

# As LAMs na Web Semântica: análise do panorama nacional

Juliana Martins de Castro Barroso<sup>1</sup> e Célia da Consolação Dias<sup>2</sup>

**A**s instituições conhecidas como LAMs (*libraries, archives and museums*, ou bibliotecas, arquivos e museus) são guardiãs de um rico acervo histórico e cultural e, para além dessa função de preservação, um dos mais valiosos objetivos dessas instituições é a promoção do acesso a esse conteúdo para o público em geral. Nesse sentido, a disponibilização de dados e informações no ambiente digital, especialmente na Web, vem se tornando, a cada dia, uma ferramenta mais poderosa na difusão do conhecimento.

A *World Wide Web*, rede mundial de computadores, teve seus primeiros desenvolvimentos a partir de 1989 e, desde então, vem caminhado de uma Web de documentos legíveis apenas por humanos para uma Web “de dados compreensíveis por máquina” (Berners-Lee, 1998). A capacidade de sistemas “compreenderem” dados e suas relações é referenciada na proposta de uma Web Semântica, na qual os “computadores teriam acesso a coleções de informações estruturadas e conjuntos de regras de inferência que poderiam ser utilizadas para conduzir um raciocínio automatizado” (Berners-Lee *et al.*, 2001, p. 37).

A publicação de acervos de acordo com os princípios da Web Semântica resulta não apenas em maior visibilidade para as instituições culturais, mas também na possibilidade ampliada de pesquisa, fruição, reúso e mixagem de dados. Porém, grande parte das LAMs ainda se encontra em um estágio anterior a esse processo, que envolve ações (e custos) para a seleção dos dados que serão publicados, sua digitalização e o reconhecimento de textos e imagens pelas máquinas (Zeng, 2019).

---

<sup>1</sup> Graduada em História pela Pontifícia Universidade Católica de Minas Gerais (PUC Minas), especialista em Gestão do Patrimônio Cultural pela Universidade Federal de Minas Gerais (UFMG), mestre e doutoranda em Gestão e Organização do Conhecimento pela Escola de Ciência da Informação da UFMG (PPGGOC/UFMG). Atuação profissional em gestão de acervos híbridos, organização da informação e do conhecimento e curadoria de dados culturais no contexto digital.

<sup>2</sup> Mestre e doutora em Ciência da Informação pela Escola de Ciência da Informação da UFMG. Coordenadora e professora permanente do Programa de Pós-Graduação em Gestão e Organização do Conhecimento (PPGGOC/UFMG), professora adjunta do Departamento de Organização e Tratamento da Informação da Escola de Ciência da Informação da UFMG e coordenadora do Grupo de Pesquisa Representação do Conhecimento e Recuperação da Informação (Recri/UFMG).

O objetivo deste estudo é analisar o panorama nacional de publicação de acervos de LAMs na Web, conferindo sua adequação aos princípios da Web Semântica. Trata-se de uma pesquisa bibliográfica e documental, que tem como fontes os dados da pesquisa TIC Cultura 2020 e dados coletados nas plataformas digitais das instituições e órgãos que representam as bibliotecas, os arquivos e os museus nacionalmente: a Biblioteca Nacional (BN), o Arquivo Nacional (AN) e o Instituto Brasileiro de Museus (Ibram). Na medida em que se responsabilizam pela consolidação de normas e políticas relacionadas à preservação de acervos no Brasil, é importante compreender de que forma a publicação de acervos na Web tem sido abordada nas práticas dessas instituições, já que isso influencia diretamente as demais LAMs nacionais.

Para isso, foram analisados: a disponibilidade de plataformas de busca de acervos na Web; o uso de *software* para a gestão e publicação dos acervos; o uso de padrões de metadados para descrição dos acervos; a existência de dados publicados de forma aberta e/ou conectada; e a existência de recomendações para a publicação de acervos culturais de forma digital – manuais, políticas ou legislações.

