



UNIVERSIDADE ESTADUAL PAULISTA  
"JÚLIO DE MESQUITA FILHO"  
Campus de Marília



CULTURA  
ACADÊMICA  
*Editora*

# Metadados de preservação digital na era do Big Data

Daniel Martínez-Ávila  
Jean Fernandes Brito

**Como citar:** MARTÍNEZ-ÁVILA, D.; BRITO, J. F. Metadados de preservação digital na era do Big Data. *In:* MARTÍNEZ-ÁVILA, D.; SOUZA, E. A.; GONZALEZ, M. E. Q. (org.).

**Informação, Conhecimento, Ação Autônoma e Big Data: Continuidade ou Revolução?**

Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2019. p. 129-142.

DOI: <https://doi.org/10.36311/2019.978-85-7249-055-9.p129-142>



All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported.

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição - Uso Não Comercial - Partilha nos Mesmos Termos 3.0 Não adaptada.

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-NoComercial-CompartirIgual 3.0 Unported.

## METADADOS DE PRESERVAÇÃO DIGITAL NA ERA DE *BIG DATA*

*Jean Fernandes Brito*  
*Daniel Martínez-Ávila*

Com o avanço das tecnologias, os suportes de informação e dispositivos tornam-se obsoletos cada vez mais rapidamente. Além dos suportes, na era dos grandes volumes de dados, a informação contida neles também apresenta as características dos *Big Data*, consideradas os cinco V's: volume, velocidade, veracidade, variedade e valor. Nesse sentido, as descrições adequadas dos objetos digitais que transmitem os *Big Data* necessitam de uma representação efetiva para garantir a preservação das informações e registros descritivos.

Nesse contexto, além de se tornarem os suportes obsoletos, há um grande volume de dados disponível, que corre o risco de desaparecer ou se distorcer, havendo a necessidade de criar e pensar meios de descrever e preservar os objetos físicos e ainda preservar os metadados de objetos digitais que correspondem aos *Big Data*.

Os *Big Data* representam quantidades maciças de diversos tipos de informação, que são constantemente

<https://doi.org/10.36311/2019.978-85-7249-055-9.p129-142>

coletadas, processadas, misturadas e analisadas estatisticamente na esperança de encontrar correlações que revelem significado (MARTÍNEZ ÁVILA *et al.*, 2018).

Nesse sentido, essas quantidades massivas de dados e informações necessitam ser representadas para garantir a preservação digital e proporcionar contexto a teorias e hipóteses científicas (pragmática). Os metadados de preservação digital têm como propósito gerenciar e administrar a representação e a descrição dos objetos digitais, disponibilizar as informações suficientes para tomada de decisões dos gestores, e manter o acesso ao conteúdo dos objetos em um ambiente tecnológico em longo prazo.

Nesse contexto, esses metadados devem ser criados de maneira adequada para que futuramente “[...] os dados possam ser copiados, ‘re-criados’ e atualizados como migrações periódicas necessárias em novos formatos para seu uso em novas gerações de *hardware* e *software*” (MÁRDERO ARELLANO, 2004, p. 21).

Desse modo, os profissionais da informação que lidam com os metadados de preservação digital devem sempre atualizar o acervo digital e acompanhar as mudanças tecnológicas e informacionais, para, assim, disponibilizar informações atuais do cotidiano para os usuários.

Castro (2012) comenta que a integração semântica dos metadados seria “o processo de utilização de uma representação conceitual dos dados e seus relacionamentos, com o objetivo de eliminar as possíveis heterogeneidades de compreensão e interpretação”.

Diante do exposto, a presente pesquisa tem por objetivo apresentar uma reflexão teórica sobre os metadados de preservação digital à luz do contexto de *Big Data*.

A pesquisa está ancorada em uma abordagem teórica, descritiva e qualitativa. Segundo Bartunek e Seo (2002), a abordagem qualitativa é útil e necessária para identificar e explorar os significados dos fenômenos estudados, possibilitando o desenvolvimento de novas compreensões e reflexões – no caso desse estudo, os metadados de preservação digital no contexto dos *Big Data*.

O trabalho está estruturado da seguinte forma para melhor entendimento: introdução, em que são apresentadas as considerações iniciais e os objetivos; seções: (1) os metadados e a preservação digital, seus conceitos e funções; (2) os metadados de preservação digital e *Big Data*: algumas reflexões; e considerações finais.

