recomendação do CCO, seja pelo elemento de metadado não ter a configuração de taxonomia ou não apresentar valor preenchido. Assim, para todas as coleções e elementos de metadados foi calculada a taxa de adequação a regra de uso de vocabulário controlado do CCO com a seguinte fórmula:

$$\text{índice}_b = (\sum \text{Valor1} / (\sum \text{Valor1} + \sum \text{Valor0})) * 100$$

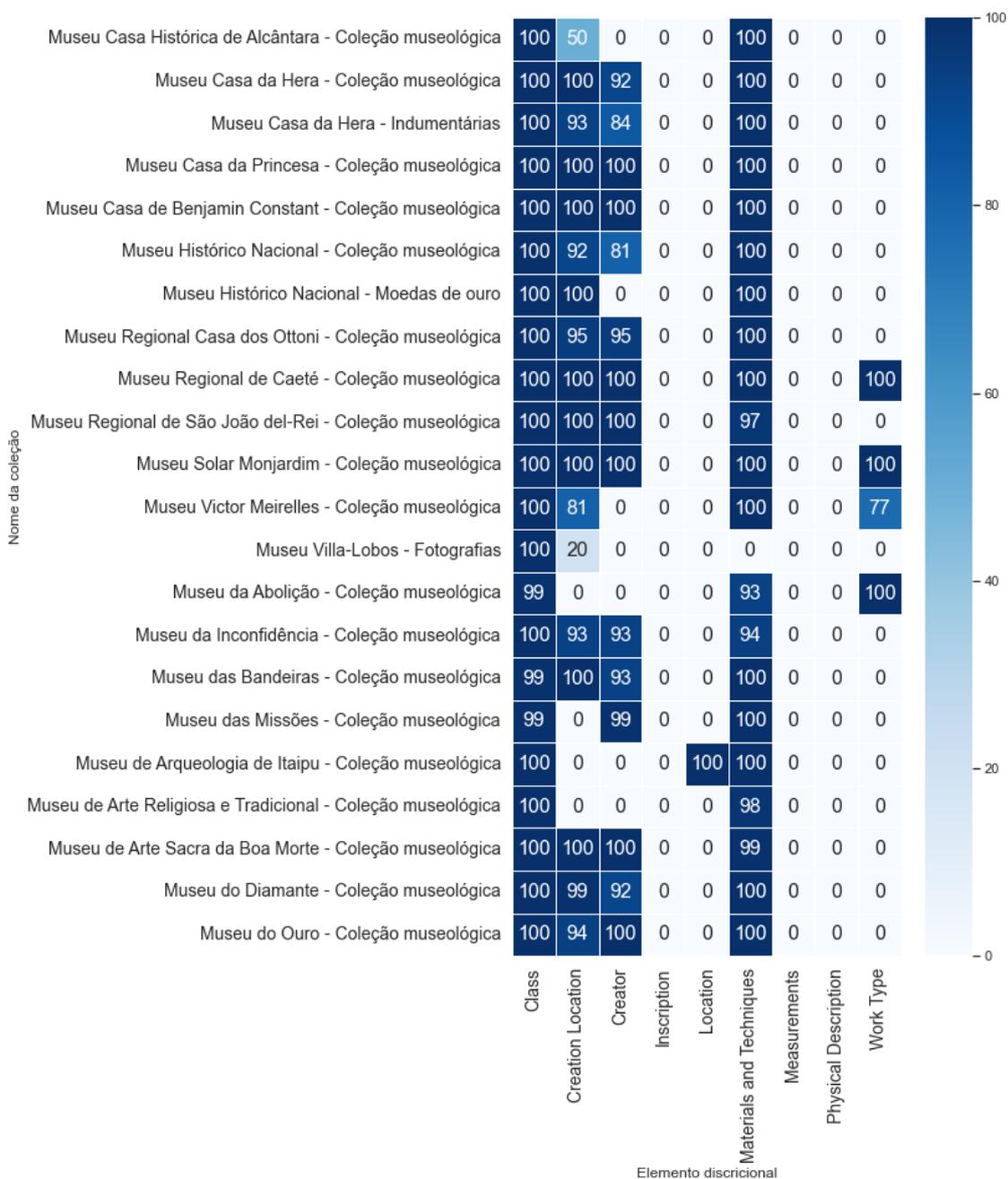
A fórmula prescreve **b** como a base de dados contendo uma coleção em particular; **índice** como o percentual de adequação obtido em relação à dimensão, elemento de metadado e regra de catalogação para um determinado museu e coleção; **Valor1** como a indicação de ocorrência do registro de dado que **se adequou** a regra; **Valor0** como a indicação de ocorrência do registro de dado que **não atendeu** a regra.