Busca-se, assim, compreender em que estágio está a publicação na Web de dados culturais no país, quais são as principais dificuldades identificadas para a difusão dessa prática e quais as oportunidades de avanços observadas, viabilizando sugestões de um possível caminho para a adoção dos princípios da Web Semântica para a difusão de informações e do conhecimento sobre o patrimônio cultural nacional.

## Análise

A Web Semântica, ou Web de dados, seria um projeto de extensão da Web tradicional, ou Web de documentos. A partir do acréscimo de contexto aos dados e de sua interconexão, seria possível a programas e dispositivos personalizados, chamados agentes, interagir com os dados e trocar informações entre si, viabilizando a automatização de uma série de tarefas. Para essa transição, Souza e Alvarenga (2004, p. 134) destacam que seria “necessária uma padronização de tecnologias, de linguagens e de metadados descritivos, de forma que todos os usuários da Web obedeam a determinadas regras comuns e compartilhadas sobre como armazenar dados e descrever a informação armazenada”.

Atingido esse grau de padronização, a proposta da Web Semântica avançaria para a conexão entre os dados publicados: “trata-se de fazer *links*, para que uma pessoa ou máquina possa explorar a teia de dados” (Berners-Lee, 2010, para. 1). Os dados conectados (*linked data*) constituem a proposta de um “banco de dados global, onde um conjunto crescente de informações possa ser acessado por um conjunto diversificado de aplicações com os mais diferentes propósitos” (Laufer, 2015, p. 70). Em 2010, Berners-Lee acrescentou ao *linked data* a recomendação de que os dados fossem publicados de forma aberta, fazendo uso de licenças abertas que viabilizem sua reutilização, constituindo a iniciativa que atualmente é conhecida como *Linked Open Data* (LOD), ou dados abertos conectados.

Os dados disponibilizados na Web sobre acervos das LAMs brasileiras ainda estão pouco alinhados à iniciativa LOD. A própria disponibilização desses dados não é uma prática majoritária entre as instituições nacionais. De acordo com a pesquisa TIC Cultura 2020, embora a maior parte dos equipamentos culturais brasileiros tenha algum tipo de presença *online* (*website* próprio, *website* de terceiros ou redes sociais), apenas 51% dos arquivos, 25% dos museus e 6% das bibliotecas disponibilizaram acervos digitais ao público por meio da Internet. Os catálogos das coleções, por sua vez, representam um número ainda menor: somente 33% dos arquivos, 15% dos museus e 12% das bibliotecas disponibilizaram seus catálogos na Internet (Comitê Gestor da Internet no Brasil [CGI.br], 2021).

Há ainda grandes entraves no uso da Internet como plataforma para a difusão dos acervos nacionais, como a falta de recursos financeiros para investimento na área de tecnologia (afeta 57% dos arquivos, 50% das bibliotecas e 49% dos museus participantes da pesquisa), baixa velocidade de conexão com a Internet (24% dos arquivos, 32% das bibliotecas e 26% dos museus), equipamentos ultrapassados (39% dos arquivos, 41% das bibliotecas e 36% dos museus) ou insuficientes (32% dos arquivos, 42% das bibliotecas e 29% dos museus), número insuficiente de computadores conectados à Internet (22% dos arquivos, 41% das bibliotecas e 26% dos museus), ausência de suporte técnico (24% dos arquivos, 32% das bibliotecas e 24% dos museus) e pouca capacitação da equipe (27% dos arquivos, 36% das bibliotecas e 26% dos museus). Os dados da pesquisa também apontam que apenas cerca de metade das instituições bibliográficas e museológicas faz uso de algum padrão de metadados para a catalogação de seu acervo – entre as instituições arquivísticas o percentual é maior, de 70% –, o que afeta diretamente a capacidade de interpretação desses dados por máquinas (CGI.br, 2021).

Dessa forma, foram consultados os *sites* das três instituições que representam as bibliotecas, os arquivos e os museus brasileiros em nível federal – a BN, o AN e o Ibram –, buscando identificar possíveis movimentos no sentido de orientar a disponibilização de dados sobre os acervos das LAMs na Web, bem como verificar de que forma essas instituições publicam seus próprios acervos.

