



# 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-8 – Informação e Tecnologia

#### REQUISITOS FUNCIONAIS PARA REPOSITÓRIOS DE DADOS: SUBSÍDIOS TEÓRICOS

#### *FUNCTIONAL REQUIREMENTS FOR DATA REPOSITORIES: THEORETICAL SUBSIDIES*

**Lucas Henrique Alves da Silva** – Universidade Federal da Paraíba (UFPB)

**Izabel França de Lima** – Universidade Federal da Paraíba (UFPB)

#### **Modalidade: Resumo Expandido**

**Resumo:** Objetiva-se identificar os requisitos para construção, manutenção e avaliação de repositórios de dados científicos, partindo-se do seguinte problema de pesquisa: quais são os elementos envolvidos no funcionamento de repositórios de dados científicos? Caracteriza-se como uma pesquisa de abordagem qualitativa e tipo descritivo, visto que busca identificar, classificar e analisar os requisitos funcionais para repositórios de dados. Por meio de pesquisa bibliográfica, selecionou-se um corpus documental composto por seis publicações sobre requisitos para repositórios digitais. O estudo realizado apontou para um conjunto de requisitos que devem ser considerados na tomada de decisão na escolha de softwares, na análise de repositórios de dados ou na avaliação dessas plataformas, para fins de implementação ou aperfeiçoamento desses sistemas.

**Palavras-Chave:** repositórios de dados; dados de pesquisa; requisitos funcionais.

**Abstract:** Aims to identify the requirements for construction, maintenance and evaluation of research data repositories, starting from the following research problem: what are the elements involved in the functioning of scientific data repositories? It is characterized as a qualitative and descriptive research, as it intends to identify, classify and analyze the functional requirements for data repositories. Through bibliographical research, a documental corpus composed of six publications on requirements for digital repositories. The study carried out pointed to a set of requirements that should be considered in decision making when choosing software, analyzing data repositories or evaluating these platforms, for purposes of implementation or improvement these systems.

**Keywords:** research data repositories; research data; functional requirements.

## 1 INTRODUÇÃO

A pesquisa científica tem como uma de suas principais características a centralidade dos dados nas engrenagens da produção de conhecimento. No contexto científico contemporâneo, marcado pela colaboração entre pesquisadores e pela ambiência de tecnologias cada vez mais avançadas de armazenamento e disseminação da informação, os dados oriundos do processo investigativo têm adquirido maior visibilidade dentro da comunidade acadêmica. Dentre os motivos para essa visibilização está o compartilhamento dos conjuntos de dados originais das pesquisas em repositórios de dados científicos.

Do ponto de vista da Ciência da Informação, os repositórios de dados são sistemas de informação com funções gerenciais e técnicas que contribuem com o compartilhamento, o arquivamento, a representação, a disponibilização, a preservação, a descoberta, o acesso, a reutilização e a citação dos dados científicos, além de proporcionar a transparência e a validação dos resultados da pesquisa científica (MONTEIRO; SANT'ANA, 2016; PAMPEL; KINDLING, 2017; SANTOS *et al.*, 2017; SAYÃO; SALES, 2018).

Os repositórios de dados científicos emergem, portanto, como plataformas digitais que possuem entre seus objetivos principais a custódia e a disseminação de conjuntos de dados provenientes da pesquisa acadêmico-científica. Essas plataformas funcionam como base tecnológica para a consecução das atividades de gestão de dados científicos, compondo, portanto, o que se tem denominado de ciberinfraestrutura da pesquisa científica (WALTERS, 2014; BORGMAN, 2015; SAYÃO; SALES, 2018).

Contudo, o empreendimento da construção de repositórios de dados é multifacetado quanto às tecnologias, aos desafios gerenciais e às práticas disciplinares de coleta, compartilhamento e reuso de dados, que variam significativamente conforme a área do conhecimento (NEWTON; MILLER; BRACKE, 2011).

Considerando que os repositórios de dados são plataformas digitais que integram os sistemas de gestão de dados científicos e que funcionam como suporte aos processos relativos ao ciclo de vida dos dados de pesquisa, levanta-se o seguinte problema de pesquisa: quais são os elementos envolvidos no funcionamento de repositórios de dados científicos? Em vista desse questionamento, este trabalho tem como objetivo identificar os requisitos necessários para o funcionamento efetivo desses ambientes digitais. Para tanto, foram realizadas buscas por contribuições de cunho teórico e prático na literatura técnico-científica da Ciência da

Informação e áreas correlatas, no intuito de realizar um levantamento de requisitos funcionais para repositórios de dados de pesquisa.