Resultados

A avaliação de uso de vocabulário controlado dos 22 acervos museológicos sob gestão do Ibram foi realizada por meio do alinhamento entre elementos de metadados do INBCM à luz das recomendações do guia de catalogação CCO (Quadro 1). Observa-se no Quadro 1, coluna “Vocabulário Controlado CCO”, que para 9 (nove) dos 15 (quinze) elementos são recomendados o uso de vocabulário controlado pelo CCO.

Nesse sentido, a taxa de adequação dos elementos de metadados das bases de dados do Ibram que tiveram a indicação de uso de taxonomia pode ser observada a seguir na Figura 1. Observam-se nas linhas, as coleções museológicas disponíveis online pelo Ibram; e nas colunas, os elementos de metadados alinhados com as regras de catalogação de uso de vocabulário controlado à luz do CCO. Cada célula apresenta a taxa de adequação das coleções à regra de uso de vocabulário controlado, com valores indo de 0 (zero) - onde foi observada a completa inadequação da coleção à regra; a 100 - em que houve a completa adequação da regra pela coleção em questão. Assim, as cores mais claras representam menores taxas de adequação e as mais fortes maiores taxas de adequação.

Figura 1 - Adequação de coleções Ibram ao uso de vocabulário controlado



Fonte: elaborado pelos autores (2022).

Pode-se destacar o maior índice do uso de vocabulário controlado no elemento de metadado *Class*, sendo a taxa de adequação em 99% nas coleções dos museus da Abolição, Bandeiras e Missões. O restante das coleções com taxa de adequação em 100%.

O elemento *Creation Location* apresentou menor taxa de adequação nas coleções dos museus da Abolição, Missões, Arqueologia de Itaipu e de Arte Religiosa e Tradicional com adequação em 0%, devido ao elemento não apresentar o uso de taxonomia nessas coleções. Ainda neste elemento, podemos observar que as coleções dos museus Casa Histórica de

Alcântara, Casa da Hera, Histórico Nacional – Coleção Museológica, Regional Casa dos Ottoni, Vitor Meirelles, Villa-Lobos, do Diamante e do ouro apresentaram os usos de taxonomia, porém com elementos de metadados com valores vazios. Neste caso, sem valor taxonômico informado. Os demais apresentaram o uso de taxonomia e todos os elementos haviam valores informados.

Para o elemento *Creator*, puderam-se observar valores superiores a 81% com exceção das coleções dos museus Casa Histórica de Alcântara, Museu Histórico Nacional – Moedas de Ouro, Victor Meirelles, Villa-Lobos, da Abolição, de Arqueologia de Itaipu e de Arte Religiosa e Tradicional com taxa de adequação em 0%.

Destacam-se negativamente os elementos *Inscription*, *Measurements* e *Physical Description* com a completa inadequação em todas as coleções.

O elemento *Location* apresentou inadequação em todas as coleções com exceção do Museu da Arqueologia de Itaipu com 100% de adequação.

Destaca-se positivamente também o elemento *MaterialsandTechniques* com adequação em 93% na coleção do Museu da Abolição e 94% no Museu da Inconfidência, possuindo as demais coleções taxas superiores a 97%, exceto para a coleção do Museu Villa-Lobos com adequação em 0%.

Outro elemento que apresentou alguma taxa de adequação é o *Work Type* com 0% na maioria das coleções, exceto para as coleções do Museu Vitor Meirelles com 77% de adequação e 100% para as coleções dos museus Regional do Caeté, Solar Monjardim – Coleção Museológica, e o da Abolição.

Dentre os elementos discricionais ora elencados, os que apresentaram maior taxa de adequação foram *Class*, *Creation Location*, *Creator* e *Materials and Techniques*. E os elementos com menor taxa de adequação foram *Inscription*, *Location*, *Measurements*, *Physical Description* e *WorkType*.

