

QUALIDADE DE DADOS EM ACERVOS DO PATRIMÔNIO CULTURAL: UMA AVALIAÇÃO DIAGNÓSTICA SEMIAUTOMÁTICA NOS OBJETOS CULTURAIS SOB GESTÃO DO INSTITUTO BRASILEIRO DE MUSEUS

Data quality in cultural heritage collection: a semi-automatic diagnostic evaluation in cultural objects under management of the Brazilian Institute of Museums

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
<https://orcid.org/0000-0003-1565-7366> 

Abeil Coelho Júnior

Universidade Federal do Espírito Santo,
Programa de Pós-graduação em Ciência da Informação
Vitória, ES, Brasil.
E-mail: abeiloc@hotmail.com
<https://orcid.org/0000-0003-1447-9537> 

A lista completa com informações dos autores está no final do artigo 

RESUMO

Objetivo: apresentar o resultado de avaliação diagnóstica de qualidade de dados realizada nas bases de dados dos museus vinculados ao Ibram frente às boas práticas de catalogação indicadas no guia de referência no campo da cultura digital, o *Cataloging Cultural Objects*.

Método: a exploração dos dados foi realizada por solução semiautomática com o uso de *script* em linguagem Python em 22 coleções de caráter museológico, representando mais de 17 mil itens de acervos das instituições sob gestão do Ibram.

Resultados: O diagnóstico permitiu aferir que os dados das coleções avaliadas carecem de um tratamento mais adequado em dimensões como características físicas do objeto de informação, descrição, localização geográfica e informações cronológicas. Por outro lado, as coleções se mostraram qualificadas em termos do uso adequado de taxonomias para a dimensão classificação.

Conclusões: Recomenda-se, assim, 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 mais sofisticados e orientados para usuários finais de sistemas de informação.

PALAVRAS-CHAVE: Organização da informação. Qualidade de dados. Padrões de documentação. Museologia. Acervos digitais.

ABSTRACT

Objective: the objective of the article is to present the result of a diagnostic evaluation of data quality carried out in the databases of museums linked to Ibram in view of the good cataloging practices indicated in the reference guide in the field of digital culture, the *Cataloging Cultural Objects*.

Methods: the exploration of the data was carried out by a semi-automatic solution using a script in Python programming language in 22 museum collections, representing more than 17 thousand items from the collections of the institutions under Ibram's management.

Results: The diagnosis made it possible to verify that the data from the evaluated collections lack a more adequate treatment in dimensions such as physical characteristics of the information object, description, geographic location and chronological information. On the other hand, the collections proved to be qualified in terms of the proper use of taxonomies for the classification dimension.

Conclusion: The diagnosis made it possible to verify that the data from the evaluated collections lack a more adequate treatment in dimensions such as physical characteristics of the information object, description, geographic location and chronological information. On the other hand, the collections proved to be qualified in terms of the proper use of

taxonomies for the classification dimension. It is therefore recommended that mature cataloging practices derived from reference instruments be incorporated into the metadata modeling of the databases of museums managed by Ibram, in order to qualify their current documentation standards through more sophisticated information organization instruments and oriented towards end users of information systems.

KEYWORDS: Information organization. Data quality. Documentation standards. Museology. Digital collections.

1 INTRODUÇÃO

Considerando que “dados” são recursos importantes e valiosos para o século XXI, 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, vislumbrando o objeto cultural digital como elemento predominante na geração de valor por meio de sua circulação e reutilização social (MARTINS *et al.*, 2022). Nesse mesmo caminho, questões ligadas à importância da qualidade de dados para publicação de conjunto de dados (do inglês, *datasets*) abertos na internet também surgiram nas últimas décadas em contextos diversos, incluindo o científico, o cultural e o governamental (BIZER; HEATH; BERNERS-LEE, 2009; SIQUEIRA *et al.*, 2021; MACEDO; LEMOS, 2021), apenas para citar alguns.

Nesse contexto, desde o ano de 2014, o Instituto Brasileiro de Museus (Ibram) vem adotando estratégias para o desenvolvimento de uma rede interoperável de agregação de repositórios digitais para o cenário cultural brasileiro envolvendo os museus sob sua gestão (SIQUEIRA; MARTINS, 2021; MARTINS *et al.*, 2021), intencionando maior acessibilidade e democratização de conhecimento científico e cultural à sociedade. Assim, em parceria com a Universidade Federal de Goiás (UFG), o Ibram aderiu, em 2016, a plataforma digital Tainacan (GOV.BR, 2021) para preservação, difusão e integração dos acervos de suas instituições, representadas por 30 museus federais, buscando maior interação por parte dos usuários com mecanismos de busca e navegação mais sofisticados e soluções mais inteligentes de curadoria e gestão de acervos na internet. Dos 30 museus sob sua gestão, o Ibram possui 20 museus que utilizam o Tainacan para disponibilizar 23 coleções na internet.

Contudo, a ausência de padrões e práticas de catalogação consistentes pelos museus ou qualquer instituição de patrimônio cultural envolvida nessa rede interoperável de agregação (SIQUEIRA *et al.*, 2021) pode comprometer a identificação de informações cruciais e necessárias para descrever um item, de modo a localizá-lo nas bases de dados para fins de busca, recuperação e interoperabilidade em rede (WYNAR, 1985; MEY, 1995; INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND

INSTITUTIONS, 2016; LEMOS; SÁ; SOUZA, 2019). Tal situação dificulta o provimento de possíveis meios de agregar os dados de acervos com outros conjuntos de dados culturais disponíveis na rede, além de inviabilizar a construção de infraestruturas de coleções culturais digitais brasileiras para reuso em outras aplicações com demandas sociais e humanas de conhecimento, como no campo emergente das Humanidades Digitais (POOLE, 2017), por exemplo.

Em suma, a obtenção da qualidade de dados em coleções de acervos é consequência da modelagem apropriada de metadados e sua aplicação pelo profissional da informação em processos de organização e representação da informação e do conhecimento (SVENONIUS, 2000; TAYLOR, 2004; HJORLAND, 2007; GILLILAND, 2016) endereçados a soluções de problemas da comunicação efetiva de conhecimento e de seus registros à sociedade (SARACEVIC, 1996).

Gilliland (2016) destaca tipologias de padrão de dados que deveriam ser levadas em consideração na ação de tratamento nos dados, focando padronização, normalização, qualidade e intercâmbio de metadados em ambiente digital, quais sejam i) **estrutura de dados**: conjunto de elementos de metadados ou esquemas de categorias que formam um registro de informação; ii) **valores dos dados**: linguagens documentárias, vocabulários controlados, arquivos de autoridade e ontologias de domínio usados para preencher os dados nos elementos de metadados; iii) **conteúdo dos dados**: regras e códigos de catalogação que orientam em formatações, sintaxes e relacionamentos para os valores de dados usados para preencher os elementos de metadados; e iv) **comunicação de dados**: esquemas e registros de metadados expressados em uma linguagem de representação legível para a máquina.

