

# WIDaT 2022

Workshop de Informação, Dados e Tecnologia

05 e 06 DE DEZEMBRO DE 2022

UFES - Vitória

## **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 – Ufes

Daniela Lucas da Silva Lemos – Ufes

# INTRODUÇÃO

Durante as últimas décadas, constatou-se considerável adesão das instituições de patrimônio cultural a processos de digitalização e disponibilização de seus dados de acervos na internet (MARTINS et al., 2022).

Entretanto investir no processo de digitalização e disponibilização de objetos culturais não é suficiente, visto que questões acerca da qualidade de dados nesses processos frequentemente não são levantadas.

Diante deste cenário, dados se tornam cada vez mais recursos importantes e valiosos para o século XXI.

# INTRODUÇÃO

Custodiadores e proprietários de dados, como, por exemplo, galerias, bibliotecas, arquivos e museus – GLAMs, são os principais responsáveis pela qualidade de seus dados, com uma boa catalogação descritiva e de assunto.

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, e, por consequência maior acesso aos documentos pelo usuário final.

# INTRODUÇÃO

Com objetivo de viabilizar a interoperabilidade, acesso e reúso de seus dados, instituições de patrimônio cultural (HARPRING, 2022) geralmente aderem a padrões de catalogação.

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.



# INTRODUÇÃO

Os vocabulários controlados podem auxiliar:

- i) nos processos de análise e descrição de documentos,
- ii) no processo de busca em um sistema de recuperação de Informação.

Exemplos de vocabulários controlados incluem: esquemas de classificação, listas de cabeçalhos de assuntos, tesauros, taxonomias e ontologias (ABBAS, 2010).

# INTRODUÇÃO

Como padrões de tratamento documental, Abbas (2010) destaca dois desses padrões que vislumbram a produção de bases de dados no âmbito do patrimônio cultural:

- 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**).

# INTRODUÇÃO

Publicado em 2006, o CCO foi 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). Embora inspirado no VRA Core e do *Getty Vocabularies*.

O CCO apresenta conceitos mais genéricos, podendo ser utilizados com outros conjuntos de metadados, como o MARC21, MODS, Dublin Core, entre outros.

# INTRODUÇÃO

O 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).



# OBJETIVO

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.

# METODOLOGIA

O INBCM possui elementos discricionais para coleções museológicas, bibliográficas e acervos arquivísticos. Porém, para os objetivos deste trabalho foram considerados apenas os elementos de carácter museológico.

Desta forma, o primeiro passo foi realizar o alinhamento (mapeamento) entre os elementos descritivos de carácter museológico do INBCM com os elementos recomendados pelo guia de catalogação CCO.

# ALINHAMENTO ENTRE INBCM E CCO

Elemento CCO	Obrigatório CCO	Vocabulário Controlado CCO	Elemento INBCM	Obrigatório INBCM
<b>Work Type</b>	Sim	Sim	Denominação	Sim
<b>Title</b>	<b>Sim</b>	Não	Título	<b>Não</b>
<b>Creator</b>	Sim	Sim	Autor	Sim
<b>Measurements</b>	Sim	Sim	Dimensões	Sim
<b>Materials and Techniques</b>	Sim	Sim	Material/Técnica	Sim
<b>Physical Description</b>	Não	Sim	Estado de Conservação	Sim
<b>Inscription</b>	Sim	Sim	Número de Registro	Sim
<b>Date</b>	Sim	Não	Data de Produção	Não
<b>Creation Location</b>	Não	Sim	Local de Produção	Não
<b>Class</b>	<b>Sim</b>	Sim	Classificação	<b>Não</b>
<b>Description</b>	Não	Não	Resumo Descritivo	Sim
<b>Other Descriptive Notes</b>	Não	Não	Condições de Reprodução	Sim
<b>Related Works</b>	Não	NA	Mídias Relacionadas	Não
<b>Location</b>	Sim	Sim	Situação	Sim
<b>NA</b>	NA	NA	Outros Números	Não

# CAPTAÇÃO DOS DADOS

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*, *BeautifulSoup* e *Requests*.

Com isso foi realizado a exportação em massa de todos os dados dos acervos dos museus no formato “CSV: inbcm-ibrammapper” dos Tainacans.

Sabendo-se que os repositórios utilizados pelos museus sob gestão Ibram adotam 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.



# CAPTAÇÃO DOS DADOS

Esta 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.