Por fim, cabe destacar, no caso do elemento *Class*, a sua adequação se deve ao fato ao uso dos 2 (dois) instrumentos de linguagem documentária praticados pelos museus sob gestão do Ibram, o tesouro com versão mais recente (publicado em 2016), e que substituiu o anterior denominado ‘tesouro para acervos museológicos’, e o *Tesouro de Objetos do Patrimônio Cultural dos Museus Brasileiros*, concebido e disponibilizado pela Secretaria de Cultura do Rio de Janeiro e pelo Museu Histórico da Cidade, sob a coordenação técnica de Helena Dodd Ferrez, e com padrão ISO 2788:1986. Nesse sentido, os dois tesouros representam terminologias de acervos de caráter histórico e artístico por meio das quais o profissional da informação classifica os itens de coleção por temáticas, assuntos ou contextos de uso. É importante destacar que o uso desses tesouros é apenas recomendado informalmente pelo Ibram, não havendo uma normativa explícita sugerindo seu uso e nem especificando qual das versões deveria ser adotada, se a de

1987 ou a de 2006. Sabe-se que atualmente ambas se encontram em uso pelos museus e possuem significativas diferenças entre si.

Discussão

A partir do levantamento realizado, observou-se que o INBCM é fundamental como instrumento de inventário, catalogação e preservação do patrimônio museológico. No entanto, o INBCM não fornece diretrizes específicas para a descrição sintática e semântica dos elementos sugeridos, nem estabelece um padrão de entrada para o valor e para o conteúdo dos dados, deixando a decisão de concepção de regras de entrada para o catalogador. Por isso, neste trabalho foi utilizado o CCO, que apesar de não ser um padrão de metadados, é um guia de referência internacional que apoia a qualidade final dos metadados resultantes de um processo de catalogação. Além disso, o CCO pode ser usado em conjunto com diferentes padrões de metadados (BACA *et al.*, 2006), como o *Machine-Readable Cataloging* (MARC21), o *Metadata Object Description Schema* (MODS), o Dublin Core, o VRA Core, entre outros. Conforme apresentado no Quadro 1, o CCO também pode ser mapeado para os elementos descritivos do INBCM, dando condições de remodelagem do modelo de inventário, pensando-o como um futuro esquema de metadados no campo da Museologia.

A análise do alinhamento mostrou que 9 dos 15 elementos descritivos recomendam o uso de vocabulário controlado, instrumento de organização da informação capaz de auxiliar nos processos de análise e descrição de documentos, permitindo a criação padronizada de metadados ao nomear, de forma consistente, os pontos de acesso aos documentos e a informação neles contida; além de auxiliar no processo de busca em um sistema de recuperação de Informação através da padronização e expansão do vocabulário controlado das consultas (MARTINS *et al.*, 2021). Entre esses 9 elementos, apenas 5 apresentaram um índice satisfatório de adequação nas coleções avaliadas, a saber: *Class*, *Creation Location*, *Creator*, *Materials and Techniques* e *Work Type*. O elemento *Class*, foi o elemento com o melhor índice de adequação.

Torna-se importante destacar que o uso de taxonomias, oriundas dos dois tesouros ora comentados, pelos museus do Ibram, por meio do metadado classificação, é um aspecto importante na qualidade de dados das coleções, pois normaliza e padroniza a terminologia que será usada nos processos de busca e recuperação da informação (LANCASTER, 2004), além de ajudar no alcance da interoperabilidade semântica dos dados entre diferentes esquemas de metadados e aplicações (ZENG, 2019). Contudo, os resultados indicam que a utilização de vocabulários controlados de maneira padronizada entre os museus ainda não é totalmente explorada pela gestão do Ibram, indo ao encontro de outras realidades do patrimônio cultural apontadas em sequência, corroborando com os achados neste estudo e reafirmando a

importância do uso de vocabulários controlados para melhoria da qualidade de dados em acervos do patrimônio cultural.

O estudo de Kapidakis (2012) destaca a importância da qualidade dos metadados na interoperabilidade das coleções e na qualidade de todos os resultados de busca. Segundo o autor, definir métricas de qualidade é uma tarefa complexa, pois é preciso considerar os diferentes tipos de metadados e coleções, além das diversas percepções que podem existir sobre o conjunto de dados. O autor aponta ainda a dificuldade de avaliar o uso quando diferentes acervos que fazem uso de diferentes vocabulários controlados são agregados em um mesmo repositório. Para solucionar esse problema, o autor desenvolveu uma métrica de qualidade de dados e de uso de vocabulário controlado nas coleções da *Europeana Local* a fim de ajudar os administradores de coleção a detectar e melhorar os pontos fracos de seus acervos. A métrica foi aplicada tanto às coleções da Europeana quanto a dados externos, para servir como controle. O sistema demonstrou ser útil para mostrar irregularidades nos metadados e manter padrões mínimos de qualidade nas coleções, visto que as coleções que distanciam do padrão são destacadas como pontos de atenção para o administrador da coleção. Dessa forma, conclui-se que a qualidade de dados é algo muito subjetivo pelas formas de conteúdo e pluralidade de acervos.