## **1 METADADOS DE PRESERVAÇÃO DIGITAL**

Os Metadados têm a função de descrever, identificar, e recuperar os objetos informacionais de acordo com seu conteúdo, contexto, estrutura ou apresentação, para que no futuro o usuário encontre a informação que deseja (ALVES, 2010).

Eles possuem a função de padronização e interoperabilidade dos sistemas de informação, como por exemplo, bibliotecas, mapas, redes sociais e qualquer outro meio, para que elas possam trocar e compartilhar informações. Se os metadados possuem essas funções, pode-se afirmar que estão relacionados à catalogação de recursos informacionais no meio digital, que descreve e representa os documentos.

Os metadados podem ser classificados em várias categorias, entretanto os mais frequentes são: os descritivos, estruturais e administrativos.

**Metadados descritivos:** é a face mais conhecida dos metadados, são eles que descrevem um recurso com o propósito de descoberta e identificação; podem incluir elementos tais como título, autor, resumo, palavras-chave e identificador persistente.

**Metadados estruturais:** são informações que documentam como os recursos complexos, compostos por vários elementos, devem ser recompostos e ordenados. Por exemplo, como as páginas de um livro, digitalizadas separadamente, são vinculadas entre si e ordenadas para formar um capítulo.

**Metadados administrativos:** fornecem informações que apoiam os processos de gestão do ciclo de vida dos recursos informacionais. Incluem, por exemplo, informações sobre como e quando o recurso foi criado e a razão da sua criação. Nessa categoria, estão metadados técnicos que explicitam as especificidades e dependências técnicas do recurso; inclui também os metadados voltados para apoio à gestão dos direitos relacionados ao recurso (SAYÃO, 2010, p. 5, grifo do autor).

Para Campos (2007, p. 30), “[...] os metadados tornam-se especialmente úteis em facilitar a navegação e o planejamento das consultas” de acordo com cada perfil diferente de usuário, para que, quando este for realizar a busca, encontre o que realmente deseja obter.

Os metadados são utilizados, por exemplo, em bibliotecas e arquivos, sendo que, nesses ambientes, possuem funções que vão além da identificação e descrição dos documentos, referindo-se também a várias seções.

Metadados referentes à administração, acesso, preservação e uso das coleções são constantemente empregados em repositórios. Isso é particularmente notável em bibliotecas e arquivos digitais. Nesses ambientes, os metadados não apenas descrevem e identificam um objeto informacional, mas explicitam as condições corretas ou ideais de seu gerenciamento, as relações do objeto com outros na coleção, sua função, utilização, comportamento, contexto de criação e condições de preservação (CAMPOS, 2007, p. 34-35).

Para não correr o risco de perder os materiais digitais deve-se utilizar *software*, *hardware*, e métodos de preservação com qualidade e que armazenem as informações e os dados de maneira adequada para permanecer em longo prazo, utilizando-se padrões de metadados.

A seguir, no Quadro 1, serão apresentados os principais padrões de metadados de preservação digital que foram apresentados na revisão de literatura por Formenton (2015), bem com a sua especificação e apresentação.

### Quadro 1- Padrões de Metadados de Preservação Digital

Padrão de Metadados	Apresentação
ANSI/NISO Z39.87, <i>Data Dictionary – Technical Metadata for Digital Still Images</i> 32	Padrão de metadados técnicos para desenvolver, intercambiar e interpretar arquivos de imagem fixa digital, facilitar a interoperabilidade entre sistemas, serviços e software, assim como apoiar a gestão a longo prazo e contínuo acesso a coleções de imagens digitais (BARBEDO <i>et al.</i> 2007; NATIONAL INFORMATION STANDARDS ORGANIZATION 2011)
<i>Dublin Core - DC</i>	Padrão para catalogação e suporte a descoberta de recursos eletrônicos na Web (CAMPOS; SARAMAGO, 2007, 2012; SAYÃO, 2010)