# CAPTAÇÃO DOS DADOS

Por fim, foram captados os dados de todas as 22 coleções de caráter museológico disponíveis online pelo Ibram. E avaliados se os elementos de metadados nestas coleções possuíam a indicação de taxonomia.

# AVALIAÇÃO REALIZADA

para cada coleção e elementos de metadados que possuíssem 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.

Por fim, o índice de adequação é dado pela seguinte fórmula:

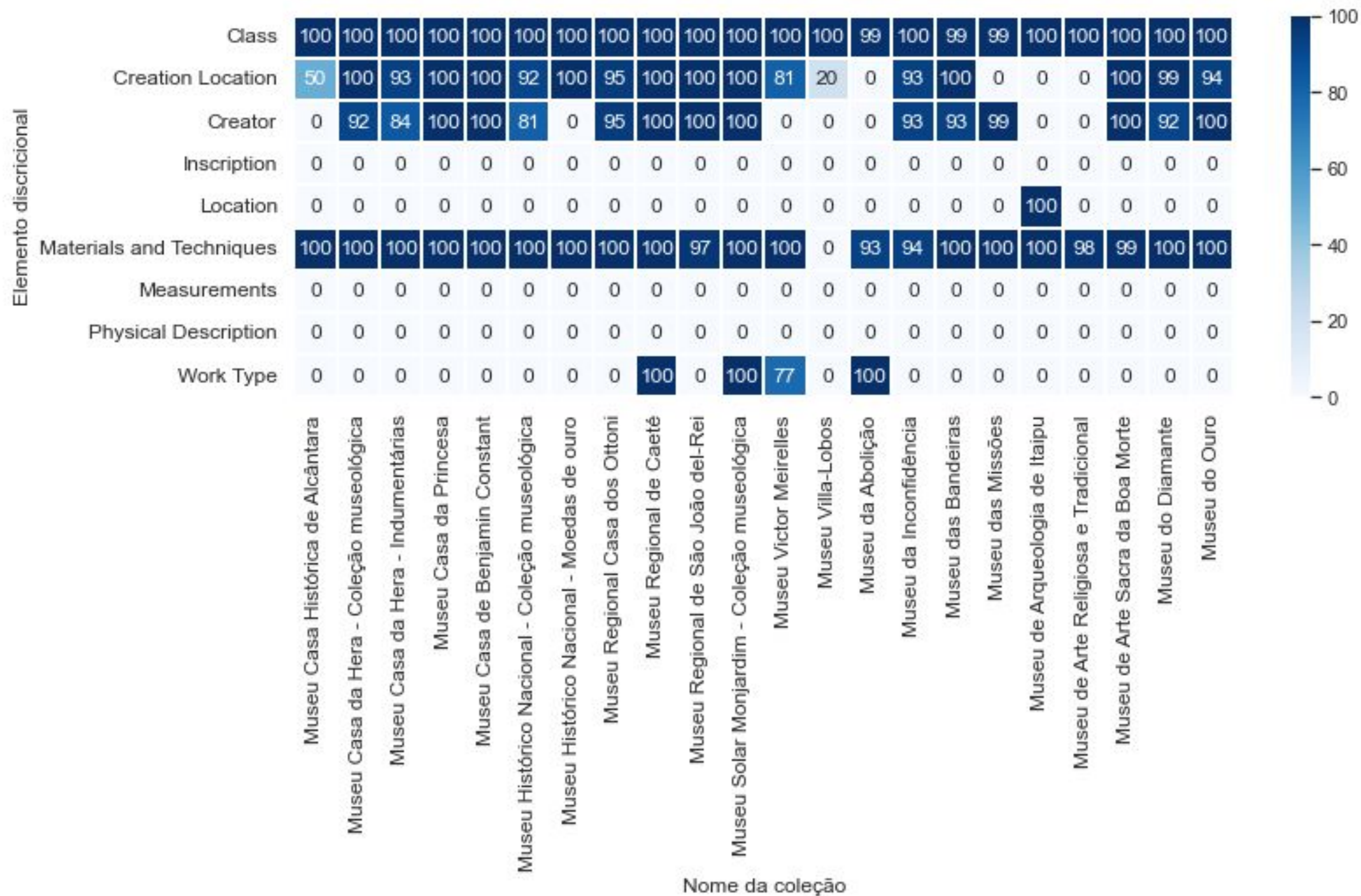
$$\text{índice b} = (\sum \text{Valor1} / (\sum \text{Valor1} + \sum \text{Valor0})) * 100$$