Destarte, este estudo parte do pressuposto de que os repositórios de dados constituem um tipo particular de repositório digital, reivindicando abordagens investigativas voltadas para as características e necessidades específicas desse tipo de repositório.

## 2 METODOLOGIA

Este trabalho é derivado de um projeto de pesquisa de mestrado do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Paraíba (PPGCI/UFPB) que tem entre suas finalidades identificar os requisitos gerais e específicos de construção e funcionamento de repositórios de dados de pesquisa.

A pesquisa emprega uma abordagem de cunho qualitativo; em relação aos seus fins, caracteriza-se como uma investigação de tipo descritiva, visto que busca apontar e classificar os requisitos funcionais para repositórios de dados, visando oferecer um panorama da constituição e do funcionamento desses sistemas.

Os requisitos foram levantados por meio de pesquisa bibliográfica realizada no Google Scholar e em bases de dados de abrangência internacional contempladas pelo Portal de Periódicos da Capes, empregando-se como estratégias de busca “*research data repositories*” e “*software*” com uso do operador *booleano* “AND”. Os requisitos foram, portanto, extraídos da literatura por meio da análise documental de cinco (5) artigos de periódicos e um (1) capítulo de livro, com data de publicação compreendida entre 2009 e 2019. O método empregado para a análise dos dados foi a análise de conteúdo, cujas diretrizes nortearam a exploração do material textual, seguida da codificação das unidades de conteúdo encontradas nos documentos analisados e, por fim, a categorização dos requisitos identificados no *corpus* documental, dentro das categorias analíticas estabelecidas ao longo do processo.

## 3 REQUISITOS FUNCIONAIS

Os repositórios de dados de pesquisa são plataformas digitais cujas especificidades técnicas, funcionais e gerenciais devem ser consideradas nos processos de construção, manutenção e aperfeiçoamento desses ambientes informacionais.

Em virtude das particularidades relativas à natureza, ao formato e às características dos conjuntos de dados, e também devido às novas dinâmicas da comunicação da informação

científica e dos dados de pesquisa, surge a necessidade de organizar um corpo de conhecimento mais estruturado e elucidativo em relação aos repositórios de dados enquanto plataformas digitais integrantes dos sistemas de gestão de dados de pesquisa.

A análise do *corpus* documental da presente pesquisa revelou uma conformidade dos autores em relação à imprescindibilidade de determinados requisitos, embora os documentos analisados diferenciem-se quanto à sua abordagem teórico-prática. O Quadro 1 evidencia as diferentes perspectivas teóricas das publicações constituintes do *corpus*.

**Quadro 1 – Abordagens teóricas sobre requisitos para repositórios**

<b>Autores(as)</b>	<b>Abordagem</b>
Sayão e Marcondes (2009)	1. Requisitos para avaliação de pacotes de software
Pyrounakis, Nikolaidou e Hatzopoulos (2014)	2. Características de repositórios digitais
Sayão e Sales (2018)	3. Requisitos para avaliação de sistemas de gestão de dados
Martins, Silva e Siqueira (2018)	4. Sistemas para criação de acervos digitais
Kim (2018)	5. Requisitos funcionais para plataformas de repositórios de dados
Morales-Vargas e Codina (2019)	6. Atributos de qualidade para repositórios de dados científicos

Fonte: Elaboração própria.

Os enfoques teóricos relacionados no Quadro 1 nortearam a forma como seus respectivos autores formularam os requisitos que vieram a constituir o objeto de interesse principal da presente pesquisa. As abordagens 1, 2 e 4 foram originalmente pensadas para repositórios digitais de uma maneira geral, ao passo que as abordagens 3, 5 e 6 são especificamente voltadas para o caso dos repositórios de dados científicos. Não obstante, todas as seis propostas são aplicáveis aos repositórios de dados. O Quadro 2 sumariza os requisitos identificados nos seis textos analisados.