Nessa perspectiva, diversos modelos manuais, semiautomáticos e automáticos de diagnóstico de qualidade de dados em acervos do patrimônio cultural têm sido propostos para exploração de serviços de agregação eficientes (FENLON *et al.*, 2012; BELLINI; NESI, 2013; HARPER, 2016; GAONA GARCIA *et al.*, 2017; ROMERO, 2019; WESTBROOK, 2021; MARTINS *et al.*, 2021), objetivando averiguar possíveis lacunas no tratamento documental em coleções provedoras de dados e estabelecer recomendações na utilização de regras de catalogação, incluindo formatos, sintaxes e valores adequados de preenchimento acerca dos elementos de metadados constitutivos de suas bases de dados.

Em estudo diagnóstico recente no Brasil, Martins *et al.* (2022) buscaram mapear as diferentes formas de organização e representação da informação aplicadas a objetos de coleções pertencentes a instituições filiadas à Secretaria Especial de Cultura do governo federal brasileiro, incluindo o próprio Ibram. A pesquisa constatou que a publicação de dados com qualidade na internet pelas instituições de memória investigadas ainda é muito precária em termos de organização e representação de informações (MARTINS *et al.*, 2022, p. 13), destacando o uso de padrões de documentação que refletem a qualidade de dados de seus acervos. São os casos do emprego adequado de regras de catalogação, linguagens documentárias e padrões de metadados, considerados elementos centrais na produção de bases de dados com qualidade para possíveis soluções de agregação.

Nesta pesquisa, portanto, considera-se que bases de dados são produtos de informação cruciais para a sociedade quando adotadas para realizar a mediação entre documentos e comunidade de usuários interessada na preservação, busca integrada, acesso e reuso, pois referenciam e divulgam o conhecimento por meio de sistemas de recuperação da informação (SRIs) (LANCASTER, 2004) a partir do uso qualificado da informação (HIGGINS, 2011; POOLE, 2017; MARTINS *et al.*, 2022).

Nesse sentido, para o procedimento de agregação de dados dos museus sob gestão do Ibram, foi necessário estabelecer um alinhamento dos metadados que descrevem as bases de dados das coleções envolvidas na agregação, para um modelo comum usado pela instituição (SIQUEIRA; MARTINS, 2021). Para tal, foi adotado o padrão de dados orientado internamente pelo Ibram, qual seja o modelo do Inventário Nacional de Bens Culturais Musealizados – INBCM (BRASIL, 2021).

Diante do contexto de uso do INBCM na arquitetura das bases de dados do Ibram, a situação problemática que se faz presente nesta pesquisa é que, a princípio, o INBCM surgiu para servir de instrumento de inventário para gestão interna de acervos. A normativa não foi concebida, a priori, como um modelo de catalogação que almeja requisitos descritivos únicos e singulares, vocabulários padronizados, indexação, localização, acesso e navegação em SRIs contemporâneos (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2016).

A instrução normativa não detalha e nem orienta, por exemplo, os aspectos sintáticos e semânticos para os elementos de descrição sugeridos, deixando em aberto a forma como os instrumentos de organização da informação (modelos conceituais, padrões de metadados, linguagens documentárias e regras de catalogação) deveriam ser

implementados pela instituição que aderiu aos mesmos. Adicionalmente, cabe destacar para um possível cenário de agregação de dados que, o preenchimento de campos ou elementos de metadados deveria ser seguido com rigor a partir de uma política de catalogação da instituição, de modo a atender os elementos mínimos obrigatórios orientados para um registro museal coerente e consistente.

No campo do patrimônio cultural digital destacam-se alguns padrões de documentação utilizados pelas comunidades envolvidas (LEMOS; SÁ; SOUZA, 2019; SILVA; LARA, 2021; LEMOS; COELHO JUNIOR; CARMO, 2021; TRUST, 2022). Tais padrões são direcionados à crescente digitalização de coleções de objetos culturais, das quais necessitam padronizar aspectos únicos de coleções culturais e de seus desdobramentos digitais.

Dentre os padrões de catalogação recomendados, o presente artigo destaca o guia *Cataloging Cultural Objects* (CCO) (BACA *et al.*, 2006) pela sua organização em sintetizar grupos de informação a partir de 9 capítulos, os quais expõem, de forma elucidativa, recomendações e regras de catalogação visando descrições normalizadas, padronizadas e intercambiáveis para ambientes de dados abertos ligados, conhecidos como *Linked Open Data* (LOD) (BIZER; HEATH; BERNERS-LEE, 2009). Nesse sentido, um dos destaques do CCO é que o padrão recomenda que as descrições sejam produzidas por meio de instrumentos de organização da informação adotados pela própria instituição e, em alguns casos, sugeridos pelo próprio guia, como são os casos dos esquemas de classificação do *The Getty Research Institute* (TRUST, 2022), os quais têm total aderência aos preceitos LOD.

Logo, diante à carência refletida na literatura de um modelo de referência abrangente no domínio da cultura que pudesse ser usado para o levantamento de regras consistentes ao processo de avaliação da conformidade dos dados de instituições do patrimônio cultural, visando serviços de agregação, a questão primeira de pesquisa que se elabora é a seguinte: *como o guia CCO consegue contribuir nos aspectos de padrões de dados em relação à estrutura, valor, conteúdo e comunicação para os metadados propostos do INBCM diante à realidade documental dos museus vinculados ao Ibram?* Uma possível questão de discussão secundária estaria em investigar *como a automação de uma avaliação diagnóstica conjugada com ações humanas poderia ser realizada e quais benefícios trariam a um processo de qualidade em bases de dados culturais?*

O objetivo do presente artigo é, então, apresentar o resultado de uma avaliação diagnóstica semiautomática de qualidade de dados, baseada no guia de catalogação de objetos culturais, nos acervos dos museus sob gestão do Ibram.

