

# TRATAMENTO DA INFORMAÇÃO EM ACERVOS CULTURAIS: avaliação do uso de vocabulários controlados em coleções museológicas sob gestão do Instituto Brasileiro de Museus

*TREATMENT OF INFORMATION IN CULTURAL COLLECTION: evaluation of the use of controlled vocabularies in museum collections managed by the Brazilian Institute of Museums*

**Abeil Coelho Júnior<sup>1</sup>, Daniela Lucas da Silva Lemos<sup>2</sup>**

(1) Universidade Federal do Espírito Santo, Vitória-ES, abeile@hotmail.com.

(2) Universidade Federal do Espírito Santo, Vitória-ES, daniela.l.silva@ufes.br.

## Resumo:

Nos últimos anos, tem-se observado considerável adesão das instituições de patrimônio cultural ao processo de digitalização e disponibilização de seus dados de acervos na internet, proporcionando maior acesso e democratização de conhecimento científico e cultural à sociedade. Porém, apenas disponibilização de dados *online* não é o bastante atualmente, considerando que a qualidade desses dados deve ser mensurada. Assim, este trabalho busca avaliar a adequação das coleções museológicas disponibilizadas *online* pelo Instituto Brasileiro de Museus acerca do uso de vocabulário controlado, à luz das orientações do guia de catalogação de objetos culturais de referência, o *Cataloging of Cultural Objects*. Para tal, foi realizado o alinhamento dos metadados utilizados pelo Instituto Brasileiro de Museus e o guia de catalogação, desenvolvido *script* em Python para processamento dos dados das coleções, além de cálculo de adequação a recomendação de uso de vocabulário controlado. Os resultados demonstram que apenas 5 (cinco) dos 9 (nove) elementos recomendados fazem uso de vocabulário controlado; apresenta ainda que, apesar desses museus e coleções estarem sob gestão do Instituto Brasileiro de Museus, as práticas de catalogação não estão alinhadas entre as instituições. Conclui-se que práticas de catalogação provenientes de instrumentos de referência sejam incorporadas na modelagem de metadados das bases de dados dos museus sob gestão do Instituto Brasileiro de Museus, visando qualificar seus atuais padrões de documentação.

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

## Abstract:

In recent years, there has been considerable adhesion of cultural heritage institutions to the process of digitizing and making their collection data available on the internet, providing greater access and democratization of scientific and cultural knowledge to society. However, just making data available online is not enough nowadays, considering that the quality of this data must be measured. Thus, this work seeks to evaluate the adequacy of museum collections made available online by the Brazilian Institute of Museums regarding the use of controlled vocabulary, in the light of the guidelines of the cataloging guide for cultural objects of reference, the *Cataloging of Cultural Objects*. To this end, the alignment of the metadata used by the Brazilian Institute of Museums and the cataloging guide was carried out, a script in Python was developed for processing the data from the collections, in addition to calculating the suitability for the recommendation of using a controlled vocabulary. The results show that only 5 of the 9 elements recommended use controlled vocabulary; also shows that, despite these museums and collections being managed by the Brazilian Institute of Museums, cataloging practices are not aligned between the institutions. It is concluded that cataloging practices from reference instruments are incorporated into the metadata modeling of the databases of museums managed by the Brazilian Institute of Museums, in order to qualify their current documentation standards.

**Keywords:** Museums. Information Organization. Information Representation. Data quality. Controlled vocabulary.

## 1. Introdução

Nos últimos anos, tem-se observado considerável adesão das instituições de patrimônio cultural ao processo de digitalização e disponibilização de seus dados de acervos na internet, proporcionando maior acessibilidade e

democratização de conhecimento científico e cultural à sociedade. Entretanto investir no processo de digitalização e disponibilização de objetos culturais não é suficiente (MARTINS et al., 2022), visto que questões acerca da qualidade de dados nesses processos frequentemente não são

levantadas, considerando os diversos tipos de bancos de dados e sistemas de informação ora envolvidos em processos de organização, modelagem e representação.

