



XXI ENANCIB

Encontro Nacional de Pesquisa em Ciência da Informação

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

Rio de Janeiro • 25 a 29 de outubro de 2021

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

GT-2 – Organização e Representação do Conhecimento

REPRESENTAÇÃO TEMÁTICA DA INFORMAÇÃO ICONOGRÁFICA: O USO DE METADADOS EM COLEÇÕES DIGITAIS DE IMAGENS DE ARTE

THEMATIC REPRESENTATION OF ICONOGRAPHIC INFORMATION: THE USE OF METADATA IN DIGITAL COLLECTIONS OF ART IMAGE

Willian de Carvalho Silva - Universidade de São Paulo (USP)

Vânia Mara Alves Lima - Universidade de São Paulo (USP)

Modalidade: Resumo Expandido

Resumo: Este trabalho tem por objetivo analisar a representação temática da informação iconográfica a partir do uso de metadados em coleções digitais de imagens de arte. Com isso, propõem-se o levantamento dos principais padrões de metadados dedicados a descrição de imagens e objetos de artes. Tendo em vista as especificidades dos documentos iconográficos traz em perspectiva a análise documentária para imagens de arte. Destaca os principais resultados obtidos com a coleta e análise de metadados de bibliotecas e museus de arte. Conclui-se a importância dos metadados para uma representação consistente da informação iconográfica.

Palavras-Chave: representação temática; metadados; imagens de arte; coleções digitais.

Abstract: This work aims to analyze a thematic representation of iconographic information from the use of metadata in digital collections of art images. Thus, it is proposed to survey the main metadata standards dedicated to the description of images and art objects. Considering the specificities of iconographic documents, it brings into perspective the documentary analysis for art images. It highlights the main results obtained with the collection and analysis of metadata from art libraries and museums. The importance of metadata for a consistent representation of iconographic information is concluded.

Keywords: thematic representation; metadata; art images; digital collections.



XXI ENANCIB

Encontro Nacional de Pesquisa em Ciência da Informação

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

Rio de Janeiro • 25 a 29 de outubro de 2021

1 INTRODUÇÃO

Viver coletivamente garantiu ao homem a herança do conhecimento deixada por seus ancestrais. Acessar esse conhecimento foi imprescindível na luta pela sobrevivência. Com isso, o registro e a transmissão das informações se desenvolveram à medida que o ser humano aprimorou suas capacidades de representar e a classificar a realidade a qual pertence.

Ao longo da história, o aumento da produção de registros pelo homem também passou a exigir novas capacidades de representar e organizar a informação. Nesse sentido, é a partir do século XIX que novos sistemas de classificação vão classificar não apenas o documento, mas a informação contida neles, isto é, a representação da informação reorienta-se em descritiva e temática.

A representação descritiva assume o propósito de normalização dos aspectos físicos que descrevem o título, o autor, o ano de publicação, a data de publicação etc., enquanto a representação temática volta-se ao conteúdo, ao assunto, ou seja, ao Tratamento Temático da Informação (TTI).

No século XX, os processos de automação da informação ampliaram as ações sobre a organização e a representação temática da informação. Para Barreto (2007, p. 28), é justamente a partir dessa década de 1990 que a informação assumiu um novo status com a internet, pois as mudanças que a tecnologia traz são novas formas de se lidar com a informação e com o conhecimento em meio às modificações do próprio tempo e espaço de transmissão.

Com a apropriação e a ampliação das tecnologias de hipermídia, muitas instituições culturais como bibliotecas e museus de arte passaram a produzir e disponibilizar imagens de arte organizadas em coleções digitais. Portanto, é nesse cenário que os dados sobre os dados se destacam, isto é, os metadados.

Fundamentais no ambiente digital, os metadados permitem representar o assunto de uma obra de arte localizada em uma coleção digital. Em muitos casos, a ausência desses metadados evidenciam as dificuldades relacionadas a representação temática da informação de arte. Para Lima, Costa e Guimarães (2017), a Organização do Conhecimento no domínio das Artes depende da elaboração de uma informação documentária que possibilite

representar documentos iconográficos, já que tradicionalmente o uso de vocabulários controlados têm sido elaborados para serem utilizados como instrumentos de indexação e recuperação da informação de natureza verbal, colocando-se assim a necessidade de se aprimorar esse instrumento para uma recuperação mais efetiva do conhecimento nos casos de natureza iconográfica (LIMA; COSTA; GUIMARÃES, 2017, p. 288).