A presente pesquisa parte da hipótese de que a automação da avaliação da qualidade de dados pode ser um ponto de partida para a celeridade de esforços para a melhoria da qualidade de dados neste domínio. Apesar dos esforços desenvolvidos pelo Ibram na qualificação dos dados e disponibilização em ambiente digital à luz da normativa INBCM, esses dados não estão bem alinhados com recomendações indicadas por um guia de catalogação de referência no âmbito da cultura (MARTINS *et al.*, 2021). Isso pode comprometer, sobretudo, o uso e reúso desses dados em serviços de agregação para além dos museus vinculados ao Ibram, isto é, com outros possíveis *datasets* culturais disponíveis em ambientes de dados abertos ligados, como os agregadores Europeana e Digital Public Library of America (DPLA) (SIQUEIRA *et al.*, 2021), e com as plataformas da Fundação Wikimedia (WIKIMEDIA FOUNDATION, 2022), o que poderia ampliar a base de usuários e a produção colaborativa em rede de seus recursos de informação.

2 PROCEDIMENTOS METODOLÓGICOS

Metodologicamente, o presente estudo foi subsidiado por uma pesquisa aplicada, quali-quantitativa, exploratória e descritiva a partir de um estudo de caso em 22 coleções de museus digitais sob gestão do Ibram. Para a fundamentação teórica e metodológica da pesquisa, usou-se de levantamento bibliográfico nas seguintes bases de dados: Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI); Repositório Institucional da UNESP, por se tratar de importante Instituição com Programa de Pós-Graduação em Ciência da Informação no país; Biblioteca Digital Brasileira de Teses e Dissertações (BDTD); *Scientific Electronic Library Online* (SciELO); e Google Acadêmico. Como fontes consultadas, incluem-se artigos científicos, anais de congressos, teses, normas e fontes documentais subjacentes aos padrões de documentação ora estudados.

Algumas decisões metodológicas são importantes de serem esclarecidas inicialmente para fins de entendimento dos dados trabalhados na pesquisa. No que diz respeito ao INBCM, foram considerados apenas os 15 elementos de descrição para identificação do bem cultural de caráter museológico. Tal decisão foi feita após análise

prévia dos dados captados dos acervos à luz das orientações do INBCM. Apenas o Museu Solar Monjardim apresentava acervo do tipo Arquivístico. Todos os outros acervos (22 coleções no total) dos 20 museus utilizavam metadados especificados pelo INBCM com caráter museológico.

Outro detalhe importante a ser destacado é que foram considerados 8 dos 9 capítulos do CCO. O Capítulo IX, denominado *View Information*, é endereçado à catalogação do substituto digital de uma obra, a exemplo de uma imagem. Os dados de catalogação dos acervos museais vinculados ao Ibram são referentes às obras presentes nos museus, logo, não descrevem as imagens representativas dessas obras.

O método de avaliação dos dados oriundos das coleções elencadas é composto pelas seguintes duas etapas: etapa 1: alinhamento entre elementos de descrição; e etapa 2: exploração semiautomática das bases de dados de coleções, descritas nas subseções a seguir.

2.1 Alinhamento entre elementos de descrição: INBCM e CCO

A primeira etapa consistiu em realizar o alinhamento (mapeamento) entre os elementos descritivos da normativa do INBCM e do guia CCO, incluindo o grupo de informação do elemento CCO; se o elemento CCO é requerido; se o elemento CCO recomenda o uso de vocabulário controlado; além dos elementos INBCM com seus requerimentos, conforme se apresentam do Quadro 1.

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	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	Não	Sim	Local de	Não

	Location			Produção	
V-Part 2	Location	Sim	Sim	Situação	Sim
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
IX-Part 2	View Description	Sim	Não	Mídias Relacionadas	Não
NA	NA	NA	NA	Outros Números	Não
Legenda: NA (não aplicado)					

Fonte: elaborado pelos autores.

O alinhamento se deu a partir de um procedimento manual e intelectual baseado na aquisição de conhecimento sobre os dois instrumentos de pesquisa, com destaque para o aspecto de natureza semântica (papel das entidades de informação) nas decisões de cotejamento dos elementos descritivos destinados a um recurso de informação.

Considera-se importante salientar que o Capítulo VI, parte 2 do guia, dedicado ao elemento central “assunto” (*Subject*), não é considerado nos elementos de descrição para identificação do bem cultural de carácter museológico do INBCM, sugerindo que esse tipo de representação temática não é relevante para o contexto dos museus vinculados ao Ibram, ou talvez não tenha sido explorado o seu potencial pelos especialistas em documentação visando os SRIs (LANCASTER, 2004). Logo, o experimento da presente pesquisa considerou 7 dimensões analíticas (Capítulos do CCO) descritas a seguir:

- *I Object Naming*: fornece maneiras de se referir a uma obra, definindo o que está sendo catalogado.
- *II Creator Information*: identifica o criador de uma obra (um ou vários), podendo ser uma pessoa, física ou jurídica, conhecida pelo nome ou anônima.
- *III Physical Characteristics*: descreve a aparência de uma obra, apresentando características de sua forma física.
- *IV Stylistic, Cultural, and Chronological Information*: descreve características estilísticas de uma obra, origens culturais e data de *design* ou criação.
- *V Location and Geography*: trata de elementos que registram informações geográficas e de localização, tais como localização atual, locais ao longo do tempo, localização de criação e localização de descoberta.

- *VII Class*: classifica uma obra específica a outras obras com características semelhantes, muitas vezes com base em esquema organizacional de um determinado repositório ou coleção.
- *VIII Description*: associa campos específicos em todo o registro, consistindo de uma nota descritiva que geralmente é um texto relativamente breve, detalhando o conteúdo e o contexto da obra.

2.2 Exploração semiautomática das bases de dados de coleções

A segunda etapa consistiu na obtenção dos dados para a presente pesquisa. Um *script* (GITHUB, 2022) foi desenvolvido por meio da linguagem de programação Python e suas bibliotecas Pandas¹, BeautifulSoup² e Requests³ para realizar a exportação em massa de todos os dados dos acervos dos museus no formato “CSV: inbcm-ibrammapper”. Ressalta-se que esse formato é um dos disponíveis para exportação no *software* Tainacan e, portanto, segue as recomendações do INBCM. Porém, há também exportação em formatos *JavaScript Object Notation* (JSON) e *HyperText Markup Language* (HTML) disponibilizados pela *Application Programming Interface* (API) do *software* de repositório (GOV.BR, 2021).

Algumas técnicas de pré-processamento de dados foram usadas a fim de deixar a base de dados do experimento padronizada em termos de quantidade de campos, a saber: foi removido das bases o elemento de descrição “Outros Números”, conforme o procedimento de alinhamento; foram renomeados os títulos dos elementos discricionais para o padrão utilizado no CCO; e por fim, todas as bases das coleções trabalhadas foram agrupadas em uma mesma base para fins de processamento.