Diante deste cenário, dados se tornam cada vez mais recursos importantes e valiosos para o século XXI. Dessa forma, considerações acerca da importância da qualidade para publicação de conjunto de dados na internet surgem em diversos contextos (BIZER; HEATH; BERNERS-LEE, 2009; WILKINSON et al., 2016; MACEDO; LEMOS, 2021; MARTINS et al., 2022).

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 corretamente 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 reúso de seus dados, instituições de patrimônio cultural (HARPRING, 2022) geralmente aderem a padrões de catalogação que produzem descrições de coleções de objetos culturais digitalizados ou nato digitais, os quais necessitam padronizar 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.

Nesse sentido, como forma de garantir a qualidade de dados de uma coleção cultural, faz-se necessário a adoção de padrões de tratamento documental atrelados ao uso de vocabulários controlados.

Vocabulário controlado é definido por Lancaster (2004) 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.

Os vocabulários controlados podem auxiliar: i) 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; e ii) no processo de busca em um sistema de recuperação de Informação através da padronização e expansão do vocabulário controlado das consultas. Exemplos de vocabulários controlados incluem: esquemas de classificação, listas de cabeçalhos de assuntos, tesouros, taxonomias e ontologias (ABBAS, 2010).

Como padrões de tratamento documental, Abbas (2010) destaca dois desses padrões que vislumbram a produção de bases de dados no âmbito do patrimônio cultural, especialmente para museus, a saber: i) o padrão semântico *Categories for the Description of Works of Art* (CDWA) e sua extensão *Cataloging of Cultural Objects* (CCO); e ii) o padrão de metadados *Visual Resources Association Core Categories* (VRA Core).

O CCO foi publicado em 2006, como resultado do consenso de profissionais das comunidades de museus, bibliotecas, galerias e arquivos que pesquisam a prática comum entre essas disciplinas (BACA et al., 2006, HARPRING, 2022) e 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), *Metadata Object Description Schema* (MODS), Dublin Core, entre outros.

No caso do Instituto Brasileiro de Museus (Ibram), suas bases de dados 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).

## 2. Objetivos

O objetivo geral desta pesquisa é fazer avaliação da qualidade de dados na perspectiva do uso de vocabulário controlado nos elementos de metadados que descrevem as coleções dos museus sob gestão do Ibram.

## 3. Procedimentos Metodológicos

O INBCM possui elementos discricionais para coleções museológicas, bibliográficas e acervos arquivísticos. Porém, para os objetivos deste trabalho foram considerados apenas os elementos de caráter museológico. 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, conforme Apêndice A. O alinhamento se deu a partir de um procedimento manual e intelectual baseado na aquisição de conhecimento sobre os dois instrumentos de pesquisa.

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*<sup>1</sup>, *BeautifulSoup*<sup>2</sup> e *Requests*<sup>3</sup>, para realizar a exportação em massa de todos os dados dos acervos dos museus no formato “CSV: inbcm-ibrammapper” do Tainacan. Sabendo-se que a ferramenta de repositório usada pelos museus sob gestão Ibram adota vocabulários controlados do tipo taxonomia na modelagem de seus metadados, fez-se a verificação do seu emprego em cada elemento de metadado, alinhado com o

CCO, com exigência de uso de vocabulário controlado. 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 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 tivesse 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 foram calculados a taxa de adequação a regra de uso de vocabulário controlado do CCO com a seguinte fórmula:

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

**Onde:** “b” é a base com dados de uma coleção em particular; **índice** é o percentual de adequação obtido a nível de dimensão, elemento de metadado e regra de catalogação para um determinado museu e coleção; **Valor1** é a indicação de ocorrência do registro de dado que **se adequou** a regra; **Valor0** é a indicação de ocorrência do registro de dado que **não atendeu** a regra.

## 4. Resultados

A luz do alinhamento entre elementos de metadados do INBCM e CCO (Apêndice A) pode-se observar pela coluna “Vocabulário Controlado CCO” que para 9 (nove) dos 15 elementos são recomendados o uso de vocabulário controlado pelo CCO.

