

EIXO 3 - UM FUTURO PARA A PROTEÇÃO DO PATRIMÔNIO

A IMPORTÂNCIA DOS METADADOS PARA A PUBLICAÇÃO DO PATRIMÔNIO CULTURAL NA WEB SEMÂNTICA

CASTRO, JULIANA M. (1); DIAS, CÉLIA C. (2)

1. Universidade Federal de Minas Gerais. Departamento de Organização e Tratamento da Informação
Av. Antônio Carlos, 6627 – Pampulha | Belo Horizonte – MG 31270-901
julianamartins@eci.ufmg.br
2. Universidade Federal de Minas Gerais. Departamento de Organização e Tratamento da Informação
Av. Antônio Carlos, 6627 – Pampulha | Belo Horizonte – MG 31270-901
celiadias@eci.ufmg.br

RESUMO

A publicação de dados sobre o patrimônio cultural na Web tem se tornado cada vez mais frequente, por meio de bibliotecas digitais, *softwares* gestores de acervo ou sistemas de gestão de conteúdo. Esse tipo de publicação, se adotados padrões e protocolos apropriados para a Web Semântica, faz com que os dados sobre o patrimônio cultural ganhem maior visibilidade, possibilitando novas descobertas e análise de grandes conjuntos de dados.

Este trabalho investiga uma parte fundamental dos padrões necessários a esse processo, representada pelos metadados. Existem diversos modelos de metadados criados por comunidades das áreas bibliográfica, arquivística e museológica disponíveis na atualidade, grande parte dos quais com acesso livre, mas que ainda não têm seu uso difundido no Brasil. Dada a demanda cada vez maior por dados abertos e interligados, torna-se de extrema importância o conhecimento das ferramentas disponíveis para a descrição e publicação de acervos on-line.

O objetivo deste estudo é mapear os diferentes tipos de metadados aplicáveis a acervos culturais, apontando suas principais características e possibilidades de uso. Trata-se de uma pesquisa qualitativa de caráter exploratório, constituída por revisão de literatura e análise documental. Como conclusão, apontam-se as potencialidades para utilização desses padrões nas instituições patrimoniais nacionais, viabilizando a gestão de acervos *online* de forma eficiente e orientada à interoperabilidade.

Palavras-chave: Patrimônio cultural; Metadados; Web Semântica; Dados Conectados; Interoperabilidade.

1 INTRODUÇÃO

A Rede Mundial de Computadores (*World Wide Web*) possibilitou uma profunda mudança no acesso a diferentes fontes de informação. Essa “teia” interliga incontáveis documentos eletrônicos que trafegam pela Internet¹ nos mais diversos formatos, como textos e hipertextos, imagens, vídeos, áudios, dentre outros. Gill (2016, p. 34) destaca que a Web “é a maior coleção de documentos que o mundo já viu, e seu crescimento não mostra sinais de desaceleração”. O autor aponta uma pesquisa do *site* britânico Netcraft que, em 2004, indicava a existência de aproximadamente 47 milhões de *hostnames* (máquinas conectadas em rede) e 22 milhões de *sites* ativos. Em 2022, 18 anos depois, o mesmo *site* indica a existência de mais de 1 bilhão de *hostnames* e 197 milhões de *sites* ativos (Netcraft, 2022).

Além do aumento exponencial no número de documentos trafegando pela Web, a própria complexificação desse ambiente gera questões para a gestão e recuperação de informações se coloquem, especialmente do ponto de vista semântico. Desenvolvida a partir da década de 2000, a visão de uma Web Semântica – hoje já uma realidade – pretendia viabilizar “uma nova forma de representar conteúdos que [permitiria] a programas, mais que simplesmente publicarem estes conteúdos, também ‘compreenderem seus significados’, podendo assim processá-los de forma mais ‘inteligente’” (Marcondes, 2021, p. 6). Nesse cenário, o uso dos metadados, ou “dados sobre dados”, se coloca como uma das respostas endereçadas a essa temática, possibilitando a operacionalização das ações de publicação e interligação dos dados dos objetos informacionais.

De acordo com Mayernik (2020, p. 3), o termo metadados “(...) teria se originado no fim dos anos 1960 no contexto do design de sistemas computacionais, se referindo ao uso de um elemento do dado para descrever ou representar algumas características do dado”. O termo foi pouco utilizado nas décadas seguintes, voltando a se tornar relevante para determinados nichos, conforme apontado pelo autor,

(...) no início da década de 1990, nas discussões sobre sistemas de gestão da informação, sistemas de informação geográfica e *design* de bases de dados. Além de um uso esporádico, o termo entrou no discurso da Ciência da Informação em meados da década de 1990, principalmente com relação ao desenvolvimento de sistemas de bibliotecas digitais e do despontar da Internet e da *World Wide Web* como grandes forças sociais (Mayernik, 2020, p. 3).

¹ A Internet trata-se, basicamente, de uma rede de computadores que compartilham dados entre si, remontando à década de 1960. Já a *World Wide Web*, criada na década de 1990, fornece ao usuário um meio de acesso ao conteúdo transferido pela Internet. (Latorre, 2018, p. 1).

Embora o termo “metadados” tenha surgido no âmbito dos sistemas computacionais, a ideia que ele representa já existe desde os primórdios da organização da informação escrita. Profissionais das áreas de bibliotecas, arquivos e museus, dentre outros, lidam com os metadados para estruturar as descrições de seus acervos, considerando aspectos como conteúdo, autoria, responsabilidade, dados de acesso, entre outros elementos que são fundamentais para a recuperação da informação na Web.

O que se faz importante destacar é que os metadados assumem uma nova importância com o avanço das tecnologias digitais on-line. A existência de um motor de buscas como o Google, criado em 1997, só é possível, entre outros fatores, pela existência de metadados que estruturam a informação, a descrevem e marcam os dados relevantes à busca. De maneira similar, objetos culturais que sejam registrados em bibliotecas digitais, bancos de dados ou *softwares* gestores de acervos precisam ser descritos de forma estruturada, conforme regras estabelecidas em normativos de cada área do conhecimento e nos próprios sistemas digitais utilizados, de forma que possam ser posteriormente recuperados com sucesso.