# RESULTADOS

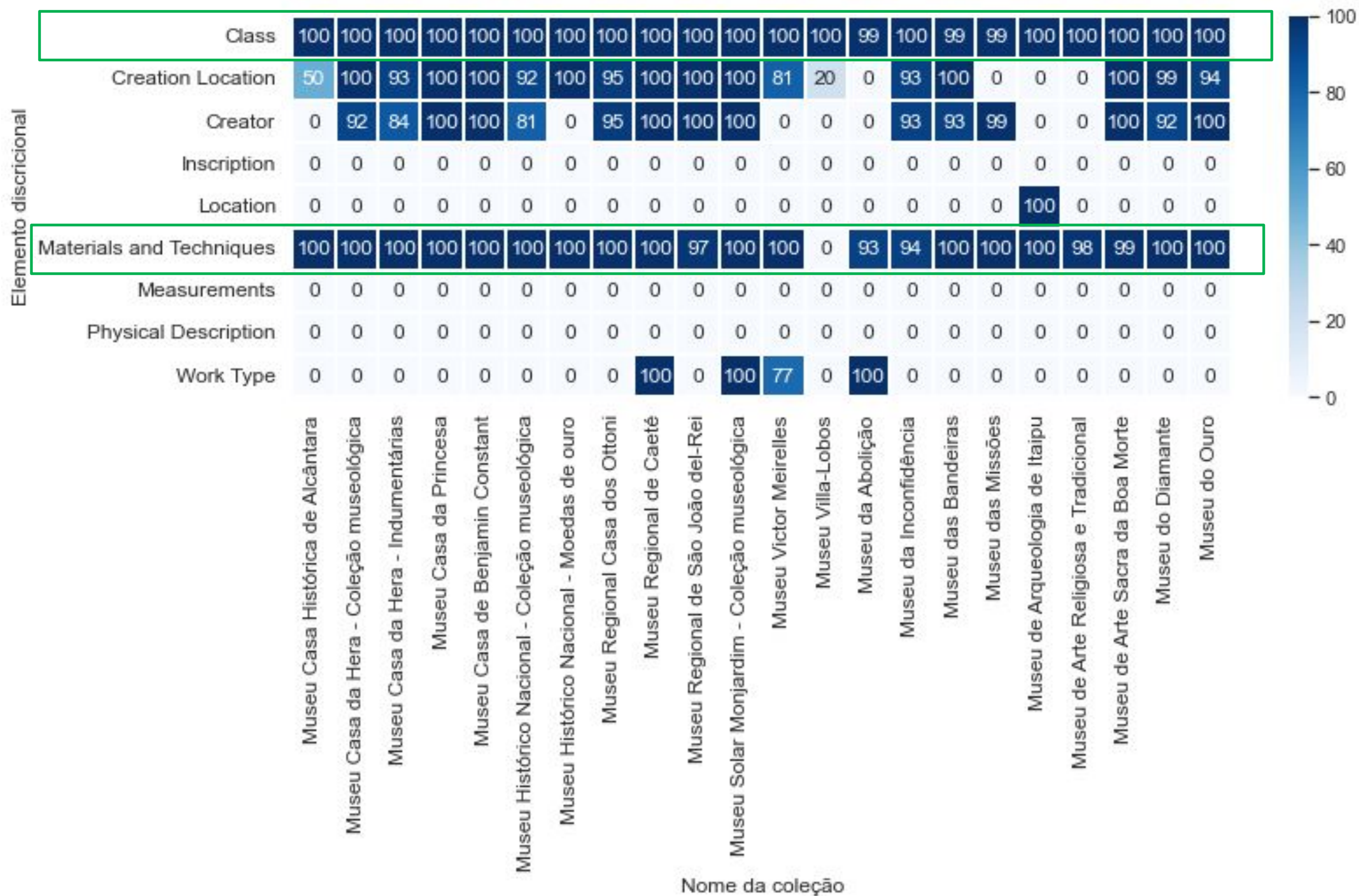
A luz do alinhamento entre elementos de metadados do INBCM e CCO, pode-se observar pela coluna “Vocabulário Controlado CCO” que para 9 (nove) dos 15 elementos são recomendados o uso de vocabulário controlado pelo CCO.