## BIBLIOTECA NACIONAL

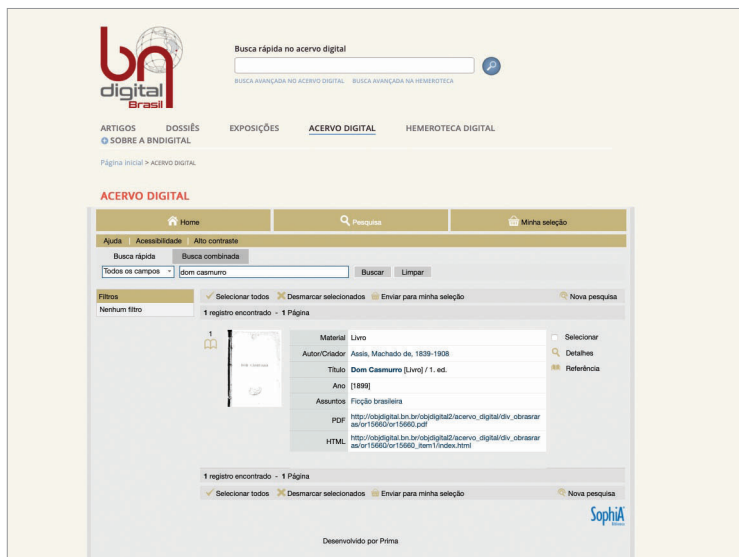
A BN, uma das mais antigas instituições culturais brasileiras, fundada em 1810, é responsável pela política nacional relacionada a bibliotecas. A BN foi uma das primeiras instituições do Brasil a disponibilizar serviços digitais relacionados a acervos, com a publicação de seus catálogos na Internet a partir de 1998. Em 2006, foi criada a Biblioteca Nacional Digital (BNDigital), disponibilizando ao público suas coleções digitalizadas.

A partir de 2017, começou a ser formulada a Política de Preservação Digital da Biblioteca Nacional (PPDBN), que recomenda a utilização de normas e padrões internacionais para balizar as ações de preservação digital. Em relação aos metadados para representação dos documentos digitais, embora não sejam especificados padrões, o documento ressalta que:

Deverão ser adotados padrões reconhecidos internacionalmente que permitam o reuso e a interoperabilidade entre sistemas contemplando metadados descritivos, administrativos (técnicos, preservação e direitos), estruturais, marcação e identificadores persistentes, tendo em vista a independência da plataforma, integridade do objeto, preservação e acesso permanente (Fundação Biblioteca Nacional [FBN], 2020, p. 13).

Na prática, a BN optou por adotar, em sua biblioteca digital, o padrão de metadados internacional Dublin Core (DC), destinado a objetos digitais em geral. O padrão, publicado pela primeira vez em 1995, conta atualmente com 15 termos, “expresso em vocabulários RDF para uso em *Linked Data*” (Dublin Core, 2022, Section 1).

FIGURA 1  
PÁGINA DE BUSCA DA BNDIGITAL



FONTE: ACERVO DIGITAL BNDIGITAL (HTTPS://BNDIGITAL.BN.GOV.BR/ACERVODIGITAL/).

A adoção desse padrão pela BN se deu com o objetivo de viabilizar a cooperação com outras bibliotecas digitais, de escopos nacional e internacional, o que constitui um importante passo com relação à publicação de dados interoperáveis na Web. O acervo também é descrito com *tags* do padrão *Machine Readable Cataloging* (Marc) e, além disso, é utilizada a norma *Anglo American Cataloguing Rules* (AACR2) para representar os pontos de acesso de autoria, havendo também uma base de terminologia utilizada para controle de vocabulário de indexação.

Os catálogos de acervos da BN estão disponíveis *online*, e são registrados por meio do *software* de gerenciamento SophiA. A busca pode ser realizada por qualquer usuário, sem necessidade de cadastro prévio, sendo possível refiná-la por título, autor, assunto, entre outros, conforme a Figura 1.