As instituições culturais, como bibliotecas, arquivos e museus, desempenham um importante papel no desenvolvimento de diferentes tipologias de metadados, tanto por sua crescente demanda em disponibilizar recursos digitais/digitalizados on-line, como em função da expertise dos profissionais dessas instituições. Essas instituições possuem um papel central no fornecimento de dados para pesquisas nas áreas de Ciências Sociais e Humanas e, mais recentemente, na área de Humanidades Digitais².

Essas instituições já alcançaram significativos avanços na elaboração de uma série de parâmetros para a gestão de acervos culturais no ambiente da Web, que se desdobram em iniciativas relevantes. Nesse sentido, é possível destacar a Europeana, biblioteca digital da União Europeia criada em 2008 e que, atualmente, fornece acesso a mais de 50 milhões de itens digitalizados de inúmeras instituições culturais europeias (Europeana, 2022).

Porém, Zeng (2019) aponta que, com relação ao uso de dados provenientes de bibliotecas, arquivos e museus, parte-se de um problema anterior ao estudado sobre outros tipos de

² De acordo com Guerreiro e Borbinha (2014, p. 2), as Humanidades Digitais constituem uma “(...) nova área de investigação e desenvolvimento, transversal a áreas como as humanidades e as ciências da informação e da computação. (...) O alcance das Humanidades Digitais ultrapassa largamente a mera transferência do analógico para o meio digital, centrando-se no desafio epistemológico e na articulação com os conhecimentos e os métodos utilizados nas Ciências Humanas com o mundo digital”.

dados, que já se concentram nas etapas de limpeza de dados, análise, visualização e síntese:

Na pesquisa em Humanidades Digitais, (...) o estágio mais desafiador é essencialmente o começo, na decisão de o que e como os dados podem ser coletados ao lidar com materiais históricos. (...) Para os pesquisadores que necessitam de dados históricos que não podem ser obtidos pela varredura da Web, fontes de dados e informações disponibilizadas por bibliotecas, arquivos e museus têm um valor extraordinário. (Zeng, 2019, p. 1).

Além disso, é importante ressaltar que, embora a catalogação descritiva de acervos bibliográficos, arquivísticos e museológicos seja uma prática consolidada, existem diferenças entre a publicação de acervos em catálogos *online* (*Online Public Access Catalog*, ou OPAC) que não são indexados por motores de busca externos – seja por não apresentar estrutura compatível com os motores de busca, seja por imporem restrições de acesso, como *login* e senha – e aqueles que são indexados, ou mesmo dados publicados diretamente em *sites* e outros tipos de plataformas on-line. As operações de busca e recuperação da informação internas aos *softwares* de catálogos on-line diferem daquelas exercidas pelos motores de busca externos, fato pelo qual a publicação de acervos culturais na Web demanda uma estrutura de metadados diferente daquela por vezes já contemplada nos processos de catalogação das instituições culturais, não sendo recomendada apenas uma transposição de conteúdo de um ambiente para o outro.

Dessa forma, este trabalho estabeleceu como objetivo mapear os diferentes tipos de metadados aplicáveis a acervos culturais, apontando suas principais características e possibilidades de uso no ambiente da Web. Para tanto, os procedimentos metodológicos utilizados foram a pesquisa bibliográfica e documental. A pesquisa bibliográfica, feita a partir de uma revisão narrativa de literatura, buscou inicialmente elucidar as diferenças entre os tipos de metadados. Além disso, a literatura também indicou os padrões de metadados existentes, viabilizando assim a pesquisa documental, que utilizou como fonte a documentação técnica dos padrões de metadados selecionados. Trata-se, portanto, de uma pesquisa qualitativa exploratória, que pretende reunir e apresentar um conjunto de padrões de metadados adequados à gestão e publicação de acervos culturais na Web, conforme abordado nas próximas seções.

2 TIPOS DE METADADOS

Para além da definição tradicional de “dados sobre dados”, é necessário destacar que os metadados possuem certa estruturação, conforme apontado por Smiraglia (2005, p. 4):

“metadados são dados estruturados e codificados que descrevem características de entidades portadoras de informações, para auxiliar na identificação, descoberta, avaliação e gerenciamento das entidades descritas”. Já o *Dublin Core Metadata Initiative* (DCMI), responsável pelo padrão Dublin Core, amplia, em sua definição, as funções atribuídas aos metadados, entendendo-os como “dados associados a um sistema informação ou a um objeto de informação para fins de descrição, administração, requisitos legais, funcionalidade técnica, utilização e uso e preservação” (Dublin Core, 2021). Sendo assim, é importante entender as diferentes formas em que se apresentam os metadados, e as diferentes finalidades a que eles se destinam.

De acordo com Riley (2017), em documento elaborado no âmbito da *National Information Standards Organization* (NISO), os metadados podem se apresentar de quatro formas diferentes com base em suas principais funções, conforme o Quadro 1.

Quadro 1 – Classificação funcional dos metadados.

Metadados descritivos	Para encontrar e entender um recurso.
Metadados administrativos - Metadados técnicos - Metadados de preservação - Metadados de direitos autorais	- Para decodificar e renderizar arquivos; - Gerenciamento de arquivos a longo prazo; - Direitos de propriedade intelectual associados ao conteúdo.
Metadados estruturais	Relações de partes dos recursos entre si.
Linguagens de marcação	Integram metadados e marcas para outros recursos estruturais ou semânticos dentro do conteúdo.

Fonte: Adaptado de Riley (2017, p. 6).