**Quadro 2 – Requisitos funcionais para repositórios de dados científicos**

<b>Autores(as)</b>	<b>Requisitos funcionais</b>
Sayão e Marcondes (2009)	1. Características gerais para avaliação de pacotes de software; 2. Distribuição do pacote de software; 3. Enquadramento tecnológico e institucional; 4. Serviços e funcionalidades; 5. Conformidade com padrões; 6. Gestão do repositório.
Pyrounakis, Nikolaidou e Hatzopoulos (2014)	1. Modelo de objeto digital; 2. Hierarquia entre coleções; 3. Metadados; 4. Indexação, pesquisa e navegação; 5. Gerenciamento de objetos; 6. Interfaces do usuário; 7. Controle

	de acesso; <b>8.</b> Idiomas; <b>9.</b> Interoperabilidade; <b>10.</b> Customização; <b>11.</b> Serviços estendidos; <b>12.</b> Preservação; <b>13.</b> Suporte comunitário; <b>14.</b> Hospedagem.
Sayão e Sales (2018)	<b>1.</b> Serviços; <b>2.</b> Interoperabilidade; <b>3.</b> Política do repositório; <b>4.</b> Institucionalização; <b>5.</b> Reconhecimento comunitário; <b>6.</b> Persistência; <b>7.</b> Visibilidade; <b>8.</b> Presença em diretórios; <b>9.</b> Licenças; <b>10.</b> Certificação; <b>11.</b> Curadoria digital; <b>12.</b> Preservação; <b>13.</b> Custo.
Martins, Silva e Siqueira (2018)	<b>1.</b> Sistema de busca; <b>2.</b> Sistema de navegação; <b>3.</b> Sistema de organização; <b>4.</b> Sistema de representação; <b>5.</b> Sistema de preservação digital; <b>6.</b> Sistema de interoperabilidade; <b>7.</b> Sistema de administração; <b>8.</b> Sistema de colaboração social; <b>9.</b> Sistema de manutenção; <b>10.</b> Características gerais.
Kim (2018)	<b>1.</b> Metadados; <b>2.</b> Identificação persistente; <b>3.</b> Autenticação e autorização; <b>4.</b> Acesso aos dados; <b>5.</b> Políticas; <b>6.</b> Publicação; <b>7.</b> Submissão, ingestão e administração; <b>8.</b> Organização das coleções; <b>9.</b> Localização; <b>10.</b> Interoperabilidade; <b>11.</b> Preservação e sustentabilidade; <b>12.</b> Interface do usuário; <b>13.</b> Qualidade dos dados.
Morales-Vargas e Codina (2019)	<b>1.</b> Interface do usuário; <b>2.</b> Recuperação; <b>3.</b> Acesso e reuso; <b>4.</b> Conservação; <b>5.</b> Disseminação; <b>6.</b> Avaliação; <b>7.</b> Compromisso institucional.

Fonte: Elaboração própria.

A abordagem de Sayão e Marcondes (2009), voltada para a seleção de softwares para a implementação de repositórios institucionais, considera pelo menos seis (6) amplas categorias de requisitos a serem observados: 1) Características gerais para avaliação de pacotes de software, que abrange requisitos como facilidade de implantação, índice de sucesso, integração com outras ferramentas, suporte técnico, limitações do sistema e documentação disponível; 2) Distribuição do pacote de software, em que se deve optar por um software proprietário ou de código aberto; 3) Enquadramento tecnológico e institucional, que abrange os requisitos de hardware, software, conhecimento técnico, sustentabilidade e conformidade com a política de tecnologia da instituição; 4) Serviços e funcionalidades, que contempla os requisitos de interface, entrada de dados, arquitetura da informação, indexação, recuperação da informação e interatividade entre usuários; 5) Conformidade com padrões, categoria em que se enquadram requisitos como metadados, preservação digital, interoperabilidade, identificação persistente, importação e exportação de dados; e 6) Gestão do repositório, que inclui gestão das coleções, estatísticas, gestão do usuário final e segurança da informação.

Cumpramos ressaltar que esses requisitos foram formulados com um olhar direcionado para a avaliação e eleição de um pacote de software para a implementação de repositórios em

contextos institucionais determinados e para a tomada de decisões gerenciais e técnicas referentes à construção desses ambientes digitais.