Em um estudo sobre a utilização do CCO, Coburn et al. (2010) examinaram a influência do CCO e sua implementação em ambientes de catalogação para comunidades de museus e bibliotecas de imagens. A prática emergente de catalogação com o CCO resultou em um significativo corpo de registros dessas comunidades, direcionados para ambientes de acesso integrado LAM (biblioteca/arquivo/museu). Os autores comentam sobre como a tomada de decisões de catalogação (por exemplo, concepções diferentes sobre o que é uma 'obra') pode impactar a convergência de registros nesses ambientes. O estudo conclui que o CCO oferece uma estrutura comum para a catalogação, permitindo que uma variedade de museus, coleções especiais de bibliotecas e coleções pictóricas usem o CCO com diferentes conjuntos de elementos de metadados descritivos e vocabulários controlados especializados para catalogar e compartilhar informações sobre obras culturais.

No Brasil, há uma ampla diversidade cultural, o que resulta em uma grande variedade de museus e assuntos catalogados por estas instituições. Além disso, há também uma gama de profissionais de diferentes áreas trabalhando na catalogação de objetos culturais dessas instituições e levando a uma variedade de práticas catalográficas (MARTINS et al., 2021). Já na Europa, conforme destaca Kapidakis (2012), a plataforma digital Europeana integra várias instituições culturais que utilizam diferentes vocabulários controlados, tornando difícil

identificar o uso correto destes vocabulários controlados e, ao identificá-los, há ainda a dificuldade em determinar qual vocabulário está sendo utilizado.

Com a variedade de vocabulários controlados existentes para diversos domínios e tipos de objetos culturais, há a necessidade de conectar e realizar o intercâmbio entre descrições que usam diferentes vocabulários controlados. Conforme apresentado por Kapidakis (2012) e Coburn et al. (2010), o alinhamento e o uso de padrões em diferentes acervos culturais melhoram a catalogação, o que por sua vez melhora o acesso à informação para o usuário final. No entanto, o grande espectro de acervos culturais possui características diferentes, o que leva ao uso de diferentes vocabulários controlados. Ainda neste cenário de utilização de padrões adequados de catalogação, há desafios na agregação desses diferentes acervos em um único repositório, pois é preciso considerar o alinhamento entre as diferentes percepções que podem existir sobre o conjunto de dados. Assim, pode-se observar os desafios que o Ibram enfrenta na missão de agregar e interligar os dados de diferentes museus no Brasil (LEMOS; COELHO JUNIOR, 2023).

Considerações finais

O diagnóstico revelou que, apesar de esses museus e coleções estarem sob a gestão do Ibram, as práticas de catalogação não estão alinhadas entre si, com grandes diferenças nas práticas utilizadas, como nos elementos discricionais *Work Type* e *Creator*. Verificou-se também que os dados das coleções avaliadas precisam de um tratamento mais adequado nos elementos *Inscription*, *Location*, *Measurements*, *Physical Description* e *Work Type*. Por outro lado, as coleções se mostraram qualificadas em termos do uso adequado de taxonomias para o elemento classificação. Logo, a questão de pesquisa se responde pela evidência dos níveis de padronização e normalização dos metadados produzidos pelos museus sob gestão do Ibram por meio do emprego de vocabulários controlados à luz das recomendações do guia de referência CCO.

Constatou-se que há uma divergência nas práticas de catalogação dos museus, especialmente no uso de vocabulário controlado, dificultando a interoperabilidade, o reúso de dados e o acesso ao usuário final.

Recomenda-se, portanto, a realização de intercâmbio de boas práticas de catalogação entre as instituições museológicas analisadas, além da incorporação de práticas de catalogação maduras oriundas de instrumentos de referência, como o CCO, na modelagem de metadados das bases de dados dos museus. Além disso, recomenda-se a qualificação dos profissionais de catalogação, por meio de formações em qualidade de dados para a catalogação de acervos museológicos em ambiente digital, afim de que esses profissionais tenham conhecimento de

boas práticas de catalogação contemporâneas visando qualificar seus atuais padrões de documentação por meio de instrumentos de organização da informação oriundos do campo da Ciência da Informação, como padrões de metadados, vocabulários controlados e regras de catalogação.