Os metadados descritivos, nessa abordagem, constituem-se como dados sobre o conteúdo de um recurso informacional, enquanto os metadados administrativos fornecem informações para a gestão desse recurso. Já os metadados estruturais se ocupam das relações entre as informações disponíveis sobre o recurso, enquanto as linguagens de marcação fornecem estilo para a formatação dos recursos, além de descrever a estrutura lógica, marcando os dados com valores semânticos e permitindo a sua compreensão pelas máquinas.

Uma visão diferente é apresentada por Gilliland (2016), que considera três dimensões principais dos objetos informacionais que podem ser traduzidos em termos de metadados: conteúdo, contexto e estrutura. O conteúdo se relaciona ao que é intrínseco ao objeto, a informações contidas nele, enquanto o contexto indica informações extrínsecas ao objeto, tais como quem o criou, há quanto tempo existe, onde foi criado, dentre outras. Já os metadados de estrutura se referem ao conjunto de associações que podem ser feitas entre

as informações de um objeto ou entre vários objetos, podendo ser intrínsecas a ele, extrínsecas ou ambas. (Gilliland, 2016, p. 11). Dessa análise resulta a divisão tipológica dos padrões de metadados contida no Quadro 2, a seguir.

Para Gilliland (2016), metadados descritivos, administrativos, de preservação, técnicos ou de uso são características funcionais dos metadados, ou para que eles se destinam, diferindo assim do que é colocado por Riley (2017). Contudo, mesmo utilizando lógicas diferentes de análise, as duas classificações possuem correspondências. Por exemplo, metadados de estrutura são correspondentes nos dois quadros, enquanto as linguagens de marcação (Quadro 1) correspondem aos padrões de formato e intercâmbio técnico (Quadro 2).

Quadro 2 – Classificação tipológica dos metadados.

Padrões de estrutura de dados (conjunto de elementos de metadados, esquemas)	São “categorias” ou “ <i>containers</i> ” de dados que formam um registro ou outro objeto informacional.
Padrões de valores de dados (vocabulários controlados, tesouros, listas controladas)	São termos, nomes e outros valores que são usados para povoar padrões de estrutura de dados ou conjunto de elementos de metadados.
Padrões de conteúdo de dados (regras de catalogação e códigos)	São diretrizes para o formato e sintaxe de valores usados para povoar elementos de metadados.
Padrões de formato e intercâmbio técnico (padrões de metadados expressos em formato legível por máquinas)	Esse tipo de padrão é geralmente a manifestação de um padrão de estrutura de dados particular, codificado ou marcado para o processamento por máquina.

Fonte: Adaptado de Gilliland (2016, p. 12).

Uma terceira forma de se entender os diferentes tipos de metadados é avaliar a que aspecto da informação ele se refere. Em Linguística, estudam-se os aspectos sintáticos e semânticos das construções lexicais e gramaticais. A sintaxe se ocupa das estruturas formais pelas quais algo é expresso, enquanto a semântica se dedica ao sentido do que é expresso, seja por meio de palavras ou da relação entre as palavras (Schneider, 2010). Não por acaso, esses termos foram apropriados pela Ciência da Computação, uma vez que a programação elaborada para que uma máquina processe conjuntos de dados é considerada uma linguagem, que possui suas regras assim como qualquer linguagem humana.

Os padrões de metadados podem, portanto, prover regras sintáticas (ou estruturais) para a descrição de informações – quais elementos descrever, como devem ser descritos, quais elementos devem ser marcados para a leitura das buscas – ou regras semânticas – que termos devem ser usados para povoar os elementos, que relações devem ser feitas entre os termos. Tal entendimento se faz útil para a visualização geral do panorama de metadados

necessários à descrição de objetos culturais, auxiliando o profissional a contemplar os dados essenciais nas diferentes etapas de decisões necessárias à seleção e implantação de padrões que melhor se adequem às especificidades da instituição e do acervo que abriga. Assim, na próxima seção, serão discutidas as opções de metadados disponíveis na literatura.

3 PADRÕES DE METADADOS PARA ACERVOS CULTURAIS

Como já mencionado anteriormente, as opções de metadados para a publicação de acervos culturais na Web são inúmeras. Porém, à luz dos diferentes propósitos que cada padrão apresenta, percebe-se que a análise dessas opções deve ser feita por etapas, precedida por um estudo das especificidades do acervo a ser registrado. É importante destacar que a publicação de acervos culturais na Web, de forma que possam ser recuperados com eficiência e intercambiados com diferentes instituições, envolve o uso de tecnologias para a marcação e interligação dos dados, e para isso o uso de diferentes padrões de metadados é fundamental.

3.1 Modelos conceituais

Existem documentos que propõem uma abordagem ampliada sobre o processo de descrição desses objetos e de suas relações com outros tipos de objetos. São modelos teóricos, “estruturas simplificadas que são imaginadas ou descritas, mas não literalmente construídas. (...) Podem ser usados para prever comportamentos; testar teorias; produzir tecnologias ou artefatos; demonstrar um projeto potencial e melhorar produtos, processos ou tecnologias” (Carlyle, 2011, p. 265). De acordo com Guarino, Guizzardi e Mylopoulos:

(...) modelos conceituais são artefatos produzidos com a deliberada intenção de descrever uma realidade contextualizada, (...) objetos de informação (tipicamente codificados em suportes físicos, como a memória dos computadores) que se baseiam em uma Linguagem de Modelagem para descrever qualquer aspecto de um domínio de interesse com o propósito de entendimento e comunicação (Guarino, Guizzardi, Mylopoulos, 2020, p. 4).

Os modelos conceituais disponíveis no campo cultural cumprem esse objetivo: propor estruturas conceituais que podem ser contempladas em seus vários aspectos. Nesse sentido, é altamente recomendável (embora não obrigatório) que uma instituição inicie seu processo de decisão sobre padrões de metadados consultando esses documentos e refletindo sobre a natureza, características e especificidade do acervo. Existem modelos para acervos arquivísticos, bibliográficos e museológicos.