Diante desse contexto, como os metadados destinados à representação temática da informação iconográfica tem sido aplicado às imagens de arte em coleções digitais de bibliotecas e museus de arte? Dessa forma, o objetivo deste trabalho consiste em analisar os metadados aplicados à representação temática da informação iconográfica. Assim, propõem-se um levantamento dos principais padrões de metadados empregados na representação de recursos visuais e objetos de arte. Além disso, a coleta e análise de metadados para se verificar quantitativamente a aplicabilidade em coleções digitais de imagens de arte.

2 ANÁLISE DOCUMENTÁRIA DE IMAGENS DE ARTE

De tradição francesa, a Análise Documentária tem sua origem em meados dos anos 1960 com autores como Maurice Coyaud e o especialista em Arqueologia da Ásia Central Jean-Claude Gardin. Para Coyaud (1966, p. 18 apud LARA, 2011, p. 97), “a análise documentária é primeiramente uma operação de reconhecimento das unidades lexicais que representam as noções importantes de um documento”.

Para Gardin (1981, p. 29 apud SILVA; FUJITA, 2004, p. 136), a análise documentária poderia ser definida como “um conjunto de procedimentos efetuados com a finalidade de expressar o conteúdo de documentos científicos, sob formas destinadas a facilitar a recuperação da informação”. O conjunto desses procedimentos terá como produto a informação documentária.

A informação documentária é o produto da Análise Documentária, ou seja, a representação de um texto para integrar sistemas documentário-informacionais. Pode assumir diferentes formas, caracterizando produtos documentários distintos, desde os mais simples – listas de palavras retiradas dos textos, palavras-chave, unitermos – até os mais elaborados como resumos e índices (LARA, 2011, p. 102).

Para a análise documentária de imagens, Smit (1987, p. 106) aponta que a grande dificuldade consiste na separação entre denotação (aquilo que a imagem mostra) e conotação (aquilo que se vê ou quer ver).

Sara Shatford Layne (2002) chama atenção para o tratamento que deve ser oferecido as imagens, que diferentemente dos textos serão sempre de uma instância específica de algo, portanto, torna-se particularmente importante fornecer o acesso ao assunto da imagem. Nesse sentido, a identificação do assunto a partir da análise documentária irá produzir a informação documentária sob a forma de resumos e índices que poderão ser acolhidos por uma estrutura de metadados em um determinado sistema de informação.

4 METADADOS

Os metadados podem ser entendidos como dados codificados e estruturados que descrevem as características de entidades portadoras de informações (ZENG; QIN, 2016, p. 6). Nesse sentido, a descrição de uma entidade em um sistema de informação poderá ser representada a partir de uma sintaxe combinante entre uma estrutura composta por propriedades e valores específicos atribuídos e controlados por dado sistema de classificação.

Utilizados em instituições de patrimônio cultural, os metadados cumprem diferentes funções e podem ser categorizados em administrativos, descritivos, preservação, técnico e uso (GILLILAND, 2016). Entre essas categorias, iremos nos concentrar nos metadados descritivos, que segundo a NISO (2004) podem ser entendidos como metadados que descrevem um recurso com o propósito de descoberta ou identificação.

Atualmente, a biblioteconomia e a museologia contam com alguns padrões de estruturas de dados e padronização de conteúdos desenvolvidos para a representação de objetos culturais e recursos visuais. Entre eles se destacam: *Categories for the Description of Works of Art (CDWA)*, *VRA Core* e o *Cataloging Cultural Objects (CCO)*.

5 COLEÇÕES DIGITAIS DE IMAGENS DE ARTE

Coleção pode ser definida como uma publicação ou uma série de publicações distintas ligadas por um título em comum, mas cada uma com autor e título que lhe são próprios (CUNHA; CAVALCANTI, 2008, p. 91). Outra definição estabelece coleção como um conjunto de documentos com características comuns que foram reunidos intencionalmente (CONSELHO NACIONAL DE ARQUIVOS, 2006, p. 14).