Os dados sobre os acervos publicados na BNDigital não possuem *links* para recursos externos, e não está claro se estão sob uma licença de uso aberta, já que os direitos de publicação estão descritos apenas como “Biblioteca Nacional”. Assim, não se pode afirmar que os acervos estão publicados no âmbito da iniciativa LOD, porém, a infraestrutura e a padronização adotadas favorecem a conectividade desses dados com outros recursos publicados na Web, facilitando iniciativas baseadas em integração e agregação de dados.

## ARQUIVO NACIONAL

A gestão arquivística brasileira é normatizada pelo AN, fundado ainda em 1838. Em 2004, foi criado o Sistema de Informações do Arquivo Nacional (Sian), base de dados informatizada que ainda é a principal fonte de acesso à documentação da instituição. Em 2006, foi elaborado o Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos (e-Arq Brasil), que busca orientar a criação de sistemas de gestão arquivística, garantindo a confiabilidade, a autenticidade e a acessibilidade aos documentos. Também em 2006 foi consolidado o conjunto de diretrizes arquivísticas nacionais, com a edição da Norma Brasileira de Descrição Arquivística (Nobrade), que estabelece 28 elementos para a descrição de documentos arquivísticos.

FIGURA 2  
PÁGINA DE BUSCA DO SIAN

The screenshot displays the SIAN search interface. At the top, the header includes the logo of the Ministério da Justiça e Cidadania and the SIAN logo. Below the header, there is a navigation menu with options like 'Fundos/Coleções', 'MAPA', 'Favoritos', 'Estatística', 'Editar Cadastro', 'Contatos', 'Ajuda', and 'Sair'. The main content area is titled 'Fundos/Coleções - Pesquisa Simples'. On the left, there are filter options for 'PERÍODO', 'REPOSITÓRIO', 'COD DE REFERENCIA', and 'ORDENAÇÃO'. The search bar contains the text 'Hom casmurro' and a 'Pesquisar' button. Below the search bar, it indicates 'Encontrados 46 registros.' The first search result is displayed as follows:

**FUNDO: Serviço de Censura de Diversões Públicas - RJ - BR RJANRIO TN**

**BR RJANRIO TN\_CPR\_PTE\_1387 - Hom casmurro - a flor e o fruto - Dossiê - ARQUIVO: BR\_RJANRIO\_TN\_CPR\_PTE\_1387\_d0001de0001.pdf** - 55,15 MB

autor: assis, joaquim maria machado de (joaquim maria machado de assis, machado de assis) adaptação: josé carlos cavalcanti borges certificado 3940/71

📄 Produção Inicial: 1971 📄 Produção Final: 📄 Assunto Inicial: 📄 Assunto Final: 📄 Ver detalhes 📄 Arquivo Digital

FONTE: SIAN (HTTPS://SIAN.AN.GOV.BR/SIANEX/CONSULTA/RESULTADO\_PESQUISA\_NEW.ASP).

A busca no acervo mantido pelo AN pode ser feita na plataforma Sian, que demanda cadastro para acesso – o que pode ser feito diretamente na plataforma ou o usuário pode utilizar seu *login* da plataforma de serviços gov.br. A busca simples é feita por termos livres e os resultados podem ser refinados por data e ordenados de diferentes maneiras (por relevância, data, título e código de referência), conforme apresentado na Figura 2. É possível também realizar buscas avançadas e consultar os instrumentos de pesquisa arquivísticos para resultados mais assertivos.

A descrição dos itens documentais é feita com base nos elementos descritos pela Nobrade, não sendo adotado nenhum padrão de metadados específico. Como o banco de dados apresenta restrição de acesso, os dados descritivos não são conectados a recursos externos. Nos itens pesquisados, não havia descrição da licença de publicação e uso adotada, mas os documentos digitalizados podem ser acessados na íntegra.