De acordo com o mapeamento realizado em todas as regras explicitadas nos Capítulos ora elencados do guia CCO (I, II, III, IV, V, VII e VIII), foram identificadas 244 regras, incluindo 122 regras distintas. Contudo, dentre o conjunto de regras mapeadas, foram elencadas apenas as regras pertencentes aos elementos descritivos alinhados com o INBCM (Quadro 1), que não apresentassem fator subjetivo e que, portanto, pudessem inviabilizar tecnicamente a avaliação por algoritmo computacional.

¹ <https://pandas.pydata.org>

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

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

As regras mapeadas (17 ao todo), apresentadas no Quadro 2, foram então implementadas em Python seguindo os fundamentos do campo da catalogação descritiva (GILLILAND, 2016) quanto às orientações acerca do uso de padrões para tratamento nos dados, a saber: padrão de conteúdo de dados e padrão de valor de dados nos acervos dos museus envolvidos na análise.

No caso do padrão de conteúdo de dados, a avaliação do dado foi implementada no algoritmo a partir de uma técnica conhecida como Expressões Regulares (CROCHEMORE; RYTTER, 1994, p. 157). Expressões Regulares (*regex*) são escritas em uma linguagem formal e podem ser interpretadas por um processador de expressão regular. Um processador de expressão regular é um programa que serve como um analisador sintático ou examinador de texto, identificando as partes que casam com a especificação dada, neste caso a *regex*. Várias linguagens de programação possuem formas diferentes de lidar com *regex*. No Python, há uma biblioteca chamada *re*⁴ que trabalha bem com *regex*, e esta foi utilizada no algoritmo do experimento.

Já para o padrão de valor de dados, a avaliação da utilização de vocabulário controlado foi feita a partir dos dados disponibilizados pela API do 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, e indicam se a configuração do elemento de metadado é do tipo taxonomia para uma determinada coleção.

Quadro 2 – Regras de catalogação e *regex* utilizados na pesquisa

Regra	Elemento de descrição	<i>regex</i>
Fazer uso de vocabulário controlado	<i>Class, Creation Location, Creator, Inscription, Location, Materials and Techniques, Measurements, Physical Description, Work Type, Inscription, Location</i>	Não se aplica (Utilizado API do Tainacan)
Evitar abreviações	<i>Class, Creation Location, Creator, Description, Location, Materials and Techniques, Other Descriptive Notes, Title, Work Type</i>	<code>([a-zA-Z\u00C0-\u00FF][\.\])*</code>
Usar o mesmo idioma do	<i>Creation Location, Date,</i>	Não se aplica

⁴ <https://docs.python.org/3/library/re.html>

catálogo	<i>Description, Location, Materials and Techniques, Other Descriptive Notes, Work Type</i>	(Utilizado <i>Python</i> – Langdetect)
Abreviar unidades métricas de acordo com o Sistema Internacional (m, cm, mm, g, kg, kb, Mb, Gb)	<i>Measurements</i>	(([0-9]{1,}([.\\][0-9]{1,})*) ((m M cm CM Cm g G kg Kg)+)(\\((m M cm CM Cm g G kg Kg)\\)): ([0-9]{1,}([.\\][0-9]{1,})*)
Capitalizar as iniciais de nomes próprios e da primeira palavra, para outros termos use letras minúsculas	<i>Creation Location, Date, Description, Location, Materials and Techniques, Other Descriptive Notes, Title, Work Type</i>	^[A-Z0-9]{1,}(.)*
Medidas geralmente incluem duas casas decimais para medidas métricas	<i>Measurements,</i>	([.\\]{1}[0-9]{2})+([.\\]{1}[0]
Não usar capitalização	<i>Measurements</i>	[A-Z]
Utilizar números inteiros ou frações decimais	<i>Measurements</i>	[0-9][.\\]
Não pode ficar vazio	<i>Class, Creator, Inscription, Materials and Techniques, Measurements, Work Type, Title, Date, Location</i>	(.){1,}
Usar singular	<i>Class, Materials and Techniques, Work Type</i>	(.[s] .[s]\$)
Anos com menos que 4 dígitos, inserir 0 a esquerda	<i>Date</i>	([0-9]{4})
Não usar pontuação, exceto hífen	<i>Work Type</i>	[a-zA-Z\u00C0-\u00FF 0-9\\-\\-]*
Não utilizar apóstrofo	<i>Date</i>	[']*
Não utilizar artigos	<i>Title</i>	(o ^o os ^os a ^a as ^as uma ^uma um ^um uns ^uns uma ^uma umas ^umas O ^O Os ^Os A ^A As ^As Uma ^Uma Um ^Um Uns ^Uns Uma ^Uma Umas ^Umas)
Seguir padrão para registro de hora, minutos e segundos	<i>Date</i>	(?<![0-9])(0?[0-9] 1[0-9] 2[0-3]):(60 [0-5][0-9])(?::(60 [0-5][0-9])?(?!:[0-9]+)b)
Seguir padrão pra registro de dia, mês e ano de data	<i>Date</i>	^(?:([0-9]{1,2})(V - .S)([0-9]{1,2})(V - .S)([0-9]{4}))
Use traço para separar intervalo de anos	<i>Date</i>	(([0-9]{4}) ([0-9]{4}))^([0-9]{4})\$

Fonte: elaborado pelos autores.

Para cada regra (documentada em GITHUB, 2022) associada ao elemento de metadado pertencente a uma dimensão, o registro de dado correspondente (*string* avaliada) recebeu o valor 0 ou 1. O valor 1 foi atribuído quando o registro de dado

atendeu ao critério (regra) recomendado pelo CCO; e o valor 0 quando não atendeu. Por fim, o índice de adequação é dado pela seguinte fórmula:

$$\text{índice } b = (\sum \text{valor1} / (\sum \text{valor1} + \sum \text{valor0})) * 100$$

Onde:

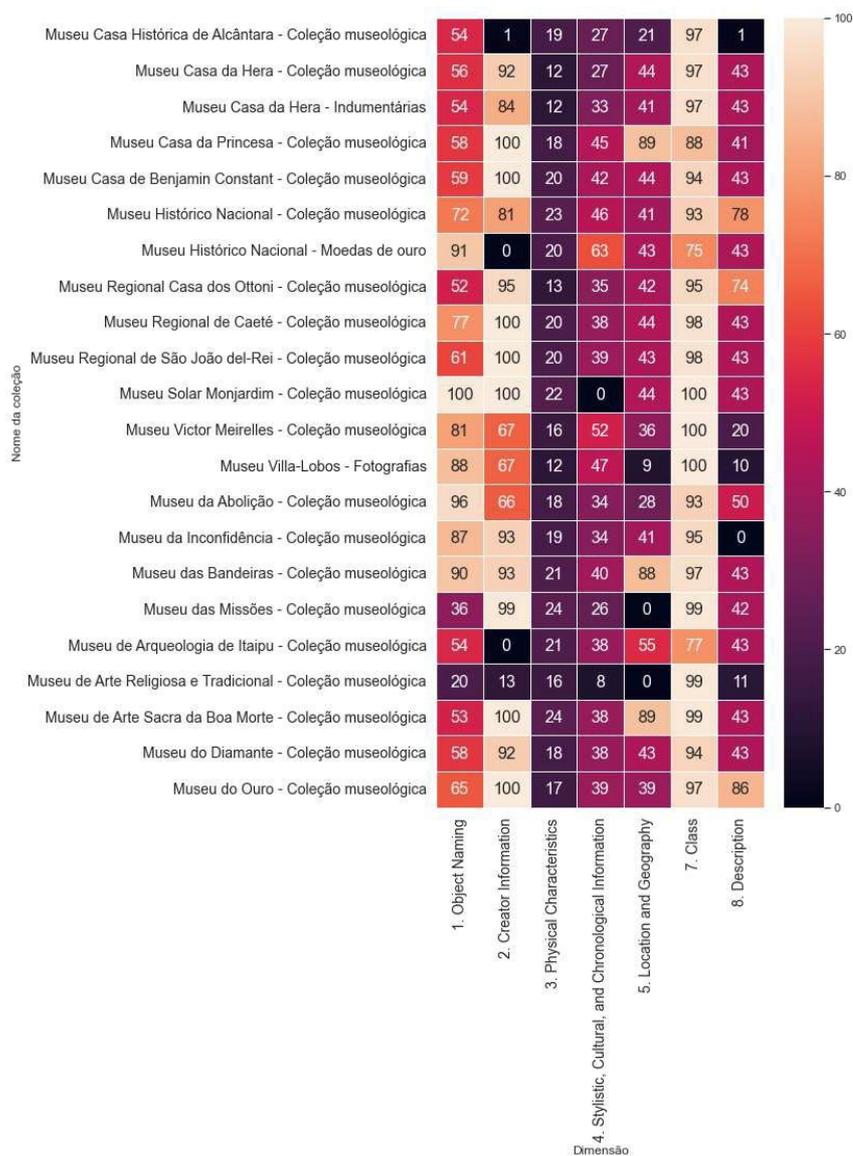
- **b** é a base com a amostra de dados de uma coleção particular;
- **índice** é o percentual de adequação obtido em relação a dimensão, a elemento de metadado e a regra de catalogação para um determinado museu e coleção;
- **valor1** é a indicação de ocorrência do registro de dado que se adequou a regra;
- **valor0** é a indicação de ocorrência do registro de dado que não atendeu a regra.

3 ANÁLISE DOS RESULTADOS

A presente seção apresenta os índices de adequação de qualidade de dados obtidos a partir das 22 coleções museológicas sob gestão do Ibram, frente às 7 dimensões analíticas oriundas dos Capítulos do CCO, os quais tiveram seus elementos descritivos alinhados com os elementos da normativa do INBCM.

Na Figura 1, cada linha do “mapa de calor” representa uma coleção e cada coluna representa uma dimensão do CCO. Os quadrados mais escuros representam (em percentuais) as dimensões com menor índice de adequação e os mais claros as dimensões com maior índice.

Figura 1 – Índice de adequação de qualidade de dados nas 22 coleções avaliadas



Fonte: elaborado pelos autores.

Na dimensão *Object Naming*, algumas coleções se destacaram com índice de adequação acima de 90%, como o Museu solar Monjardim, o Museu da Abolição, o Museu Histórico Nacional - coleção de moedas de ouro e o Museu das Bandeiras. Contribuíram para essa adequação, as regras “evite abreviações”, “use o mesmo idioma do catálogo” e “não pode ficar vazio” nos elementos *Title* e *Work Type*. Como adequação de qualidade negativa, destaca-se as coleções com índice de adequação inferior a 40%, como o Museu das Missões (36%) e o Museu de Arte Religiosa e Tradicional (20%). Esses valores são justificados devido à ausência do valor de dado no elemento *Work Type* na descrição dos recursos. Outro destaque negativo é a completa ausência de título

nas coleções do Museu de Arqueologia de Itaipu, do Museu de Arte Sacra da Boa Morte, do Museu Regional Casa dos Ottoni e do Museu Casa da Hera - Indumentárias.

Na dimensão *Creator information*, o destaque é dado às coleções do Museu Regional de Caeté, do Museu do Ouro, do Museu Casa da Princesa, do Museu Solar Monjardim, do Museu Casa de Benjamin Constant, do Museu de Arte Sacra da Boa Morte e do Museu Regional de São João del-Rei, com os índices de qualidade em 100%, atendendo completamente as regras “evite abreviações”, “não pode ficar vazio” e “faça uso de vocabulário controlado”. Como adequação de qualidade negativa, destaca-se as coleções com índice em 0% do Museu Histórico Nacional e do Museu de Arqueologia de Itaipu. Esses valores são justificados devido à ausência do valor de dado no elemento *Creator* na descrição dos recursos.

Na dimensão *Physical Characteristics*, observa-se que o maior índice de adequação chegou a 24% para as coleções Museu das Missões e Museu de Arte Sacra da Boa Morte, sendo que grande parte das coleções ficou igual ou abaixo de 20%. O baixo índice de adequação nesta dimensão ocorreu devido às regras de catalogação “faça uso de vocabulário controlado”, “medidas incluem duas casas decimais para medidas métricas”, “abrevie unidades métricas de acordo com o Sistema Internacional” e “não use capitalização” com índices inferiores a 10%.

Na dimensão *Stylistic, Cultural, and Chronological Information*, o destaque é dado às coleções que apresentaram o índice de adequação superior a 50%, sendo estas as coleções do Museu Histórico Nacional com 63% e do Museu Victor Meirelles com 52%. Na contramão, grande parte das coleções ficou abaixo de 50% por não ter atendido as regras elencadas para esta dimensão, tais como “anos com menos que 4 dígitos, insira 0 a esquerda”, “use traço para separar intervalo de anos”, “siga um padrão pra registro de dia, mês e ano de data”, “siga um padrão para registro de hora, minutos e segundos” e “não utilize apóstrofo” no elemento *Date*.