Esta pesquisa, que consiste em uma análise exploratória e descritiva de cunho quanti-qualitativo, buscou coletar os registros de coleções digitais de bibliotecas, museus de arte e agregadores como *Digital Public Library of America (DPLA)* e *Europeana*. Para isso,

utilizou-se o *Jupiter Notebook*, *software web* em código aberto preparado para linguagens de programação como *Python*, que oferece um ambiente dinâmico para a escrita de códigos necessários à coleta dos dados.

Figura 1: Exemplo de requisição

```
https://api.europeana.eu/record/v2/search.json?theme=art&qf=TYPE%3A%22IMAG
E%22&query=drawings&reusability=open&rows=100 &cursor=* &wskey=000000000
```

Fonte: elaborado pelo autor

Na figura acima (figura 1), destaca-se um *Uniform Resource Locator* (URL) como exemplo de uma requisição para *Application Programming Interface* (API) da Europeana que pode ser utilizada em um ambiente *Jupiter Notebook*. A partir de requisições como essa, que possuem diferentes parâmetros predeterminados pelo desenvolvedor, se obteve os registros das seguintes instituições selecionadas:

Tabela 1: Registros coletados

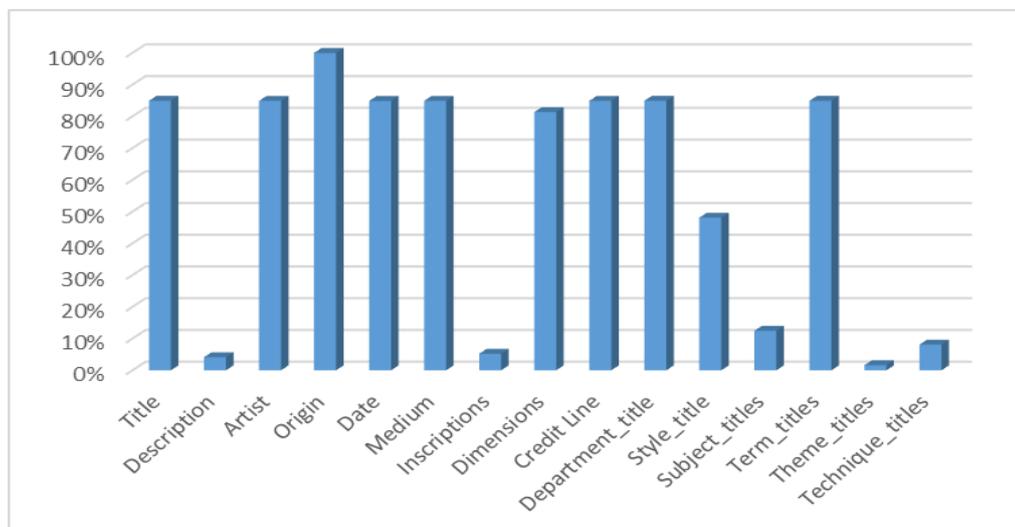
Instituição	Tipo	Registros coletados
Art Institute of Chicago	Museu	16.543
The Cleveland Museum of Art	Museu	6.474
Library of Congress	Biblioteca	18.793
New York Public Library	Biblioteca	20.532
Digital Public Library of America	Agregador	9.104
Europeana	Agregador	11.458
Total		82.904

Fonte: elaborado pelo autor

Fundamentais nas operações da *World Wide Web*, as APIs foram desenvolvidas para facilitar o acesso a serviços de diferentes *softwares* através de requisições e repostas que são geralmente fornecidas no formato de intercâmbio de dados *JavaScript Object Notation* (JSON). Com as APIs, muitas instituições de patrimônio cultural conseguem fornecer o acesso aos registros de suas coleções. Na tabela 1, identifica-se que o Art Institute of Chicago, principal museu de arte da cidade de Chicago, e detentor de mais de 300.000 obras de arte que variam em diferentes formas, estilos e artistas, teve 16.543 registros coletados. Outro importante museu de arte é o The Cleveland Museum of Art, que teve 6.474 registros