<i>Encoded Archival Description (EAD)</i>	Padrão para codificação de instrumentos arquivísticos de acesso, tais como inventário, índices, dentro outros (ANDRADE, 2007; BARBEDO <i>et al.</i> ; CHAN ZENG, 2006)
<i>Learning Object Metadata (LOM)</i>	Padrão que define um conjunto mínimo de atributos para gerir, localizar e avaliar recursos ou objetos de aprendizagem (CHAN; ZENG, 2006, NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004)
<i>Machine Readable Cataloging (MARC)</i>	Padrão para a representação e a comunicação de registros ou informações bibliográficas entre bibliotecas, através da utilização de computadores (CHAN, ZENG, 2006; GRACIO, 2012; THOMAZ, 2004)
<i>Metadata Encoding and Transmission Standard (METS)</i>	Padrão para codificar metadados descritivos, administrativos e estruturais sobre objetos digitais num repositório (ALMEIDA; CENDÓN; SOUZA 2012, BARBEDO <i>et al.</i> 2007)
<i>Metadata Object Description Schema (MODS)</i>	Padrão de metadados descritivos derivado do esquema bibliográfico MARC 21, cujo enfoque são os recursos eletrônicos e os serviços de biblioteca. (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004)
<i>ISO/IEC Moving Picture Experts Group (MPEG) Multimedia</i>	Padrões para a descrição e a representação de recursos ou objetos audiovisuais e multimídiaicos digitais tais como MPEG-7 e MPEG-21 (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004; SAYÃO, 2010)
<i>PREservation Metadata Implementation Strategies (PREMIS)</i>	Padrão que define um conjunto de elementos básicos para codificar, armazenar, gerir e intercambiar metadados de preservação no contexto de um sistema de repositório para a preservação digital (ARELLANO, 2008; CAMPOS; SARAMAGO, 2007, SAYÃO, 2010)
<i>Record Export for Art and Cultural Heritage (REACH)</i>	Padrão para a descrição de objetos no âmbito de coleções de museus (GRACIO, 2012)

**Fonte:** (FORMENTON, 2015, p. 62)

Para entender o processo de preservação digital é necessário compreender primeiro o que é um objeto digital, sendo assim, “Um objeto digital é aquele que foi criado em computador, podendo ser original ou uma versão depois de haver sido convertido (ou digitalizado).” (MÁRDERO ARELLANO, 2004, p. 16). Para Hedstrom (1998, p. 190, tradução nossa), a preservação digital é “[...] o planejamento, alocação de recursos e aplicação de métodos e tecnologias para assegurar que a informação digital de valor contínuo permaneça acessível e utilizável”.

A preservação digital busca armazenar as informações e os dados de uma instituição para que permaneça sempre em acesso e uso contínuo, preserva também a memória e a cultura de uma determinada comunidade. “A preservação digital compreende os mecanismos que permitem o armazenamento em repositórios de dados digitais que garantiriam a perenidade dos seus conteúdos” (MÁRDERO ARELLANO, 2004, p. 17).

## **2 OS METADADOS DE PRESERVAÇÃO DIGITAL NA ERA DOS *BIG DATA*: ALGUMAS REFLEXÕES**

De modo a elucidar as reflexões dos metadados de preservação digital à luz dos *Big Data*, apresenta-se uma conceituação desse termo:

O Big Data é um termo derivado dos avanços recentes relativos à massificação da utilização de recursos tecnológicos e da farta produção de dados. Em suma, é um conceito que caracteriza volumosos conjuntos de dados heterogêneos, os quais não são passíveis de processamento por soluções computacionais tradicionais considerando dinamismo e sua complexidade (RAUTENBERG; CARMO, 2019, p 57).

Rautenberg e Carmo (2019), ao apresentarem e discutirem esse termo, dialogam com aportes teóricos da Ciência da Informação, no que diz respeito à otimização do acesso e à forma como os metadados se comportam nesse processo de representação da informação.

Os autores supracitados apresentam, ainda, alguns questionamentos sobre a aplicação dos dados e metadados, ancorando sua reflexão à luz dos metadados de preservação digital. São questionamentos dos autores: Como armazenar os dados e metadados em ecossistemas de *Big Data*? Como organizar e catalogar dados e metadados nesses ecossistemas? Ao colocar esses questionamentos diante do grande volume de dados e ancorá-los à luz dos metadados de preservação digital e de seu processo de armazenamento, deve-se pensar numa política dos padrões de metadados de preservação digital que discuta esse grande volume de dados.

Grácio (2012, p. 75) caracteriza os aspectos da preservação digital em três grupos: organizacional, legal e técnico que se subdividem:

**O Organizacional** em: missão, visão, objetivos institucionais; equipe multidisciplinar; responsabilidades; recursos financeiros e atos administrativos.

**O Legal** em: legislação vigente e direitos autorais;

**O Técnico** em: modelos, padrões, iniciativas; infraestrutura tecnológica; repositórios institucionais; seleção, descarte; estratégias de preservação; autenticidade; metadados e suporte digital.