Na dimensão *Location and Geography*, algumas coleções se destacaram com índice de adequação superior a 80% por terem atendido consideravelmente as regras elencadas para esta dimensão, a saber: “capitalize as iniciais de nomes próprios e a primeira letra do texto; para outros termos, use apenas letras minúsculas”; “use o mesmo idioma”; e “evite abreviações” nos elementos *Creation Location* e *Location*. Foram os casos das coleções do Museu de Arte Sacra da Boa Morte e do Museu Casa da Princesa, ambos com 89% no índice de qualidade, e do Museu das Bandeiras com 88%. Nestas

coleções, houve perda de pontos pelo não cumprimento da regra “faça uso de vocabulário controlado” no elemento *Location*. Como adequação de qualidade negativa, destaca-se as coleções do Museu das Missões e do Museu de Arte Religiosa e Tradicional, ambas com adequação em 0%. Esse índice é justificado pela ausência do valor de dado nos elementos *Creation Location* e *Location* na descrição dos recursos.

Na dimensão *Class* pode-se observar os índices de qualidade mais altos dentre as dimensões avaliadas no experimento com índices acima de 70%, destacando 19 das 22 coleções avaliadas acima de 90% por terem atendido consideravelmente as regras “evite abreviações”, “não pode ficar vazio” e “faça uso de vocabulário controlado” no elemento *Class*.

Na dimensão *Description*, as coleções pertencentes ao Museu do Ouro (86%), Museu Histórico Nacional (78%) e Museu Regional Casa dos Ottoni (74%) obtiveram maior índice de qualidade. Tal cenário evidenciou-se devido ao cumprimento das regras “evite abreviações”, “capitalize as iniciais de nomes próprios e a primeira palavra” e “use o mesmo idioma”, perdendo pontos, por outro lado, na regra “faça uso de vocabulário controlado”. Como adequação de qualidade negativa, destaca-se a completa ausência de valores nessa dimensão para a coleção Museu da Inconfidência (0%). Por fim, as demais coleções também foram impactadas no índice de qualidade devido à ausência de valor de dado nos elementos descritivos dos recursos.

4 DISCUSSÕES

Um dos destaques do CCO é que o padrão apresenta conceitos genéricos que podem ser empregados a qualquer conjunto de metadados (BACA *et al.*, 2006), como, por exemplo, o *Machine-Readable Cataloging* (MARC), o *Metadata Object Description Schema* (MODS), o *Dublin Core*, o *VRA Core*, e, inclusive, com os elementos descritivos do INBCM, conforme se comprovou na ação de alinhamento (Quadro 1). Assim, os elementos do INBCM podem servir de base para a denominação de um conjunto de categorias que podem ser usadas para criar uma **estrutura** no formato de campos em um banco de dados ou de propriedades de um recurso em um modelo *Resource Description Framework* (RDF), por exemplo, o que reforça o aspecto do tratamento para **comunicação** de dados em ambiente digital (GILLILAND, 2016). Embora uma estrutura de dados seja o primeiro passo lógico no desenvolvimento de esquemas de metadados, uma estrutura por si só não alcançará uma alta taxa de consistência descritiva por parte

dos catalogadores, muito menos uma alta taxa de recuperação por parte dos usuários finais (BACA *et al.*, 2006), sendo necessário, portanto, outros meios de tratamento sintático e semântico nos dados.

Vale destacar que o INBCM não detalha e nem orienta a respeito dos aspectos sintáticos ou semânticos para os elementos de descrição sugeridos, deixando bastante em aberto a forma como esses elementos devem ser implementados pela instituição. Assim sendo, padrões que regem a sintaxe e a semântica da linguagem (**valor** de dados) empregada no sistema de informação e sua seleção, organização e formatação (**conteúdo** de dados) são dois outros tipos de padrões que devem ser usados em conjunto com uma estrutura de dados acordada para a aplicação. Sabe-se que trabalhos no desenvolvimento de padrões para valores de dados (LEMOS; COELHO JUNIOR; CARMO, 2021; TRUST, 2022) são muito mais evidentes do que para conteúdo de dados, normalmente na forma de tesouros, vocabulários controlados e ontologias.

Os museus vinculados ao Ibram até possuem a prática de uso de linguagens documentárias (taxonomias, no Tainacan) para preenchimento dos valores de dados para entidades associadas à classificação de temas, assuntos ou contextos de uso (MARTINS *et al.*, 2021), conforme se comprovou nos bons índices de adequação da dimensão CCO *Class* (classificação no INBCM) nas coleções avaliadas (todas acima de 70%, conforme Figura 1). Porém, o INBCM não faz menção a qualquer orientação acerca de qual versão do tesouro usar, podendo acarretar significativas diferenças terminológicas no processo de indexação em âmbito geral dos museus, o que poderia acarretar dificuldades numa solução de agregação de dados projetada. Adicionalmente, as taxonomias do Ibram não estão representadas para consumo computacional, isto é, as descrições produzidas por meio desses vocabulários controlados não estão com as suas terminologias configuradas a partir de um identificador único no formato *Uniform Resource Identifier* (URI), como se recomenda no guia CCO. Tal funcionalidade consegue estabelecer interligações e anotações sobre dados sob licença aberta (princípio dos dados abertos ligados), o que os confere possibilidades de reúso e interoperabilidade com outros conjuntos de dados na internet no âmbito do patrimônio cultural (LEMOS; COELHO JUNIOR; CARMO, 2021).

Para além do INBCM e das linguagens documentárias usadas pelos museus vinculados ao Ibram, cabe destacar a evidência de problemas de catalogação no que tange ao conteúdo dos dados frente às dimensões avaliadas pelo experimento, destacando características físicas do objeto de informação (*Physical Characteristics*),

informações cronológicas (*Chronological Information*), localização geográfica (*Location and Geography*) e descrição (*Description*).

Uma das regras que soou alerta para as coleções museológicas avaliadas em termos da não adesão à prática de catalogação foi a “não pode ficar vazio”. Esse tipo de situação pode acarretar a ausência do registro de um item numa dada situação de busca e recuperação de informação e, conseqüentemente, prejudicar a criação de possíveis índices numa solução de busca agregada (SIQUEIRA; MARTINS, 2021). Pondera-se, portanto, a importância do uso de uma norma de catalogação para o universo dos museus, destacando os que foram avaliados no presente estudo. Assim, a adoção de padrões, tal como o CCO, é de extrema importância para regulamentar como o valor de cada metadado deve ser construído, facilitando sua padronização, normalização e melhorando substancialmente as experiências de busca e recuperação da informação por parte dos usuários (LANCASTER, 2004; INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2016); além de ajudar no alcance da interoperabilidade semântica dos dados entre diferentes esquemas de metadados e aplicações em ambiente web (LEMOS; SÁ; SOUZA, 2019).