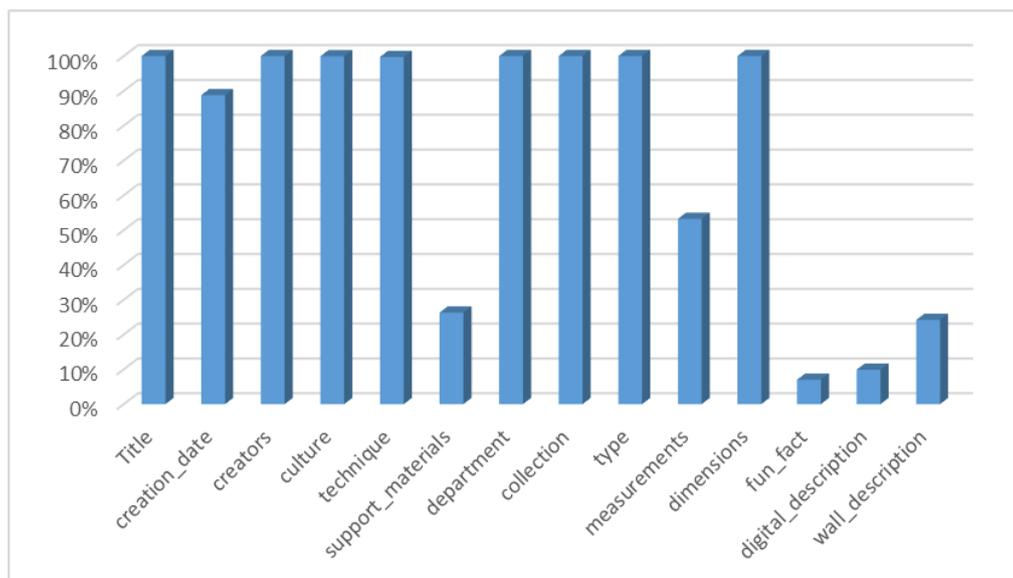
coletados. Com isso, obtivemos a projeção dos seguintes gráficos a partir da identificação dos principais metadados utilizados na descrição dos itens das coleções digitais desses museus:

Gráfico 1: Metadados Art Institute of Chicago



Fonte: elaborado pelo autor a partir dos registros coletados

Gráfico 2: Metadado The Cleveland Museum of Art

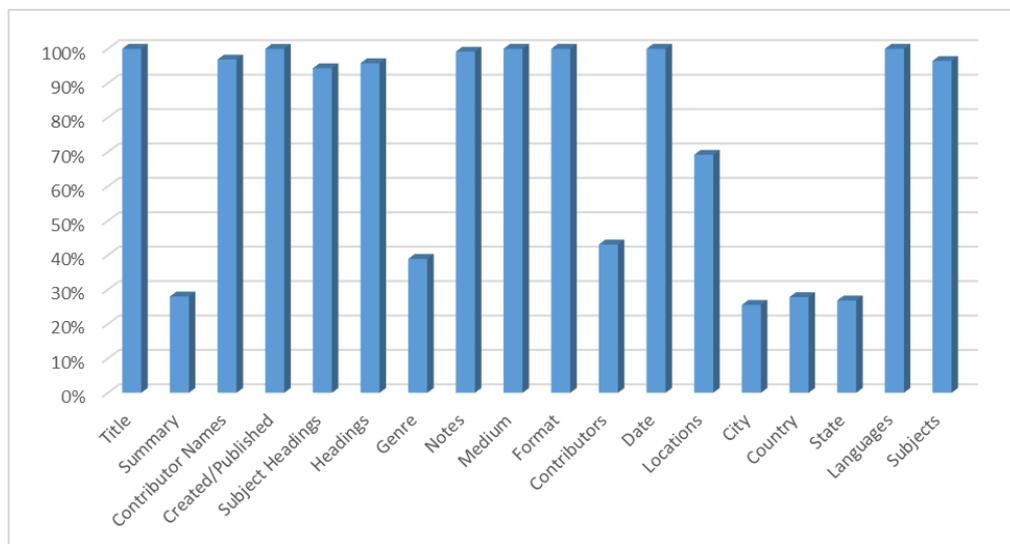


Fonte: elaborado pelo autor a partir dos registros coletados

Podemos identificar significativas diferenças entre os metadados dos gráficos 1 e 2. Essas diferenças residem nos nomes atribuídos aos elementos de cada instituição. Observe ainda que o The Cleveland Museum of Art não possui nenhum elemento denominado *Subject*. No caso do museu Art Institute of Chicago os elementos *Subject_title* e *Term_title*