<sup>1</sup> <https://pandas.pydata.org>

<sup>2</sup> <https://beautiful-soup-4.readthedocs.io/>

<sup>3</sup> <https://requests.readthedocs.io/en/master/>

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 no Apêndice B. Observa-se nas linhas, os elementos de metadados alinhados com regras de catalogação de uso de vocabulário controlado e nas colunas as coleções museológicas disponíveis online pelo Ibram. Cada célula apresenta a taxa de adequação das coleções à regra de uso de vocabulário controlado, com valores de 0 (zero), onde foi observado a completa inadequação das coleções à regra; a 100, em que houve a completa adequação da regra pela coleção. Assim, as cores mais claras representam menores taxas de adequação e as mais fortes maiores taxas de adequação.

Pode-se destacar o uso de vocabulário controlado no elemento de metadado *Class* com menor taxa de adequação em 99% nas coleções dos museus da Abolição; Bandeiras e Missões. Possuindo as demais coleções a adequação em 100%. Destaca-se positivamente também o elemento *Materials and Techniques* 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 a coleção do Museu Villa-Lobos com adequação em 0%.

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

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 objetos com valores vazios. Neste caso sem valor taxonômico informado. Os demais apresentaram o uso de taxonomia e todos os objetos haviam valores informados.

Para o elemento *Creator* pode-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%.

Outro elemento que apresentou alguma taxa de adequação é o *Work Type* com 0% na maioria das coleções, exceto 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.

## 5. Considerações Finais

O CCO não é um padrão de metadados, mas os conceitos e elementos apresentados podem ser mapeados para vários elementos de metadados (BACA et al., 2006), como, por exemplo, o *Machine-Readable Cataloging* (MARC21), o *Metadata Object Description Schema* (MODS), o Dublin Core, o VRA Core, e, desta forma, também pode ser mapeado para os elementos descritivos do INBCM, conforme se comprovou no alinhamento (Apêndice A).

Destaca-se ainda que o uso de vocabulário controlado é recomendado por 9 (nove) dos 15 elementos de descrição alinhados entre o CCO e INBCM. Desses 9 (nove), apenas 5 (cinco) deles apresentaram algum índice de adequação nas coleções avaliadas, a saber: *Class*, *Creation Location*, *Creator*, *Materials and Techniques* e *Work Type*. O elemento *Class*, alinhada com o elemento INBCM Classificação, foi a que apresentou melhor índice de adequação dentre as coleções analisadas. Pondera-se, assim, que o uso de taxonomia 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).

Finalmente, o diagnóstico permitiu aferir que apesar desses museus e coleções

estarem sob gestão do Ibram, as práticas de catalogação não estão alinhadas entre elas, com grandes discrepâncias nas práticas utilizadas a exemplo dos elementos discricionais *Work Type* e *Creator*. Podemos aferir ainda que os dados das coleções avaliadas carecem 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. Desta forma, recomenda-se, que práticas de catalogação maduras oriundas de instrumentos de referência sejam incorporadas na modelagem de metadados das bases de dados dos museus sob gestão do Ibram, visando qualificar seus atuais padrões de documentação por meio de instrumentos de organização da informação e orientados para usuários finais de sistemas de informação.