A figura a seguir exemplifica melhor a relação entre esses grupos.

**Figura 1:** Aspectos da Preservação digital



**Fonte:** Fonte: Grácio (2012, p. 76).

A prática de preservação digital é necessária em todas as unidades de informação para que os usuários tenham acesso *on-line* e para que a informação esteja sempre preservada e ativa, pois os documentos ou conteúdos que estão no suporte físico (as informações impressas, livros, periódicos, teses, dissertações, entre outros) com o passar do tempo podem se desintegrar ou perder, nunca mais sendo recuperadas, lembrando que, o usuário não pode ter acesso ao documento impresso a todo o momento e nem poderá pesquisar em qualquer lugar que esteja (ARAKAKI *et al.*, 2018)..

## CONSIDERAÇÕES FINAIS

Nas discussões sobre *Big Data* no contexto documental, apresentamos um novo V que caracteriza os *Big Data* e a informação em ambientes digitais: a *volatilidade*. Neste sentido, consideramos os metadados de preservação digital como

uma ferramenta que facilitaria a preservação e melhor contextualização dos dados.

Os metadados de preservação digital assumem um papel primordial nas descrições e representações dos dados e da informação possibilitando a interoperabilidade entre sistemas. As políticas de preservação digital no contexto dos *Big Data* possuem grande importância global, pois têm o papel de garantir o armazenamento e a descrição não ambígua de informações, permitindo também uma melhor significação contextual dos *Big Data* em relação ao conhecimento acumulado.

## REFERÊNCIAS

ALMEIDA, M.B.; CENDÓN, B.V. & SOUZA, R.R. Metodologia para implantação de programas de preservação de documentos digitais a longo prazo. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, Florianópolis, v. 17, n. 34, p. 103-130, ago. 2012. ISSN 1518-2924. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17n34p103>. Acesso em: 27/05/2019.

ALVES, R. C. V. *Metadados como elementos do processo de catalogação*. 2010. 132 f. Tese (Doutorado em Ciência da Informação)–Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, 2010.

ANDRADE, R. S. Aspectos introdutórios da representação de informação arquivística: a Norma Brasileira de Descrição Arquivística (Nobrade), a Descrição Arquivística Codificada (EAD-DTD) e o projeto Archives Hub. *Ponto de Acesso*, Salvador, v. 1, n. 2, p. 70-100, jul./dez. 2007. Disponível em: <http://www.brapci.inf.br/index.php/article/view/0000005065/7cd7f7403d73975ed2619d2d8310720/>. Acesso em: 23/05/2019.

ARAKAKI, F. et al. Web Semântica e Preservação digital: o padrão de metadados PREMIS na proposta do Liked Data. *Informação e Tecnologia*. v. 2, 2018. Disponível em: <http://www.periodicos.ufpb.br/ojs/index.php/itec/article/view/38118>. Acesso em: 22/05/2019.

- ARELLANO, M. Á. M. *Crerios para a preservação digital da informação científica*. 2008. 354 f. Tese (Doutorado em Ciência da Informação) – Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2008. Disponível em: [http://repositorio.unb.br/bitstream/10482/1518/1/2008\\_MiguelAngelMarderoArellano.pdf](http://repositorio.unb.br/bitstream/10482/1518/1/2008_MiguelAngelMarderoArellano.pdf). Acesso em: 11/01/2015
- BARBEDO, F. *et al.* RODA: repositório de objectos digitais autênticos. *Actas dos Congressos Nacionais de Bibliotecários, Arquivistas e Documentalistas*, Lisboa, n. 9, 2007. Disponível em: <https://www.bad.pt/publicacoes/index.php/congressosbad/article/view/535>. Acesso em: 13/05/ 2019.
- BARTUNEK, J. M. & SEO, M. Qualitative research can add new meanings to quantitative research. *Journal of Organizational Behavior*, v. 23, n.2, 2002.
- CAMPOS, L. F. de B. Metadados digitais: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 12, n. 23, p. 16-46, 1. sem. 2007. Disponível em: <http://www.brapci.ufpr.br/download.php?dd=11673>. Acesso em: 26/05/2017.
- CAMPOS, F. M. G.; SARAMAGO, M. L. Preservação digital de longo prazo em instituições patrimoniais: reutilização e adaptação de metadados. *Actas dos Congressos Nacionais de Bibliotecários, Arquivistas e Documentalistas*, Lisboa, n. 9, p. 1-7, 2007. Disponível em: <https://www.bad.pt/publicacoes/index.php/congressosbad/article/view/540>. Acesso em: 11/05/2019.
- CASTRO, F. F. de. *Elementos de interoperabilidade na catalogação descritiva: configurações contemporâneas para a modelagem de ambientes informacionais digitais*. 2012. 202 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2012.
- CHAN, L. M.; ZENG, M. L. Metadata interoperability and standardization: a study of methodology part i: achieving interoperability at the schema level. *D-Lib Magazine*, v. 12, n. 6, june 2006. Disponível em: <http://www.dlib.org/dlib/june06/chan/06chan.html>. Acesso em: 23/05/2019.