No contexto **arquivístico**, desenvolveu-se o modelo conceitual *Records in Contexts: A Conceptual Model For Archival Description* (RIC-CM), pelo *International Council on Archives* (ICA) a partir da década de 2010, tendo sua primeira versão publicada em 2016, mas ainda em processo de expansão. A versão do modelo atual inclui as entidades descritivas centrais, suas propriedades ou atributos e relações essenciais entre elas – trata-se, portanto, de um modelo Entidade-Relação (ER)³ –, sendo todas incorporadas com base nos quatro padrões internacionais de descrição arquivística: *General International Standard Archival Description* (ISAD(G)), *International Standard Archival Authority Records - Corporate Bodies, Persons, and Families* (ISAAR(CPF)), *International Standard for Describing Functions* (ISDF) e *International Standard for Describing Institutions with Archival Holdings* (ISDIAH). O modelo RIC-CM assume o formato de uma “descrição multidimensional”, ao invés de uma descrição hierárquica, podendo ser apresentadas em forma de gráficos e redes. Há também a previsão de uma ontologia (RIC-O) como desdobramento desse modelo (International Council on Archives, 2016).

Com relação a acervos **bibliográficos**, o *Functional Requirements for Bibliographic Records* (FRBR) foi desenvolvido ainda na década de 1990 no âmbito da *International Federation of Library Associations and Institutions* (IFLA) e, assim como o RIC-CM, trata-se de um modelo ER, que contempla as entidades, atributos e relacionamentos para os registros bibliográficos (Tillet, 2003). A “família FR” ainda era composta por outros dois desdobramentos, o *Functional Requirements for Authority Data* (FRAD), para registros de autoridade de nomes, e o *Functional Requirements for Subject Authority Data* (FRSAD), para registros de autoridade de assuntos.

Com o objetivo de atualizar e unificar os três modelos desenvolvidos no âmbito da IFLA, foram realizados vários encontros de grupos de revisão, a partir de 2010, o que culminou com a edição do modelo *Library Reference Model* (LRM), entre 2016 e 2017, que

(...) incorporou a abordagem adotada no estudo original do FRBR, (...) [mas] não faz distinção entre dados tradicionalmente armazenados em registros bibliográficos ou de coleções e dados tradicionalmente armazenados em registros de autoridade de nome ou assunto. Para os propósitos do modelo, todos esses dados estão incluídos no termo informação bibliográfica e, como tal, estão dentro do escopo do modelo. (Riva, Le Bœuf, Žumer, 2017, p. 5).

³ “A modelagem ER é uma técnica que especifica a estrutura de um modelo conceitual. Em outras palavras, ele especifica o tipo de coisas que devem estar nele e as propriedades que essas coisas podem ter. Uma explicação simplificada dessa estrutura é que três tipos de coisas são permitidos: entidades, atributos e relacionamentos. Entidades são coisas, físicas ou abstratas; (...) relacionamentos são interações entre entidades; e atributos são propriedades ou características de entidades ou relacionamentos” (Carlyle, 2011, p. 266).

Já o BIBFRAME (*Bibliographic Framework*) tem por finalidade implementar um novo ambiente para as bibliotecas na Web, favorecendo a interconexão de informações e o intercâmbio de dados, e para isso contando com classes, propriedades e relacionamentos expressos em RDF⁴. Trata-se de um projeto que teve início em 2012 no âmbito da *Library of Congress* como um modelo de dados interligados, a fim de tornar as informações bibliográficas mais úteis dentro e fora da comunidade de bibliotecas (The Library of Congress, 2021a).

No âmbito dos acervos **museológicos**, o *CIDOC Conceptual Reference Model* (CRM) foi desenvolvido pelo *International Council of Museums* (ICOM) a partir de 2000, sendo um modelo teórico e prático para a integração de informações na área do patrimônio cultural. Ele fornece definições e estrutura formal para descrever os conceitos e relações implícitas e explícitas usados na documentação do patrimônio cultural, podendo ser descrito como uma ontologia formal⁵. Essas descrições formais permitem a integração de dados de fontes múltiplas em *softwares* e esquemas diversos (CIDOC CRM, 2022).

3.2 Esquemas de metadados

Após a identificação dos modelos conceituais que abrangem o acervo que se pretende descrever, faz-se necessário analisar a estrutura dos metadados que serão utilizados. Essa estrutura deverá contar, minimamente, com um esquema de metadados e uma linguagem de marcação (no caso de publicação de conteúdos on-line) e, idealmente, também um padrão para a descrição de conteúdo.

Os esquemas de metadados são os conjuntos de elementos necessários para descrever um determinado objeto informacional. No campo cultural existem diversos esquemas disponíveis, aplicáveis a diferentes tipos de objetos informacionais, oriundos de acervos distintos. Dentre essa diversidade, destacam-se o padrão EAD, para acervos arquivísticos; os padrões CDWA e VRA Core, para acervos artísticos, arquitetônicos e museológicos; os

⁴ O RDF (*Resource Description Framework*) trata-se de um modelo de dados que constitui atualmente uma das bases da Web Semântica, já que viabiliza a formatação de dados de forma a gerar significado para as máquinas. O RDF descreve uma forma de representação da informação muito parecida com as sentenças em linguagem natural, fazendo uso de sujeitos, predicados e objetos, estrutura conhecida como “triplas” (W3C, 2022).

⁵ De acordo com Campos (2010, p. 223), uma ontologia formal é “aquela que define vocabulário com lógica. A sintaxe exata e a semântica dependem da representação da linguagem, que é expressa, sintaticamente, com uma linguagem equivalente à lógica de primeira ordem. Desta forma, uma ontologia consiste em termos, definições, e axiomas relativos a ele. (...) As ontologias formais, ligadas ao conceito de modelagem de conhecimento, podem ser consideradas um mecanismo de representação, como um meio próprio de observação do conhecimento de um dado domínio”.

padrões MARC, METS e MODS, para acervos bibliográficos; o padrão TEI, para acervos textuais; e, para recursos digitais em geral, o Dublin Core. O Quadro 3 apresenta as principais características de cada um desses padrões.