# RESULTADOS

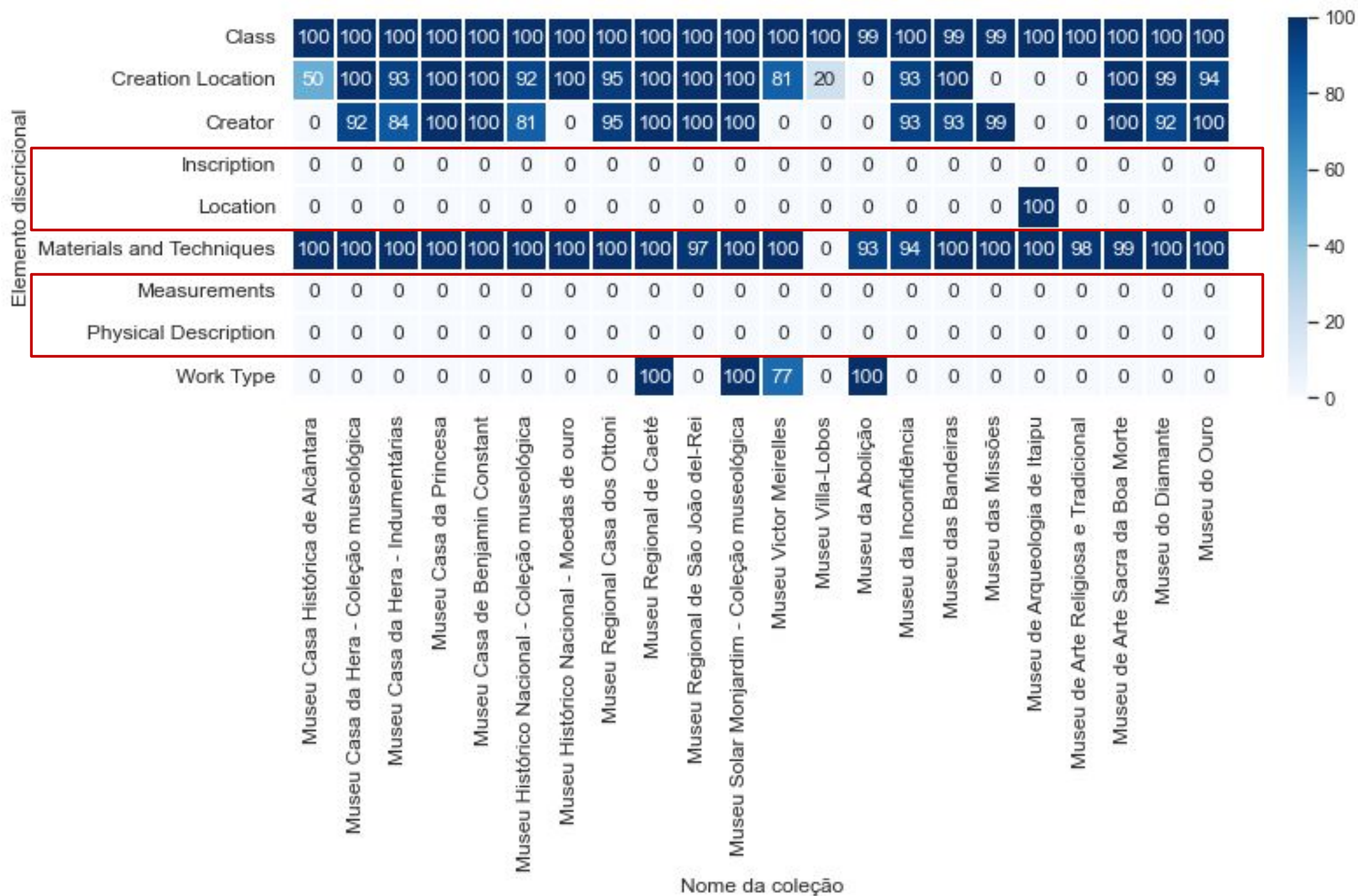


# RESULTADOS





# RESULTADOS



# 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).

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) apresentaram algum índice de adequação nas coleções avaliadas, a saber: Class, Creation Location, Creator, Materials and Techniques e Work Type.



# CONSIDERAÇÕES FINAIS

O elemento Class, alinhada com o elemento INBCM Classificação, foi a que apresentou melhor índice de adequação dentre as coleções analisadas.

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),

# CONSIDERAÇÕES FINAIS

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.

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.

# CONSIDERAÇÕES FINAIS

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.

Com foco em qualificar seus atuais padrões de documentação por meio de instrumentos de organização da informação e orientados para usuários finais.



# REFERÊNCIAS

ABBAS, June. **Structures for organizing knowledge**: exploring taxonomies, ontologies, and other schema. New York: Neal-Schuman Publishers, 2010.

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.

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.

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.