FOMERTON, Danilo. *Identificação de padrões de metadados de preservação digital*. Dissertação de Mestrado em Tecnologia, Informação e Sociedade: Universidade Federal de São Carlos - Ufscar – São Carlos, 2015.

GRÁCIO, J. C. A. *Preservação digital na gestão da informação: um modelo processual para as instituições de ensino superior*. São Paulo, SP: Cultura Acadêmica, 2012. Programa de Pós-Graduação em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. Disponível em: <http://www.HYPERLINK> “[http://www.culturaacademica.com.br/\\_img/arquivos/Preservacao\\_digital\\_na\\_gestao\\_da\\_inform%20acao-WEB\\_v2.pdf](http://www.culturaacademica.com.br/_img/arquivos/Preservacao_digital_na_gestao_da_inform%20acao-WEB_v2.pdf)” w.culturaacademica.com.br/\_img/arquivos/Preservacao\_digital\_na\_gestao\_da\_inform HYPERLINK “[http://www.culturaacademica.com.br/\\_img/arquivos/Preservacao\\_digital\\_na\\_gestao\\_da\\_inform%20acao-WEB\\_v2.pdf](http://www.culturaacademica.com.br/_img/arquivos/Preservacao_digital_na_gestao_da_inform%20acao-WEB_v2.pdf)” acaoHYPERLINK “[http://www.culturaacademica.com.br/\\_img/arquivos/Preservacao\\_digital\\_na\\_gestao\\_da\\_inform%20acao-WEB\\_v2.pdf](http://www.culturaacademica.com.br/_img/arquivos/Preservacao_digital_na_gestao_da_inform%20acao-WEB_v2.pdf)”-WEB\_v2.pdf. Acesso em: 26/05/2019.

HEDSTROM, M. Digital preservation: a time bomb for digital libraries. *Computers and the Humanities*, Michigan, U.S.A, n. 31, p. 189–202, 1998. Disponível em: <http://deepblue.lib.umich.edu/bitstream/handle/2027.42/42573?sequence=1>. Acesso em: 24/05/2019.

MÁRDERO ARELLANO, M. Á. Preservação de documentos digitais. *Ciência da Informação*, Brasília, DF, v. 33, n. 2, p. 15-27, maio/ago. 2004. Disponível em: <http://www.brapci.ufpr.br/download.php?dd0=9026>. Acesso em: 24/05/2017.

MARTÍNEZ ÁVILA, D. *et al.* CLASSIFYING THE LOD CLOUD. *Brazilian Journal of Information Science: research trends*, v. 12, n. 4, p. 06-10, 12 dez. 2018. Disponível em: <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/8328>. Acesso em: 29/05/2019

NATIONAL INFORMATION STANDARD ORGANIZATION. *Understanding Metadata*. Bethesda, MD: NISO Press, 2004. Disponível em: <https://www.niso.org/publications/understanding-metadata-2017>. Acesso em: 12/11/2018.

RAUTENBERG, S.; CARMO, P. R. BIG DATA E CIÊNCIA DE DADOS. *Brazilian Journal of Information Science: research trends*, v. 13, n. 1, p. 56-67, 29 mar. 2019. Disponível em: <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/8315>. Acesso em: 29/05/2015.

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 15, n. 30, p. 1-31, 2010. Disponível em: <http://www.brapci.ufpr.br/download.php?dd0=12621>. Acesso em: 18/05/2019.

THOMAZ, K. P. *A preservação de documentos eletrônicos de caráter arquivístico: novos desafios, velhos problemas*. 2004. 388 f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004. Disponível em: <http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/VALA-68ZRKF>. Acesso em: 11/01/2019.