Pyrounakis, Nikolaidou e Hatzopoulos (2014), por sua vez, elaboram seu constructo em torno da criação de repositórios que hospedam coleções de objetos digitais. Nessa abordagem, são elencados os recursos e as características consideradas essenciais em um software de repositório moderno. Nessa contribuição, são identificados os quatorze (14) seguintes requisitos: 1) Modelo de objeto digital, referente à estrutura do objeto, que integra metadados, conteúdo e relacionamentos com outros objetos; 2) Hierarquia entre coleções; 3) Metadados; 4) Indexação, pesquisa e navegação; 5) Gerenciamento de objetos, que considera os fluxos de trabalho para envio de objetos de digitais e as interfaces exibidas na criação, atualização e exclusão de conteúdos e metadados; 6) Interfaces do usuário; 7) Controle de acesso; 8) Idiomas; 9) Interoperabilidade; 10) Customização; 11) Serviços estendidos, como *plug-ins* e outros complementos; 12) Preservação, que abarca funções como *backup*, replicação, migração e versionamento; 13) Suporte comunitário, referente à existência de uma comunidade de especialistas e usuários que possa assegurar a possibilidade de assistência técnica na instalação e nas eventuais atualizações do repositório; 14) e Hospedagem, a qual envolverá custos mensais ou anuais, além de alternativas como armazenamento de dados na nuvem.

À semelhança dos requisitos propostos por Sayão e Marcondes (2009), os critérios levantados por Pyrounakis, Nikolaidou e Hatzopoulos (2014) são majoritariamente relacionados com os atributos oferecidos pela tecnologia empregada, não obstante os requisitos que são mais atrelados à gestão do repositório, que dizem respeito predominantemente às competências e atribuições da equipe institucional incumbida das atividades rotineiras de administração, alimentação e manutenção do repositório.

Na abordagem de Sayão e Sales (2018), prevalece o enfoque nos sistemas de gestão de dados de pesquisa como um todo, o que não se restringe às funções dos repositórios de dados, abrangendo, portanto, variáveis externas a essas plataformas digitais. Assim, os requisitos propostos pelos autores para a avaliação desses sistemas podem ser agrupados em treze (13) categorias, a saber: 1) Serviços, categoria que inclui serviços de referência, desenvolvimento de coleções, competência informacional, identificação persistente, citação, controle de versões, publicação, representação, reformatação, segurança e análise, todos esses serviços voltados para o tratamento dos conjuntos de dados que são fornecidos pelos

pesquisadores; 2) Interoperabilidade; 3) Política do repositório; 4) Institucionalização; 5) Reconhecimento comunitário; 6) Persistência; 7) Visibilidade; 8) Presença em diretórios; 9) Licenças; 10) Certificação; 11) Curadoria digital; 12) Preservação; e 13) Custo.

Nesse último modelo, observa-se uma significativa influência de fatores gerenciais e institucionais, ou seja, processos relativos a elementos externos aos repositórios. Essa distinção em relação aos outros modelos teóricos supramencionados pode ser explicada em virtude da visão dos autores, que consideram os repositórios de dados como componente integrante de um sistema de gestão de dados mais abrangente, e não como ferramentas autogeridas.

Em Martins, Silva e Siqueira (2018), são propostas dez (10) dimensões analíticas para análise de softwares para criação de acervos digitais. As dimensões são as seguintes: 1) Sistema de busca; 2) Sistema de navegação; 3) Sistema de organização; 4) Sistema de representação; 5) Sistema de preservação digital; 6) Sistema de interoperabilidade; 7) Sistema de administração; 8) Sistema de colaboração social; 9) Sistema de manutenção; e 10) Características gerais, referentes a requisitos diversificados que não se enquadram nas demais categorias.

Constata-se, nesse conjunto de dimensões, uma prevalência de requisitos associados à arquitetura da informação, embora os autores também considerem aspectos levantados pelos outros três modelos anteriormente mencionados, como preservação digital, interoperabilidade e suporte técnico. Cabe ressaltar, contudo, que essa contribuição teórica não é voltada para especificamente os repositórios de dados, mas sim para acervos digitais de um modo geral, ainda que os requisitos propostos sejam aplicáveis ao contexto dos repositórios de dados.