A padronização da descrição arquivística existe e está de acordo com a norma nacional da área, porém a não utilização de padrões de metadados e a adoção de uma plataforma fechada fazem com que haja pouca aderência, ainda, a iniciativas relacionadas ao LOD, à integração e à agregação de dados.

## INSTITUTO BRASILEIRO DE MUSEUS

O Ibram foi criado em 2009, por meio da Lei n. 11.906, para elaborar, implementar e gerenciar as políticas públicas do setor museológico nacional. Diferentemente da BN e do AN, o Ibram não se trata de uma instituição de salvaguarda de acervos, embora tenha, sob sua responsabilidade, uma série de instituições museológicas em todo o país.

Com relação à documentação dos objetos de museus, a lei de criação do Ibram determina que é competência do órgão “promover o inventário sistemático dos bens culturais musealizados” (Lei n. 11.906/2009). Assim, foi instituído o Inventário Nacional dos Bens Culturais Musealizados (INBCM), por meio do Decreto n. 8.124, de 2013, normatizado pela Resolução Normativa n. 6 do Ibram, de 2021.

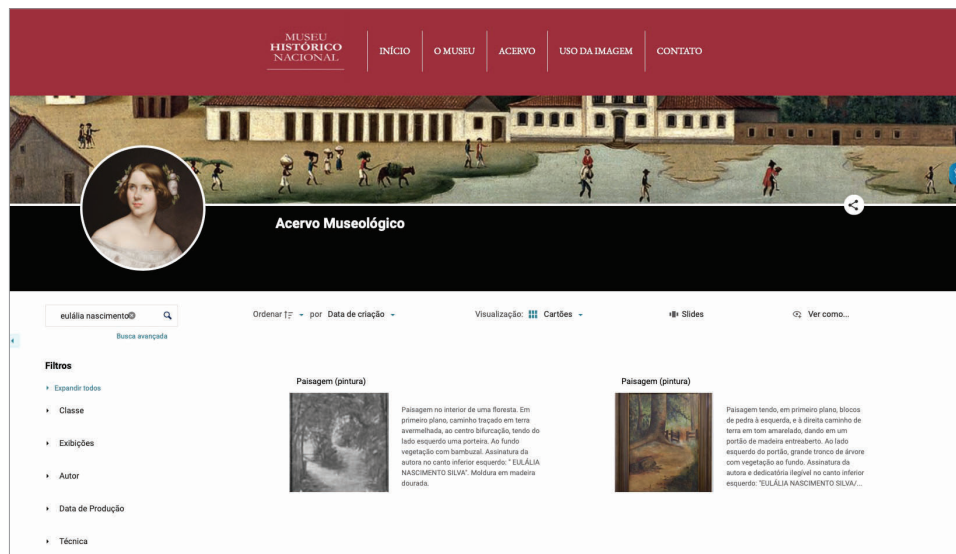
A Resolução Normativa n. 6 define os elementos de descrição para identificação do bem cultural musealizado no INBCM, levando em conta as especificidades das áreas de Museologia, Biblioteconomia e Arquivologia. São definidos 15 elementos para a descrição de objetos museológicos, 19 elementos para itens bibliográficos e 16 elementos para documentos arquivísticos.

Não há, até o momento, normatização elaborada por parte do Ibram para a inclusão da documentação museológica em ambientes digitais. Entretanto, em 2013, o Instituto lançou o Programa Acervo em Rede, com o objetivo de “promover a democratização do acesso digital aos bens culturais musealizados, promovendo também a digitalização e a documentação dos acervos das instituições museológicas na Internet” (Ibram, 2022, para. 1). No âmbito desse Programa, foi realizada uma parceria, a partir de 2016, com a Universidade Federal de Goiás (UFG) para o desenvolvimento do Tainacan, um *plugin* para WordPress, sistema de gerenciamento de conteúdo (SGC) utilizado para construção de *sites*.

Até o momento, 20 museus federais já contam com acervos publicados por meio do Tainacan. A busca pode ser feita de forma simples ou avançada, podendo ser refinada pelos elementos descritivos do acervo, como apresentado na Figura 3.