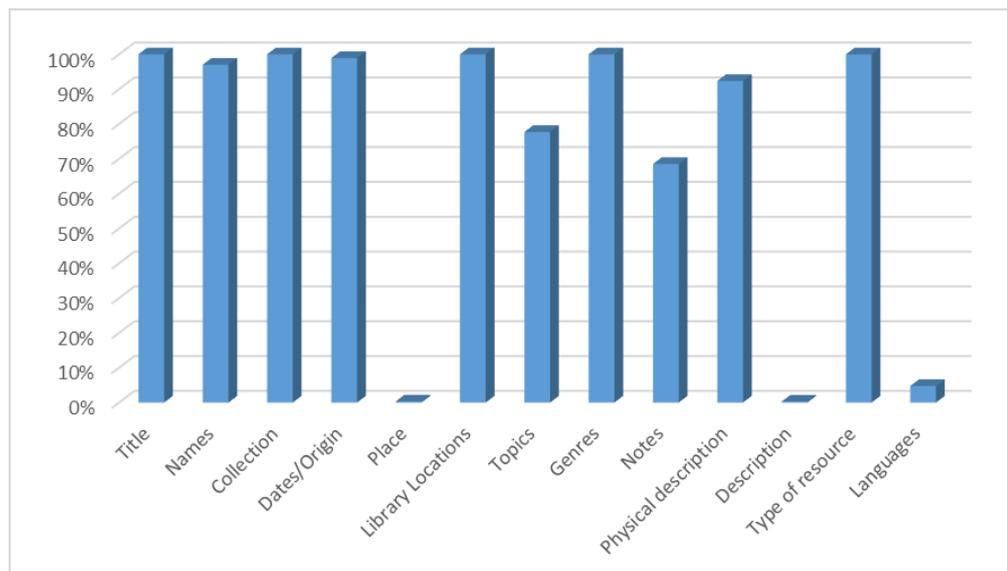
estão disponíveis apenas para os registros que são recuperados via API, isto é, esses elementos não estão visíveis no *template* para os usuários dessas coleções. Em relação a descrição das obras, existem elementos como *description* e *wall_description*.

Gráfico 3: Metadado Library of Congress



Fonte: elaborado pelo autor a partir dos registros coletados

Gráfico 4: Metadado New York Public Library

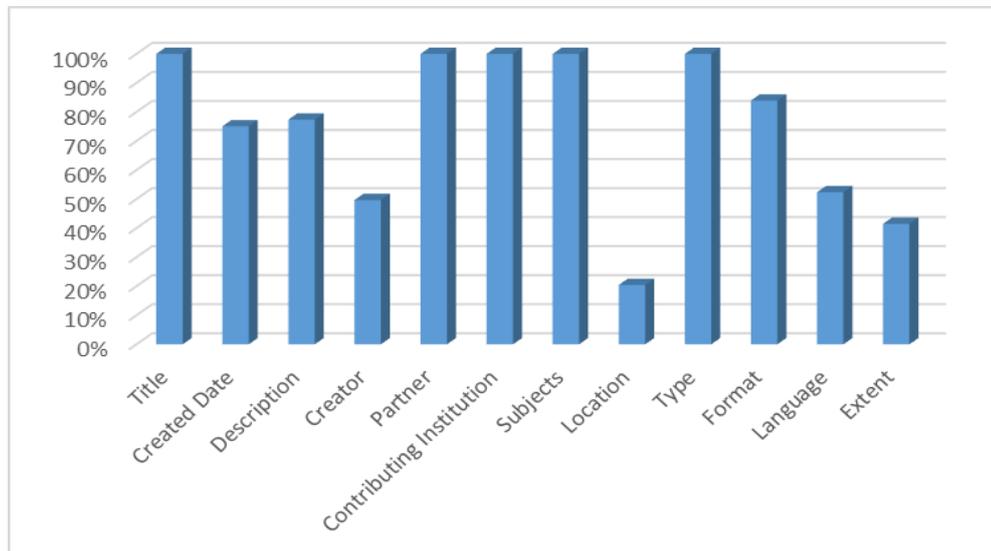


Fonte: elaborado pelo autor a partir dos registros coletados

Os gráficos 3 e 4 demonstram respectivamente os metadados coletados da Library of Congress (LC) e da New York Public Library (NYPL). Inaugurada em 1800, a LC conta com coleções de desenhos, impressões e fotografias organizados por divisões como a Prints and