A catalogação, portanto, pelo seu longo trajeto histórico, riqueza e maturidade metodológica (WYNAR, 1985; MEY, 1995; JOUDREY; TAYLOR; MILLER, 2015; INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2016), deve ser considerada como requisito para uma política de qualidade de dados em instituições de patrimônio cultural interessadas em agregar seus dados de acervos, como é o caso do Ibram, compartilhá-los de forma descentralizada e fornecer uma melhor experiência de usuário, criando coleções completas e coerentes. Conforme assinalam Siqueira *et al.* (2021, p. 107, tradução nossa), a política de qualidade de dados “possui caráter formativo, visto que os parceiros são encorajados a não apenas enviar um mínimo de metadados e qualidade de conteúdo, mas a buscar metadados ricos e a mais alta qualidade de dados possível”. Tal ação, portanto, não se realiza sem boas práticas de catalogação.

5 CONSIDERAÇÕES FINAIS

A partir do alinhamento entre os elementos descritivos do INBCM e do CCO, pôde-se realizar a implementação de uma porção de regras de catalogação do guia CCO em *regex* por meio da linguagem Python. A aplicação possibilitou apurar o índice de

adequação da qualidade de dados em todos os registros de metadados das 22 coleções museológicas vinculadas ao Ibram, sendo mais de 17 mil itens processados. Tal resultado pode levar a uma economia de tempo para o profissional da informação na ação de avaliar a qualidade de bases de dados legadas, visando serviços de agregação, e executar possíveis ações preventivas e corretivas a partir das informações diagnósticas aferidas, respaldando, assim, a hipótese da pesquisa de que a automação dessa avaliação pode ser um ponto de partida para a celeridade de esforços para a melhoria da qualidade de dados no domínio da cultura.

A a questão primária elaborada no presente artigo se responde com as contribuições do guia CCO conjuntamente com os princípios teórico-metodológicos da Ciência da Informação em relação ao tratamento adequado de **bases de dados**. O profissional da informação envolvido no processo de modelagem de metadados geralmente utiliza padrões terminológicos para prover um vocabulário comum que descreva uma variedade de estruturas de dados capazes de satisfazer a várias comunidades, e, geralmente, são estruturados seguindo modelos para tratamento dos dados, o que redundando em normalização, qualidade e intercâmbio de suas descrições. Com o aporte do COO, incluindo seus grupos de informação a partir de 9 capítulos, os elementos descritivos do INBCM podem se tornar um esquema formal de metadados em pesquisas futuras. Adicionalmente, sistemas de organização do conhecimento contemporâneos (exs.: *Simple Knowledge Organization System – SKOS*) também são recomendados pelo CCO nessa perspectiva de modelagem, cujas terminologias (padrão de valor de dados) apresentam em suas estruturadas um URI semântico para estabelecer interligações e anotações sobre dados sob licença aberta, o que os confere possibilidades de reuso e interoperabilidade (padrão de comunicação de dados) com outros conjuntos de dados associados ao campo do patrimônio cultural. Por fim, mas não menos importante, o uso de regras de catalogação, como as previstas no CCO, determinam como elaborar o conteúdo da descrição de um recurso de informação, os pontos de acesso e os relacionamentos entre estes, tornando-se práticas essenciais na padronização, na descrição e, portanto, na agregação semântica de recursos de informação.

A questão secundária também se responde com a automação vista como uma aliada na acurácia dos resultados diagnósticos para cada coleção de museu, economizando recursos e direcionando esforços dos especialistas na tomada de decisões

com maior exigência de atenção. Como exemplo, com resultados diagnósticos de qualidade de dados projetados automaticamente, a instituição poderia investir esforços no emprego de técnicas de Ciência de Dados para melhorar a qualidade dos dados descritivos e temáticos nas bases de dados, sobretudo quando catalogados manualmente, incluindo normalização, limpeza, inclusão de valores ausentes, entre outros tratamentos, sendo bem úteis para aplicações de aprendizagem de máquina não-supervisionadas e supervisionadas (MARTINS *et al.*, 2022).

Logo, o objetivo do artigo também se consolida na apresentação da avaliação diagnóstica semiautomática de qualidade de dados nas coleções museológicas sob gestão do Ibram a partir de regras consistentes e abrangentes oriundas do CCO.

A avaliação diagnóstica semiautomática permitiu aferir que os dados das coleções carecem de um tratamento mais adequado em dimensões como características físicas do objeto de informação, descrição, localização geográfica e informações cronológicas. Por outro lado, as coleções se mostraram qualificadas em termos do uso adequado de taxonomias para a dimensão classificação. Recomenda-se, portanto, que práticas de catalogação maduras oriundas de modelos 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 mais sofisticados e orientados para usuários finais de sistemas de informação.

Por fim, como trabalhos futuros recomendam-se o refinamento das regras *regex* e a inclusão de novas regras do CCO para o modelo avaliativo. Sugere-se também a disponibilização do modelo de diagnóstico de qualidade de dados em forma de *plug-in* para ser utilizado em diferentes plataformas de repositório digital no domínio da cultura, viabilizando o cruzamento de diferentes modelos de metadados às dimensões do CCO.

REFERÊNCIAS

BACA, Murtha. *et al.* **Cataloging cultural objects: a guide to describing cultural works and their images**. Chicago: American Library Association, 2006.

BELLINI, Emanuele; NESI, Paolo. Metadata Quality Assessment Tool for Open Access Cultural Heritage Institutional Repositories. *Em*: NESI, Paolo; SANTUCCI, Raffaella (org.). **Information Technologies for Performing Arts, Media Access, and Entertainment**. Lecture Notes in Computer Science Berlin, Heidelberg: Springer Berlin Heidelberg, v. 7990p. 90–103, 2013. Disponível em: http://link.springer.com/10.1007/978-3-642-40050-6_9. Acesso em: 3 ago. 2022.

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.

CROCHEMORE, Maxime; RYTTER, Wojciech. **Text algorithms**. New York: Oxford University Press, 1994.

FENLON, Katrina; EFRON, Miles; ORGANISCIAK, Peter. Tooling the aggregator's workbench: Metadata visualization through statistical text analysis: Tooling the Aggregator's Workbench: Metadata visualization through statistical text analysis. **Proceedings of the American Society for Information Science and Technology**, [S. l.], v. 49, n. 1, p. 1–10, 2012.

GAONA GARCÍA, Paulo Alonso; FERMOSE GARCÍA, Ana; UNIVERSIDAD PONTIFICIA DE SALAMANCA; SÁNCHEZ ALONSO, Salvador; UNIVERSIDAD DE ALCALÁ. Exploring the Relevance of European Digital Resources: Preliminary Ideas on European Metadata Quality. **Revista Interamericana de Bibliotecología**, [S. l.], v. 40, n. 1, p. 59–69, 2017.