FIGURA 3

## PÁGINA DE BUSCA DO TAINACAN, NO SITE DO MUSEU HISTÓRICO NACIONAL (MHN)



FONTE: ACERVO DO MHN ([HTTPS://MHN.ACERVOS.MUSEUS.GOV.BR/](https://mhn.acervos.museus.gov.br/)).

Embora a caixa de busca avançada se refira aos elementos descritivos como “metadados”, não há uso de um padrão de metadados propriamente dito, e sim dos elementos indicados pelo INBCM. Não é indicado também o uso de esquemas específicos de regras para catalogação e vocabulários controlados. No caso do MHN, apenas o campo “Classe” apresenta controle, conforme o guia *Thesaurus para Acervos Museológicos* (Ferrez & Bianchini, 1987).

Os dados não são conectados com recursos externos, mas há indicação das condições de uso e reprodução. O fato de ser utilizada uma plataforma aberta facilita a tarefa de conexão com dados externos, sendo necessário, futuramente, um esforço de enriquecimento semântico dos dados, além da padronização de metadados descritivos e de conteúdo (vocabulários controlados).

## Conclusão

Como anteriormente apontado, as instituições nacionais, em geral, fazem pouco uso de princípios da Web Semântica e da iniciativa LOD para a publicação de acervos online. A BN, o AN e o Ibram, que são órgãos representativos e orientadores para as LAMs nacionais, possuem plataformas para a publicação de acervos, mas apenas a BN utiliza um padrão de metadados internacionalmente consolidado para sua descrição. Em nenhuma das plataformas analisadas, os dados foram publicados no formato de links, conectados a recursos externos. Na BNDigital e na plataforma Tainacan, os links remetem a outros recursos internos, apresentando, por exemplo, todas as obras de um mesmo autor.

Sobre a encontrabilidade dos dados, a plataforma Sian apresenta um entrave ao demandar o acesso identificado dos usuários, não permitindo assim a varredura por buscadores externos, possível nas plataformas BNDigital e Tainacan. Esse é um ponto importante para a visibilidade da instituição e de seu acervo, já que nem sempre o usuário sabe, de antemão, quais instituições possuem informações significativas para sua pesquisa, partindo de buscas mais genéricas em plataformas como o Google.

A BN e o AN possuem documentos que normatizam a gestão, a preservação e a publicação de acervos digitais, o que representa um guia importante para as demais instituições bibliográficas e arquivísticas nacionais. O Ibram ainda não conta com um documento nesse sentido, e, embora a plataforma Tainacan disponibilize orientações para sua implementação, identifica-se a necessidade de uma política mais ampla para a gestão de acervos em ambientes digitais.

A consolidação das informações obtidas na pesquisa realizada junto aos *sites* da BN, do AN e do Ibram encontra-se na Tabela 1, a seguir.

TABELA 1

**CATEGORIAS DE ANÁLISE DA PUBLICAÇÃO DE DADOS NA WEB**

Categoria	Instituição		
	Biblioteca Nacional	Arquivo Nacional	Museus Ibram
Existência de plataforma <i>online</i> para busca de acervos	Sim – BNDigital	Sim – Sian	Sim – Tainacan
Uso de <i>software</i> de registro de acervos	Sim – SophiA ( <i>software</i> pago)	Sim – Sian (desenvolvimento próprio)	Sim – WordPress (sistema pago)
Uso de padrão de metadados	Sim – Dublin Core	Não (uso de elementos descritivos da Nobrade)	Não (uso de elementos descritivos do INBCM)
Publicação de dados abertos e/ou conectados (LOD)	Não	Não	Não
Existência de recomendações para a gestão e/ou publicação de acervos digitais	Sim – PPDBN	Sim – e-Arq Brasil	Não

FONTE: ELABORADO PELOS AUTORES.