Quadro 3 – Esquemas de metadados para acervos culturais.

Tipo de acervo	Esquema de metadados	Caracterização
Arquivístico	EAD	O <i>Encoded Archival Description</i> (EAD) é um padrão internacional de metadados arquivísticos. Desenvolvido pelo EAD <i>Working Group</i> da <i>Society of American Archivists</i> , em parceria com a <i>Library of Congress</i> , foi publicado pela primeira vez em 1998 e encontra-se na versão EAD3 1.1.1, publicada em 2019. Ele é expresso em XML e reflete a estrutura hierárquica das coleções arquivísticas, permitindo a descrição tanto da coleção como um todo quanto de suas partes individuais. Possui 165 elementos descritivos e 85 atributos. Aceita descrições multilíngues e já descreve a transposição para seus elementos da norma ISAD(G) e dos padrões MARC21 e MODS (EAD, 2019, p. 5-6).
Bibliográfico	MARC	O <i>Machine Readable Cataloging</i> (MARC) é um padrão para a representação e comunicação de informações bibliográficas e relacionadas em formato legível por máquina. Um registro MARC envolve três elementos: a estrutura do registro, a designação de conteúdo (<i>tags</i> numéricas, indicadores e códigos de subcampo) e o conteúdo em si. O padrão foi criado no âmbito da <i>Library of Congress</i> ainda na década de 1960 e possui variações que mudam de acordo com o país que o adota, sendo o mais utilizado atualmente o MARC21, uniformização dos padrões norte-americano e canadense (THE LIBRARY OF CONGRESS, 2021b).
	METS	O <i>Metadata Encoding and Transmission Standard</i> (METS) é uma especificação de codificação e transmissão de dados, expressa em XML, que fornece os meios para transmitir os metadados necessários para o gerenciamento de objetos digitais dentro de um repositório (como bibliotecas digitais) e a troca de tais objetos entre repositórios. O esquema foi criado em 2001 sob o patrocínio da <i>Digital Library Federation</i> (DLF) e apoio da <i>Library of Congress</i> . Em 2004 recebeu o Registro NISO. Possui 36 elementos e 57 atributos (THE LIBRARY OF CONGRESS, 2021c).
	MODS	O <i>Metadata Object Description Schema</i> (MODS) é um esquema XML para um conjunto de elementos bibliográficos que pode ser usado para uma variedade de propósitos e, particularmente, para aplicativos de biblioteca. É um derivado do formato bibliográfico MARC21 e, como tal, inclui um subconjunto de campos MARC, usando <i>tags</i> baseadas na linguagem em vez de numéricas. Possui 108 elementos e subelementos, suportando descrições multilíngues (THE LIBRARY OF CONGRESS, 2021d).
Museológico	CDWA	O <i>Categories for the Description of Works of Art</i> (CDWA) trata-se de uma estrutura conceitual para a descrição de obras de arte, arquitetura, cultura material, grupos/coleções de obras e imagens relacionadas. O CDWA não é um modelo de dados, mas pode ser usado para projetar modelos de dados, bancos de dados e para acessar informações. Inclui 540 categorias e subcategorias de informações, mas existem 13 categorias consideradas o núcleo do esquema, que devem ser preenchidas. Sua versão resumida, denominada CDWA Lite, é expressa em XML. Aceita descrições multilíngues e já descreve a transposição para seus elementos dos padrões VRA Core, MARC, MODS, Dublin Core e EAD (BACA, HARPRING, 2016).

Tipo de acervo	Esquema de metadados	Caracterização
Museológico	VRA Core	O VRA Core é um padrão de dados para a descrição de obras de cultura visual, bem como das imagens que as documentam. É reconhecido internacionalmente e usado como um formato autônomo ou um esquema de extensão aprovado para o padrão METS. O VRA Core foi desenvolvido pela <i>Visual Resources Association</i> em 1996 e atualmente encontra-se na versão 4.0, lançada em 2007, expressa como um esquema XML. O padrão possui 3 entidades primárias, 19 elementos e 9 atributos globais. Aceita descrições multilíngues e, em sua versão anterior (3.0), descrevia a transposição de seus elementos para o padrão MARC (THE LIBRARY OF CONGRESS, 2021e).
Textual	TEI	O <i>Text Encoding Initiative</i> (TEI) é um esquema XML modular e extensível para representar recursos estruturais, de renderização e conceituais dos textos. Se concentram (embora não exclusivamente) na codificação de documentos das ciências humanas e sociais e, em particular, na representação de materiais de fonte primária para pesquisa e análise. Desenvolvido nas décadas de 1980 e 1990 pela <i>Association for Computers and the Humanities</i> e mantido pelo <i>TEI Consortium</i> , encontra-se na versão 4.2.1, ou TEI P5, atualizada em 2021 (TEI CONSORTIUM, 2021).
Geral	Dublin Core	O Dublin Core original de 13 elementos foi publicado pela primeira vez em 1995. Em 1998 foi formalizado no padrão RFC-5791 e começaram as discussões para torná-lo um padrão NISO. Atualmente, o <i>Dublin Core Metadata Element Set</i> conta com 15 elementos e dezenas de propriedades, classes, tipos de dados e esquemas de codificação de vocabulário. Os termos são expressos em vocabulários RDF para uso em projetos de dados abertos conectados. Os criadores de metadados não-RDF podem usar os termos em contextos como XML, JSON, UML ou bancos de dados relacionais, desconsiderando o identificador global e as implicações formais dos aspectos específicos de RDF das definições de termos. Esses usuários podem considerar as relações de domínio, intervalo, subpropriedade e subclasse como sugestões de uso e se concentrar no texto de linguagem natural de definições, notas de uso e exemplos (DUBLIN CORE, 2021).

Fonte: elaboração própria, 2022.