## 6. Referências

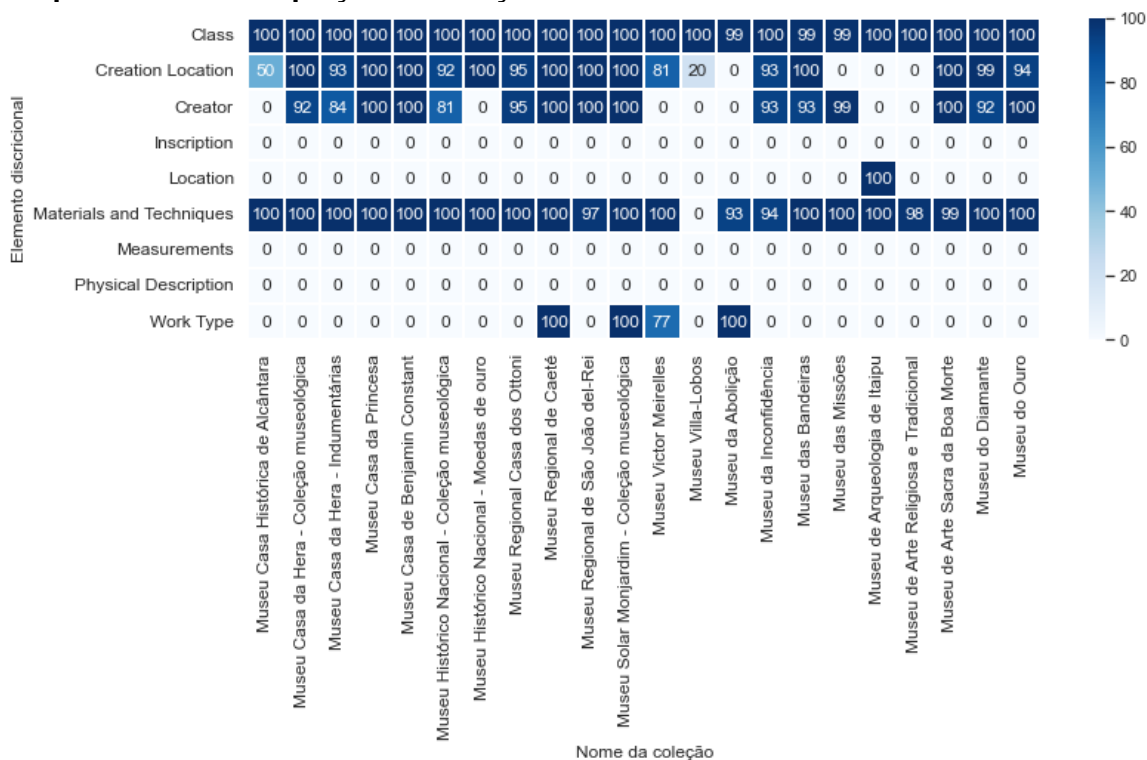
- ABBAS, June. **Structures for organizing knowledge: exploring taxonomies, ontologies, and other schema**. New York: Neal-Schuman Publishers, 2010.
- BACA, Murtha; HARPRING, Patricia; LANZI, Elisa; MCRAE, Linda; WHITESIDE, Ann. **Cataloging cultural objects: a guide to describing cultural works and their images**. Chicago: American Library Association, 2006.
- BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. 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, Arthur D. Principles of Data Quality. Copenhagen, 2005. DOI: 10.15468/DOC.JRGG-A190. Disponível em: <https://www.gbif.org/document/80509>. Acesso em: 28 jul. 2022.
- COELHO, Abeil. **Qualidade\_dados\_IBRAM**. 2022. Disponível em: [https://github.com/AbeilCoelho/Qualidade\\_dados\\_IBRAM](https://github.com/AbeilCoelho/Qualidade_dados_IBRAM). Acesso em: 10 set. 2022.
- GILLILAND, Anne J. Setting the Stage. In: BACA, Murta. (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, Patricia. **Metadata Standards Crosswalks**. [S. l.], 2022. Disponível em: [https://www.getty.edu/research/publications/electronic\\_publications/intrometadata/crosswalks.html#endnote1CCO](https://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html#endnote1CCO). Acesso em: 17 jul. 2022.
- LANCASTER, Frederick Wilfrid. **Indexação e resumos: teoria e prática**. Brasília: Briquet de Lemos, 2004.
- MACEDO, Dirceu Flávio; LEMOS, Daniela Lucas da Silva. Dados abertos governamentais: iniciativas e desafios na abertura de dados no Brasil e outras esferas internacionais. **AtoZ: novas práticas em informação e conhecimento**, [S. l.], v. 10, n. 2, p. 14, 2021.
- MARTINS, Dalton Lopes 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.
- WILKINSON, Mark 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, Marcia Lei. Interoperability. **Knowledge Organization**, v.46, n.2, p. 122-146, jan. 2019.

## Apêndice A – Alinhamento entre elementos descritivos – INBCM e CCO

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

Fonte: elaborado pelos autores.

## Apêndice B – Adequação de coleções Ibram ao uso de vocabulário controlado



Fonte: elaborado pelos autores.