A digitalização e a publicação de acervos culturais na Web é cada vez mais parte fundamental da missão das LAMs. A população brasileira em geral possui, de acordo com a pesquisa TIC Domicílios 2020, um amplo acesso à Internet<sup>3</sup>, e, portanto, a democratização do acesso aos bens culturais também deve passar por esse espaço. É importante ressaltar, além disso, que o esforço para a disponibilização desses bens de forma digital pode e deve ser acompanhado pela adoção de padrões relacionados à Web Semântica, o que possibilita uma melhor descoberta dos acervos culturais pela sociedade, que tem o direito de usufruir, de forma intuitiva, acessível e aberta, dos dados sobre bens que podem colaborar para seu desenvolvimento educacional, profissional e cultural.

---

<sup>3</sup> 82% dos domicílios brasileiros possuíam acesso à Internet em 2021, embora, quando investigadas as classes econômicas, as classes DE apresentavam apenas 61%, contra 89% da classe C, 98% da classe B e 100% da classe A (CGI.br, 2022).

## Referências

- Berners-Lee, T. (1998). *The World Wide Web: A very short personal history*. W3C. <http://www.w3.org/People/Berners-Lee/ShortHistory.html>
- Berners-Lee, T. (2010). *Linked data*. W3C. <https://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34-43.
- Comitê Gestor da Internet no Brasil. (2021). *Pesquisa sobre o uso das tecnologias de informação e comunicação nos equipamentos culturais brasileiros: TIC Cultura 2020*. [https://cetic.br/media/docs/publicacoes/2/20210616181537/tic\\_cultura\\_2020\\_livro\\_eletronico.pdf](https://cetic.br/media/docs/publicacoes/2/20210616181537/tic_cultura_2020_livro_eletronico.pdf)
- Comitê Gestor da Internet no Brasil. (2022). *Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros: TIC Domicílios 2021*. [https://cetic.br/media/docs/publicacoes/2/20211124201233/tic\\_domicilios\\_2020\\_livro\\_eletronico.pdf](https://cetic.br/media/docs/publicacoes/2/20211124201233/tic_domicilios_2020_livro_eletronico.pdf)
- Dublin Core. (2022). *DCMI metadata terms*. <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#section-1>
- Ferrez, H. D., & Bianchini M. H. S. (1987). *Thesaurus para acervos museológicos* (Vol. 1). Fundação Nacional Pró-Memória.
- Fundação Biblioteca Nacional. (2020). *Política de preservação digital da Biblioteca Nacional*. <https://www.gov.br/bn/pt-br/acao-a-informacao-2/acoes-e-programas/documentos/politica-de-preservacao-digital-da-biblioteca-nacional-ppdbn>
- Instituto Brasileiro de Museus. (2022). *Acervo em Rede e Projeto Tainacan*. <https://www.gov.br/museus/pt-br/acao-a-informacao/acoes-e-programas/acervo-em-rede-e-projeto-tainacan>
- Laufer, C. (2015). *Guia de Web Semântica*. Governo do Estado de São Paulo; Governo do Reino Unido; Núcleo de Informação e Coordenação do Ponto BR. <http://www.governoaberto.sp.gov.br/wp-content/uploads/2016/05/Book-Web-Semantica.pdf>
- Lei n. 11.906, de 20 de janeiro de 2009*. (2009). Cria o Instituto Brasileiro de Museus – IBRAM, cria 425 (quatrocentos e vinte e cinco) cargos efetivos do Plano Especial de Cargos da Cultura, cria Cargos em Comissão do Grupo-Direção e Assessoramento Superiores - DAS e Funções Gratificadas, no âmbito do Poder Executivo Federal, e dá outras providências. [http://www.planalto.gov.br/ccivil\\_03/\\_ato2007-2010/2009/lei/11906.htm](http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2009/lei/11906.htm)
- Souza, R. R., & Alvarenga, L. A. (2004). Web Semântica e suas contribuições para a ciência da informação. *Ciência da Informação*, 33(1), 132-141.
- Zeng, M. L. (2019). Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article. *El profesional de la información*, 28(1), 1-35.