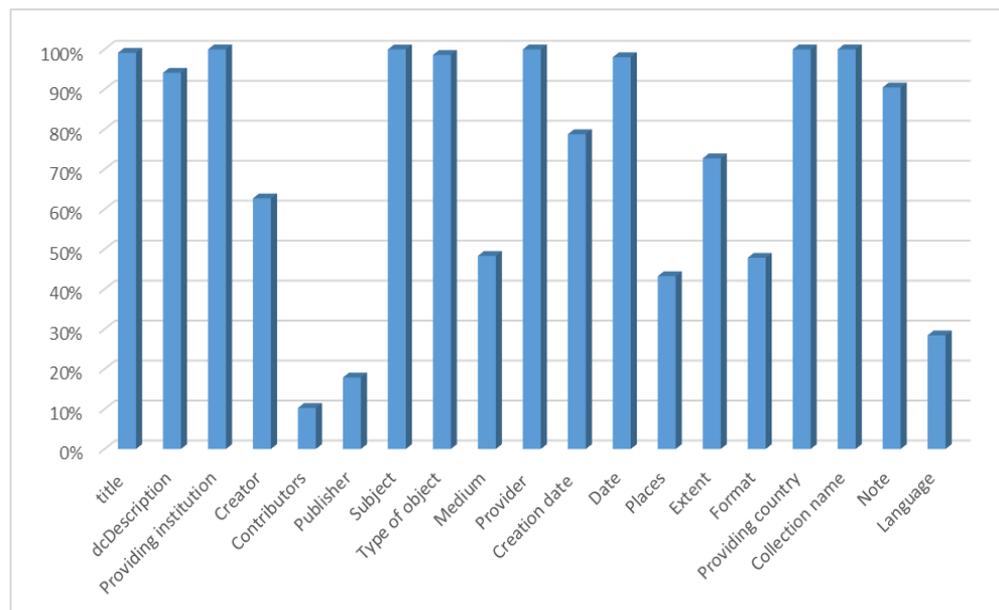
Photographs Division. Entre os elementos destinados aos pontos de acesso de assunto da LC encontramos o *Subject Headings, Headings, Genre, Locations* e *Subjects*.

Gráfico 5: Metadado Digital Public Library of America



Fonte: elaborado pelo autor a partir dos registros coletados

Gráfico 6: Metadado Europeana



Fonte: elaborado pelo autor a partir dos registros coletados

Já a NYPL dispõe de coleções digitais como Farm Security Administration Photographs que retrata a vida americana entre os anos de 1935 e 1944 a partir de fotógrafos como Dorothea Lange e Russell Lee. Para descrição dos termos associados ao assunto a NYPL destina

o elemento *Topics* (gráfico 4) que pode receber valores do tipo *geographic, name, temporal* e *occupation*. Assim como os museus (gráficos 1 e 2), a LC e a NYPL registraram baixa porcentagem quanto o elemento *description*, que no caso dos elementos da LC apresenta-se como *Summary*.

Por último, os gráficos 5 e 6 trazem os metadados ligados aos agregadores Digital Public Library of America (DPLA) e a Europeana. Lançado em 2008, a biblioteca virtual Europeana agrega itens de diferentes instituições assim como DPLA. Diante de diferentes padrões de metadados a DPLA e a Europeana adotam o Europeana Data Model (EDM) para a padronização dos dados das suas coleções digitais. Comparativamente, o elemento responsável pela descrição do assunto da DPLA e da Europeana aparecem em maior porcentagem em relação aos museus e bibliotecas apresentados nos gráficos de 1 a 4. Nesse sentido, podemos destacar que a ausência de metadados dedicados a representação temática da informação evidenciam as complexidades envolvidas na representação de documentos de natureza iconográfica.