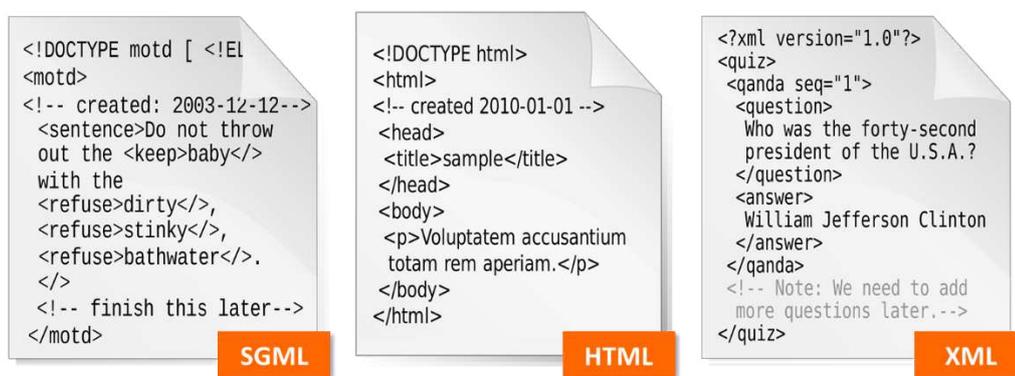
3.3 Linguagens de marcação

Os esquemas de metadados geralmente são associados a uma linguagem de marcação, que atua tanto no formato com que os dados serão apresentados, demarcando as diferenças entre título, parágrafos e outros segmentos textuais, como na sinalização dos termos relevantes no esquema de elementos, que serão varridos em um processo de busca. É importante destacar que esses tipos de marcação são fundamentais para a recuperação de conteúdos publicados na Web, uma vez que é por meio deles que as máquinas conseguem compreender diferenças de estilo e identificar as *tags* (palavras-chave) que representam o recurso publicado.

As linguagens de marcação surgiram ainda na década de 1960, com a criação da *Generalized Markup Language* (GML), ou linguagem de marcação generalizada, que foi, nas décadas de 1970 e 1980, definida como linguagem de marcação padrão para anotações em

documentos digitais, passando a se chamar *Standard Generalized Markup Language* (SGML). A SGML é considerada uma metalinguagem, a partir das quais existe a possibilidade de se desdobrarem outras linguagens mais específicas. Isso ocorreu com a criação, ainda na década de 1980, da *Hypertext Markup Language* (HTML), ou linguagem de marcação hipertextual – que possibilitou a conexão entre diferentes documentos via links, transformando o texto em um hipertexto – e da *eXtensible Markup Language* (XML), ou linguagem de marcação extensível – criada na década de 1990 como um subconjunto simplificado da SGML (Ramalho, Martins, Sousa, 2017, p. 23-24). As linguagens mencionadas encontram-se representadas na Figura 1, a seguir.

Figura 1 – Exemplos da estrutura de escrita das linguagens SGML, HTML e XML.



Fonte: Dreftymac, 2007a, 2007b e 2007c.

Muitos dos esquemas de metadados citados anteriormente fazem uso da linguagem XML (como o METS, o MODS, o CDWA, o VRA Core), enquanto outros optam por partir diretamente da SGML (como o TEI e o EAD). Alguns modelos conceituais e esquemas fazem uso também do modelo de dados RDF, anteriormente citado, e de metalinguagens como o JSON (*JavaScript Object Notation*).

3.4 Metadados de conteúdo

Os metadados de conteúdo descrevem as regras a serem seguidas para o preenchimento dos elementos do esquema de metadados escolhido. No domínio bibliográfico podem ser citados como exemplos o *Anglo-American Cataloguing Rules* (AACR), o *Resource Description and Access* (RDA) e o *International Standard Bibliographic Description* (ISBD), enquanto o *Describing Archives: A Content Standard* (DACS) se encontra no domínio arquivístico e o *Cataloging Cultural Objects* (CCO) se destina a objetos culturais em geral.

A maioria desses padrões já define previamente de qual esquema de metadados irá derivar seus elementos, como o CCO, por exemplo, que faz uso de elementos do VRA Core e do CDWA. Mas é possível, no caso de haver a intenção de utilizar um esquema diferente, realizar a correspondência entre os elementos.

Não há obrigatoriedade no uso de um padrão de conteúdo para a descrição dos acervos, porém alguns domínios, como a biblioteconomia, tradicionalmente já fazem uso de padrões, que tornam a descrição de recursos de informação mais consistente, possibilitando o compartilhamento destes dados com outras instituições. Além destes aspectos, destaca-se ainda que esses padrões são, em sua maioria, elaborados com vistas às tecnologias de dados abertos conectados (*Linked Open Data* – LOD), já alinhados aos objetivos da Web Semântica. Com isso, há maior possibilidade de que os dados publicados sejam mais facilmente recuperados e remetam a conteúdos externos.

3.5 Metadados semânticos

Passando à “esfera semântica”, os metadados concentram-se principalmente no conteúdo dos elementos de dados. Metadados descritivos (ou metadados de valor), por exemplo, são basicamente os valores dos campos a serem preenchidos, o que pode ser feito de forma livre ou com o uso de vocabulários controlados. São múltiplos os vocabulários controlados disponíveis para acervos culturais, tais como o *Art & Architecture Thesaurus* (AAT); o *Thesaurus of Geographic Names* (TGN) e o *Cultural Objects Name Authority* (CONA), todos do *Getty Institute*. No Brasil, destacam-se o Tesouro de Objetos do Patrimônio Cultural nos Museus Brasileiros e a Base de Terminologia da Fundação Biblioteca Nacional, dentre vários outros.

Metadados administrativos são destinados à gestão de acervos, enquanto metadados de preservação controlam as condições físicas e digitais de preservação dos objetos informacionais. Já os metadados técnicos referem-se à documentação sobre *hardwares* e *softwares* utilizados, enquanto os metadados de uso referem-se aos dados sobre buscas, acessos e empréstimos registrados pelo sistema. Esses quatro tipos de metadados são mais específicos, dependendo de situações próprias às instituições que abrigam acervos e aos sistemas utilizados para seu registro e gestão, sendo ideal que decisões locais sejam tomadas – e registradas – para sua implementação.