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

GITHUB. AbeilCoelho. **Qualidade_dados_IBRAM**. 2022. Disponível em: https://github.com/AbeilCoelho/Qualidade_dados_IBRAM. Acesso em: 22 jul. 2022.

GOV.BR. **Acervo em Rede e Projeto Tainacan. Ministério do Turismo - Instituto Brasileiro de Museus (Ibram)**, 2021. Disponível em: <https://www.gov.br/museus/pt-br/aceso-a-informacao/acoes-e-programas/acervo-em-rede-e-projeto-tainacan>. Acesso em: 18 jul. 2022.

HARPER, Corey A. Metadata Analytics, Visualization, and Optimization: Experiments in statistical analysis of the Digital Public Library of America (DPLA). **The Code4Lib Journal**, [S. l.], n. 33, 2016. Disponível em: https://journal.code4lib.org/articles/11752?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+c4lj+%28The+Code4Lib+Journal%29. Acesso em: 3 ago. 2022.

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

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: 22 jul. 2022.

LANCASTER, Frederic Wilfrid. **Indexação e resumos: teoria e prática**. 2. ed. Brasília: Briquet de Lemos, 2004.

LEMOS, Daniela Lucas da Silva; SÁ, Asla Medeiros E.; SOUZA, Renato Rocha. Padrões para documentação de réplicas digitais em 3D: o caso de esculturas modernistas no espaço público do Rio de Janeiro. **Tendências da Pesquisa Brasileira em Ciência da Informação**; v. 12 n. 2, 2019.

LEMOS, Daniela Lucas da Silva; COELHO JÚNIOR, Abeil; CARMO, Danielle do. Ontologias para anotação semântica em mídias: uma construção colaborativa de redes de conhecimento do patrimônio cultural. **Fronteiras da Representação do Conhecimento**, v. 1, n. 1, p. 94-125, 2021.

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**, Curitiba, v. 10, n. 2, p. 14 - 26, abr. 2021. Disponível em: <https://revistas.ufpr.br/atoz/article/view/77737>. Acesso em: 20 jul. 2022.

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

MARTINS, Dalton Lopes et al. Requisitos de qualidade para dados de agregação em museus: o caso do Instituto Brasileiro de Museus. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 14, 2021.

MEY, Eliane Serrão A. **Introdução à catalogação**. Brasília: Briquet de Lemos Livros, 1995.

BRASIL (País). Ministério da Cultura. **Resolução Normativa n. 6, de 31 de agosto de 2021**. Disponível em: <https://www.in.gov.br/web/dou/-/resolucao-normativa-ibram-n-6-de-31-de-agosto-de-2021-342359740>. Acesso em: 12 ago. 2022.

POOLE, Alex H. The conceptual ecology of digital humanities. **Journal of Documentation**, v. 73, n. 1, p. 91-122, 2017.

ROMERO, Gustavo Candela. **Publicación y enriquecimiento semántico de datos abiertos en bibliotecas digitales**. 2019. UNIVERSIDAD DE ALICANTE, Espanha, 2019. Disponível em: <https://rua.ua.es/dspace/handle/10045/97353>. Acesso em: 1 ago. 2022.

SARACEVIC, Tefko. Ciência da Informação: origem, evolução, relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.1, n.1, p.41-62, 1996.

SILVA, Camila Aparecida Da; LARA, Marilda Lopes Ginez De. Esquema básico de metadados para representação descritiva de obras de arte em museus brasileiros. **Transinformação**, [online], v. 33, 2021.

SIQUEIRA, Joyce; CARMO, Danielle do; MARTINS, Dalton Lopes; LEMOS, Daniela Lucas da Silva; MEDEIROS, Vinicius Nunes; OLIVEIRA, Luis Felipe Rosa. Elements for the construction of a data quality policy for the aggregation of digital cultural collections: the cases of the Digital Public Library of America.Inc and the Europeana Foundation. In: ÁLVAREZ, Edgar Bisset. (eds) **Data and Information in Online Environments: Second EAI International Conference- DIONE 2021**. Springer International Publishing, 2021.

SIQUEIRA, Joyce; MARTINS, Dalton Lopes. Painel de visualização analítica dos acervos digitais integrados do instituto brasileiro de museus: o uso das tecnologias Tainacan e Elastic Stack. In: **XXI Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação**, 2021, Rio de Janeiro. XXI Enancib, 2021. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/95>. Acesso em 20 jul. 2022.

SVENONIUS, Elaine. **The intellectual foundation of information organization**. Cambridge: The MIT Press, 2000.

TAYLOR, Arlene. G. **The organization of the information**. 2nd ed. Westport: Libraries Unlimited, 2004.

TRUST, Jean Paul Getty. **The Getty Research Institute - Getty Vocabularies**. 2022. Disponível em: <https://www.getty.edu/research/tools/vocabularies/>. Acesso em: 17 jul. 2022.

WESTBROOK, R. Niccole; JOHNSON, Dan; CARTER, Karen; LOCKWOOD, Angela. Metadata Clean Sweep: A Digital Library Audit Project. **D-Lib Magazine**, [S. l.], v. 18, n. 5/6, 2012. Disponível em: <http://www.dlib.org/dlib/may12/westbrook/05westbrook.html>. Acesso em: 3 ago. 2022.

WIKIMEDIA FOUNDATION. **List of Wikipedias - Meta**. Disponível em: https://meta.wikimedia.org/wiki/List_of_Wikipedias. Acesso em: 21 jul. 2022.

WYNAR, Bohdan S. **Introduction to cataloging and classification**. 7ª ed. Colorado: Libraries Unlimited Inc., 1985.

NOTAS

AGRADECIMENTOS

Agradecimentos à equipe Tainacan do Laboratório de Inteligência de Redes.

CONTRIBUIÇÃO DE AUTORIA

Concepção e elaboração do manuscrito: D. L. Lemos

Coleta de dados: A. Coelho Junior

Análise de dados: D. L. Lemos, A. Coelho Junior

Discussão dos resultados: D. L. Lemos

Revisão e aprovação: D. L. Lemos

LICENÇA DE USO

Os autores cedem à **Encontros Bibli** os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution](#) (CC BY) 4.0 International. Esta licença permite que **terceiros** remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os **autores** têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

PUBLISHER

Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no [Portal de Periódicos UFSC](#). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES

Edgar Bisset Alvarez, Ana Clara Cândido, Patrícia Neubert e Genilson Geraldo.

HISTÓRICO

Recebido em: 08-11-2022 – Aprovado em: 30-01-2023 - Publicado em: 10-02-2023.