Referências

ANSI/NISO Z39.19-2005 (R2010). **Guidelines for the construction, format, and management of monolingual controlled vocabularies**. Baltimore: NISO Press, p. 184, 2005.

BACA, M.; HARPRING, P.; LANZI, E.; MCRAE, L.; WHITESIDE, A. **Cataloging cultural objects: a guide to describing cultural works and their images**. Chicago: American Library Association, 2006.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked Data - The Story So Far. **International Journal on Semantic Web and Information Systems (IJSWIS)**, v. 5, n. 3, p. 1–22, 2009.

BRASIL. MINISTÉRIO DO TURISMO. INSTITUTO BRASILEIRO DE MUSEUS – Ibram. **Museus Ibram – Instituto Brasileiro de Museus**. Brasília, 2021. Disponível em: <https://antigo.museus.gov.br/museus-ibram/>. Acesso em: 28 nov. 2021.

CHAPMAN, A. D. **Principles of Data Quality**. Copenhagen, 2005. DOI: [10.15468/DOC.JRGG-A190](https://doi.org/10.15468/DOC.JRGG-A190). Disponível em: <https://www.gbif.org/document/80509>. Acesso em: 28 jul. 2022.

COBURN, E. *et al.* The Cataloging Cultural Objects experience: Codifying practice for the cultural heritage community. **IFLA Journal**, v. 36, n. 1, p. 16–29, mar. 2010.

COELHO, A. **Qualidade dados IBRAM**. 2022. Disponível em: https://github.com/AbeilCoelho/Qualidade_dados_IBRAM. Acesso em: 06 mar. 2023.
DODEBEI, V.L.D. **Tesouro: linguagem de representação da memória documentária**. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002.

GILLILAND, A. J. Setting the Stage. In: BACA, M. (ed.). **Introduction to metadata**. 3. ed. Los Angeles: Getty Research Institute, 2016. Disponível em: <https://www.getty.edu/publications/intrometadata/setting-the-stage/>. Acesso em: 22 jul. 2022.

HARPRING, P. **Metadata Standards Crosswalks**. [S. l.], 2022. Disponível em: https://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html#endnote1CCO. Acesso em: 17 jul. 2022.

HJØRLAND, B. Semantics and Knowledge Organization. **Annual Review of Information Science and Technology**, v. 41, p. 367–405, 2007.

INSTITUTO BRASILEIRO DE MUSEUS (IBRAM). **Sobre o órgão – Instituto Brasileiro de Museus – Ibram**. Disponível em: <https://www.museus.gov.br/sobre-o-orgao/>. Acesso em: 06 mar. 2023.

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: 06 mar. 2023.

KAPIDAKIS, S. **Comparing metadata quality in the Europeana context.** In: PETRA2012: THE 5TH INTERNATIONAL CONFERENCE ON PERVASIVE TECHNOLOGIES RELATED TO ASSISTIVE ENVIRONMENTS. Heraklion Crete Greece: ACM, 6 jun. 2012. **Proceedings.** Disponível em:
<<https://dl.acm.org/doi/10.1145/2413097.2413129>>. Acesso em: 10 jan. 2023.

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

LEMOS, D.L.D.S.; COELHO JUNIOR, A. 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.

MACEDO, D.F.; LEMOS, D.L.D.S. Dados abertos governamentais: iniciativas e desafios na abertura de dados no Brasil e outras esferas internacionais. **AtoZ: novas práticas em informação e conhecimento**, Curitiba, v. 10, n. 2, p. 14, 2021.

MARTINS, D. L.; LEMOS, D.L.S.; ANDRADE, M. C. 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**, v. 26, n. 2, p. 569-595, 2021.

MARTINS, D.L. *et al.* 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.

MARTINS, D. L.; LEMOS, D. L. S.; CARMO, D.; SIQUEIRA, J.; OLIVEIRA, L. F. R. Requisitos de qualidade para dados de agregação em museus. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 14, 2021.

SILVA, D.L.D.; SOUZA, R.R.; ALMEIDA, M.B. Ontologias e vocabulários controlados: comparação de metodologias para construção. **Ciência da informação**, Brasília, v. 37, p. 60-75, 2008.

SVENONIUS, E. **The Intellectual Foundation of Information Organization.** Boston: MIT Press, 2000

WILKINSON, M.D. et al. The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, [S. l.], v. 3, n. 1, p. 160018, 2016.

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