Consolidando todas essas informações, apresentam-se, no Quadro 4, as opções disponíveis em cada uma das categorias de metadados, divididas por tipo de acervo contemplado. As

categorias foram numeradas de 1 a 5, na ordem sugerida por este trabalho para a análise dos padrões, partindo do mais amplo (modelos conceituais) aos mais restritos (metadados descritivos)⁶.

Quadro 4 – Padrões de metadados disponíveis para cada tipo de acervo cultural.

Tipo de acervo	1. Modelos conceituais	2. Esquemas de metadados	3. Linguagens de marcação	4. Metadados de conteúdo	5. Metadados descritivos
Arquivístico	RIC-CM	EAD, TEI (textual), Dublin Core (recursos digitais)	SGML, HTML, XML, XHTML	DACS	Dicionário Brasileiro de Terminologia Arquivística
Bibliográfico	FRBR/LRM, BIBFRAME	MARC, METS, MODS, TEI (textual), Dublin Core (recursos digitais)	SGML, HTML, XML, XHTML	AACR, RDA, ISBD	Base de Terminologia da Fundação Biblioteca Nacional
Museológico	CIDOC CRM	CDWA, VRA Core, TEI (textual), Dublin Core (recursos digitais)	SGML, HTML, XML, XHTML	CCO	Tesouro de Objetos do Patrimônio Cultural nos Museus Brasileiros

Fonte: Elaboração própria, 2021.

Os profissionais envolvidos em projetos de publicação de acervos culturais na Web podem fazer uso desse quadro para iniciar a análise dos instrumentos necessários para tanto, levando em conta a realidade dos acervos e das instituições que os abrigam. Cada uma das colunas representa um momento diferente de tomada de decisão, como, por exemplo: 1. Será adotado um modelo conceitual para a descrição do acervo? Se sim, qual? 2. Quais são os esquemas de metadados mais adequados à descrição do acervo? Eles se adequam ao modelo conceitual escolhido? 3. Qual linguagem de marcação pode ser usada para expressar o esquema de metadados escolhido? 4. Serão utilizados metadados de conteúdo para descrever as regras de catalogação? Quais? 5. Serão utilizados metadados descritivos (ou de valor) para controlar os termos utilizados? Quais?

Percebe-se, portanto, que o uso de metadados implica uma série de etapas, que envolvem a determinação dos padrões mais adequados aos tipos de acervos e a implementação desses padrões para a descrição e publicação dos recursos. Embora sejam várias as

⁶ Os metadados descritivos, que representam os vocabulários controlados, foram exemplificados no Quadro 4 apenas com as publicações nacionais, uma vez que são inúmeros os padrões para cada tipo de acervo no contexto internacional. Ressalta-se também que não foram adicionados a esse quadro os metadados administrativos, técnicos, de preservação e uso, por serem muito dependentes do contexto institucional e de sistema empregado. É importante destacar também que os metadados do campo museológico também podem ser utilizados para acervos arquetônicos e patrimoniais.

possibilidades – e conseqüentemente a demanda de análise de cada uma delas –, é importante ressaltar que a publicação de acervos culturais na Web é uma das formas mais eficientes de se garantir a democratização do acesso ao patrimônio cultural, aumentando sua visibilidade e possibilitando o uso e reuso dessas informações.

4 CONSIDERAÇÕES FINAIS

Este trabalho buscou apresentar, de forma ampliada e introdutória, questões fundamentais para o embasamento de iniciativas de estruturação, descrição, gestão e publicação de acervos culturais na Web, de forma aberta e interligada. O levantamento dos padrões de metadados disponíveis para acervos arquivísticos, bibliográficos e museológicos foi aqui sintetizado de forma a oferecer aos profissionais da informação um leque de opções a serem consideradas em seus projetos. A consolidação desses padrões colabora para que sejam feitas escolhas embasadas em evidências, evitando retrabalhos e a perda de tempo e de informações sobre os recursos.

É importante apontar que não houve a pretensão de exaurir o tema, e assim este estudo possui algumas limitações. Não foram tratadas as recomendações e os protocolos técnicos, como o protocolo Z39.50, que permite a pesquisa e a recuperação de informações em redes de computadores, e a recomendação SKOS (*Simple Knowledge Organization System*), para representação de vocabulários controlados, uma vez que estão implícitos nos padrões aqui abordados, apesar da importância para seu entendimento. Dois modelos de dados que também não foram aqui contemplados são o OAI-ORE (*Open Archives Initiative Object Reuse and Exchange*), que define padrões para a descrição e troca de agregações de recursos da Web, e o EDM (*Europeana Data Model*), modelo de dados da biblioteca digital Europeia que possibilita que os dados de outras instituições sejam “agregados” a seu esquema. A questão da agregação de metadados, fazendo a transposição de um esquema para outro, demanda um aprofundamento que não seria possível no escopo deste artigo, sendo uma temática para estudos futuros.

Percebe-se, dessa forma, que existem inúmeras possibilidades para a adoção de uma série de padrões de metadados coordenados pelas instituições culturais, com vistas à estruturação, descrição, gestão, preservação e publicação de seus acervos culturais. O conhecimento dos metadados – suas características, tipologias e para que se destinam – e das tecnologias por trás desses padrões são o primeiro passo para que os objetos culturais

sejam inseridos significativamente na Web Semântica, aumentando seu alcance e cumprindo sua vocação cultural, educativa, informativa e fruitiva.

REFERÊNCIAS

BACA, M., HARPRING, P. (eds.). *CDWA: Categories for the Description of Works of Art*. Los Angeles: Getty Institute, 2016.

CAMPOS, M. L. A. O papel das definições na pesquisa em ontologia. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 15, n. 1, p. 220-238, 2010.