6 CONSIDERAÇÕES FINAIS

As mudanças promovidas pelo desenvolvimento tecnológico têm exigido significativo esforço das instituições de arte quanto a disponibilização de suas coleções na *World Wide Web*. Com isso, os metadados vêm ocupando papel central nas discussões pertinentes à quebra dos silos de informação. Atualmente o desenvolvimento de novos padrões de metadados tem ganhado força em um contexto que visa a comunicação entre as coleções digitais de diferentes instituições. Contudo, uma revisão sobre como as imagens de arte estão sendo organizadas e representadas se faz necessária, já que boa parte dos métodos e ferramentas disponíveis ainda se dedicam quase que exclusivamente ao tratamento textual. Por fim, os metadados demonstraram-se fundamentais para uma consistente organização e representação da informação iconográfica.

REFERÊNCIAS

BARRETO, Aldo de Albuquerque. Uma história da Ciência da Informação. *In*: TOUTAIN, Lídia Maria Batista Brandão (org.). **Para entender a Ciência da Informação**. Salvador: EDUFBA, 2007. p. 13-34. Disponível em:

<https://repositorio.ufba.br/ri/bitstream/ufba/145/1/Para%20entender%20a%20ciencia%20da%20informacao.pdf>. Acesso em: 28 ago. 2021.

CUNHA, M.B. da; CAVALCANTI, C.R. de O. **Dicionário de Biblioteconomia e Arquivologia**. Brasília: Briquet de Lemos, 2008.

GILLILAND, Anne J. Setting the Stage. *In*: BACA, M. (ed.) **Introduction to Metadata**. 3rd ed. Los Angeles: Getty Publications, 2016. Disponível em:
<https://www.getty.edu/publications/intrometadata/>. Acesso em: 28 ago. 2021.

LARA, Marilda Lara Lopes Ginez de. Conceitos de Organização e Representação do Conhecimento na ótica das reflexões do Grupo Temma. **Informação & Informação**, n. 2, v. 16, p. 92-121, 2011.

LIMA, Vânia Mara Alves; COSTA, Ivani Di Grazia; GUIMARÃES, Magda de Oliveira. A Organização do Conhecimento no Domínio das Artes: O Fazer Terminológico na Gestão do Vocabulário Controlado. *In*: PINHO, Fábio Assis; GUIMARÃES, José Augusto (org). **Memória, Tecnologia e Cultura na Organização do Conhecimento**. Recife: Ed. UFPE, 2017 p. 288-296. Disponível em: <https://brapci.inf.br/index.php/res/v/122099>. Acesso em: 28 ago. 2021.

NISO. **Understanding Metadata**. Bethesda, MD: NISO Press, 2004. Disponível em: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf> Acesso em: 6 março 2021.

CONSELHO NACIONAL DE ARQUIVOS (Brasil). **NOBRADE: Norma Brasileira de Descrição Arquivística**. Rio de Janeiro: Arquivo Nacional, 2006. Disponível em:
<http://www.siga.arquivonacional.gov.br/images/publicacoes/nobrade.pdf>. Acesso em: 28 ago. 2021.

SHATFORD LAYNE, S. S. Subject Access to Art Images. *In*: BACA, Murtha (ed.) **Introduction to art image access: issues, tools, standards, strategies**. Los Angeles: Getty Research Institute, 2002. p. 1-19 Disponível em
<http://www.getty.edu/publications/virtuallibrary/0892366664.html>. Acesso em 27 ago. 2021.

SILVA, Maria dos Remédios da; FUJITA, Mariângela Spotti Lopes. Prática de indexação: análise da evolução de tendências teóricas e metodológicas. **Transinformação**, Campinas, v. 16, n. 2, p. 133-161, maio/ago. 2004.

SMIT, Johanna W. (org.). **Análise documentária: a análise da síntese**. Brasília: IBICT, 1987. Disponível em:
<https://livroaberto.ibict.br/bitstream/1/1011/1/An%C3%A1lise%20document%C3%A1ria.pdf> f. Acesso em: 28 ago. 2021.

ZENG, Marcia Lei, QIN, Jian. **Metadata**. 2. ed. Chicago: Neal-Schuman/American Library Association, 2016.