CARLYLE, A. Understanding FRBR as a conceptual model. *Library Resources & Technical Services*, v. 50, n. 4, p. 264-273, 2011.

CIDOC CRM. *What is the CIDOC CRM?* Disponível em: <<http://www.cidoc-crm.org/>>. Acesso em: 19 mar. 2022.

DREFTYMAC. *A graphical depiction of a very simple HTML document*. In: WIKIMEDIA COMMONS, a midiateca livre. Flórida: Wikimedia Foundation, 2007a. Disponível em: <<https://commons.wikimedia.org/wiki/File:HTML.svg>>. Acesso em: 19 mar. 2022.

DREFTYMAC. *A graphical depiction of a very simple SGML document*. In: WIKIMEDIA COMMONS, a midiateca livre. Flórida: Wikimedia Foundation, 2007b. Disponível em: <<https://commons.wikimedia.org/wiki/File:SGML.svg>>. Acesso em: 19 mar. 2022.

DREFTYMAC. *A graphical depiction of a very simple XML document*. In: WIKIMEDIA COMMONS, a midiateca livre. Flórida: Wikimedia Foundation, 2007c. Disponível em: <<https://commons.wikimedia.org/wiki/File:XML.svg>>. Acesso em: 19 mar. 2022.

DUBLIN CORE. *Dublin Core Metadata Initiative*. Disponível em: <<https://dublincore.org/specifications/dublin-core/>>. Acesso em: 02 dez. 2021.

EAD. *Encoded Archival Description: Tag Library*. v. EAD3 1.1.1. Chicago: Society of American Archivists, 2019. 422p.

EUROPEANA. *About Europeana*. Disponível em: <<https://www.europeana.eu/pt/about-us>>. Acesso em: 19 mar. 2022.

GILL, T. Metadata and the Web. In: BACA, M. (ed.). *Introduction to metadata*. 3 ed. Los Angeles: Getty Publications, 2016. p. 32-50.

GILLILAND, A. J. Setting the stage. In: BACA, M. (ed.). *Introduction to metadata*. 3 ed. Los Angeles: Getty Publications, 2016. p. 10-31.

- GREENBERG, J. Understanding metadata and metadata schemes. *Cataloging & Classification Quarterly*, v. 40, n. 3-4, p. 17-36, 2005.
- GUARINO, N.; GUIZZARDI, G.; MYLOPOULOS, J. On the philosophical foundations of conceptual models. *Information Modelling and Knowledge Bases*, v. 31, n. 321, p. 1-14, 2020.
- GUERREIRO, D.; BORBINHA, J. L. Humanidades Digitais: novos desafios e oportunidades. *Revista Internacional del Libro, Digitalización y Bibliotecas*, v. 2, n. 2, p. 1-16, 2014.
- INTERNACIONAL COUNCIL ON ARCHIVES. *Records in Contexts: A Conceptual Model For Archival Description*. Consultation Draft, v 0.1, set. 2016.
- LATORRE, M. *Historia de las Web 1.0, 2.0, 3.0 y 4.0*. Universidad Marcelino Champagnat, Lima, p. 1 - 8, 2018.
- MARCONDES, C. H. *Dados abertos interligados: publicação, recuperação e integração de acervos de arquivos, bibliotecas e museus na web*. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2021.
- MAYERNIK, M. S. Metadata. *Knowledge Organization*, v. 47, n. 8, p. 696-713, 2020.
- NETCRAFT. *July 2022 Web Server Survey*. Disponível em: <<https://news.netcraft.com/archives/category/web-server-survey/>>. Acesso em: 20 jul. 2022.
- RAMALHO, R. A. S.; MARTINS, P. G. M.; SOUSA, J. L. Evolução das linguagens de marcação: um breve histórico à luz da área de Ciência da Informação. *Informação@Profissões*, v. 6, n. 2, p. 20-34, 2017.
- RILEY, J. *Understanding metadata*. Washington DC: National Information Standards Organization, v. 23, 2017.
- RIVA, P.; LE BŒUF, P.; ŽUMER, M. *IFLA Library Reference Model: Um Modelo Conceitual para a Informação Bibliográfica*. IFLA, 2017.
- SCHNEIDER, L. Sintaxe e semântica na reformulação de algumas prescrições gramaticais. *Travessias*, v. 4, n. 3, p. 565-569, 2010.
- SMIRAGLIA, R. P. Introducing metadata. *Cataloging & Classification Quarterly*, v. 40, n. 3-4, p. 1-15, 2005.
- TEI CONSORTIUM. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Disponível em: <<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>>. Acesso em: 02 dez. 2021.

THE LIBRARY OF CONGRESS. *Bibliographic Framework Initiative*. Disponível em: <<https://www.loc.gov/bibframe/>>. Acesso em: 02 dez. 2021a.

THE LIBRARY OF CONGRESS. *MARC Standards*. Disponível em: <<https://www.loc.gov/marc/marcqinf.html#intro>>. Acesso em: 02 dez. 2021b.

THE LIBRARY OF CONGRESS. *Metadata Encoding and Transmission Standard*. Disponível em: <<https://www.loc.gov/standards/mets/>>. Acesso em: 02 dez. 2021c.

THE LIBRARY OF CONGRESS. *Metadata Object Description Schema*. Disponível em: <<https://www.loc.gov/standards/mods/>>. Acesso em: 02 dez. 2021d.

THE LIBRARY OF CONGRESS. *VRA Core Schemas and Documentation*. Disponível em: <<https://www.loc.gov/standards/vracore/schemas.html>>. Acesso em: 02 dez. 2021e.

TILLET, B. *O que é FRBR? Um modelo conceitual para o universo bibliográfico*. Washington, DC: Library of Congress, 2003.

W3C. *Resource Description Framework (RDF)*. Disponível em: <<https://www.w3.org/RDF/>>. Acesso em: 19 mar. 2022.

ZENG, M. L. Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article. *El profesional de la información*, v. 28, n. 1, p. 1-35, 2019.