O modelo de Kim (2018), por sua vez, é voltado especificamente para plataformas de repositórios de dados científicos. O autor propõe um conjunto de treze (13) requisitos funcionais, são eles: 1) Metadados; 2) Identificação persistente; 3) Autenticação e autorização; 4) Acesso aos dados, o que inclui aspectos como mecanismos de pesquisa e download dos conjuntos de dados; 5) Políticas, especialmente em relação ao processamento dos dados; 6) Publicação, a qual inclui estatísticas de acesso aos dados e informações sobre citação; 7) Submissão, ingestão e administração; 8) Organização das coleções; 9) Localização, a qual remete à computação de alto desempenho; 10) Interoperabilidade; 11) Preservação e sustentabilidade; 12) Interface do usuário; e 13) Qualidade dos dados.

Já a abordagem de Morales-Vargas e Codina (2019) tem como objetivo a definição de atributos de qualidade para repositórios de dados científicos, ou seja, assim como a proposta anterior, também é direcionada para os repositórios de dados. Para esses autores, sete (7) categorias de requisitos devem ser considerados na análise da qualidade de repositórios de dados, os quais são: 1) Interface do usuário, a qual abrange elementos alusivos à arquitetura da informação; 2) Recuperação, a qual abarca atributos como metadados, identificação persistente, citação e descoberta; 3) Acesso e reuso; 4) Conservação, a qual engloba aspectos como armazenamento, preservação e versionamento dos conjuntos de dados; 5) Disseminação, que remete aos mecanismos de busca e à divulgação do repositório; 6) Avaliação, referente à curadoria, certificação e métricas; e 7) Compromisso institucional.

Como é possível constatar, as duas últimas abordagens são especificamente voltadas para a construção, análise ou avaliação de repositórios de dados de pesquisa, o que as permitiu pontuar aspectos próprios desse tipo de repositório. De um modo geral, observa-se que os requisitos funcionais para repositórios de dados envolvem aspectos técnicos, gerenciais, institucionais, políticos, financeiros, entre outros.

#### **4 CONSIDERAÇÕES FINAIS**

Os modelos teóricos discutidos neste trabalho foram postos em evidência com o objetivo de oferecer subsídios teóricos sobre os repositórios de dados de pesquisa, especialmente no que diz respeito aos requisitos funcionais para esse tipo de plataforma digital. Embora as abordagens teóricas referenciadas manifestem diferenças em relação ao modo como enxergam os repositórios digitais, e nem todas tenham originalmente pretendido tratar sobre repositórios de dados, a análise dos dados da pesquisa em andamento revelou a convergência das seis perspectivas abordadas em torno de alguns requisitos em comum que parecem se instituir como universais, como metadados, acesso, interoperabilidade, preservação, interface do usuário e arquitetura da informação.

Portanto, o objetivo principal da pesquisa foi alcançado, visto que foram identificados na literatura científica um conjunto de requisitos pertinentes aos repositórios de dados, que devem ser considerados na tomada de decisão na escolha de softwares, na análise de repositórios de dados ou simplesmente na avaliação dessas plataformas, para fins de implementação ou aperfeiçoamento. Para estudos posteriores, sugere-se aprofundar a análise sobre cada um desses requisitos, no intuito de gerar conhecimentos mais detalhados

acerca dos mecanismos que constituem os repositórios de dados e dos processos que determinam ou influenciam sua criação e seu funcionamento.

## REFERÊNCIAS

BORGMAN, Christine L. **Big data, little data, no data: scholarship in the networked world**. Cambridge: The MIT Press, 2015.

KIM, Suntae. Functional requirements for research data repositories. **International Journal of Knowledge Content Development and Technology**, [S. l.], v. 8, n. 1, p. 25-36, mar. 2018.

MARTINS, Dalton Lopes; SILVA, Marcel Ferrante; SIQUEIRA, Joyce. Comparação entre sistemas para criação de acervos digitais: análise dos softwares livres DSpace, EPrints, Fedora, Greenstone e Islandora a partir de novas dimensões analíticas. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 9, n. 1, p. 52-71, mar./ago. 2018. Disponível em: <https://www.revistas.usp.br/incid/article/view/134333>. Acesso em: 28 mar. 2021.

MONTEIRO, Elizabete Cristina de Souza de Aguiar; SANT'ANA, Ricardo César Gonçalves. Repositório de dados científicos nas universidades brasileiras e portuguesas. *In: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO*, 6., 2016, Londrina. **Anais [...]**. Londrina: UEL, 2016. p. 652-664.

MORALES-VARGAS, Alejandro; CODINA, Lluís. Atributos de calidad web para repositorios de datos de investigación en universidades. **Hipertext.net**, Barcelona, n. 19, p. 49-62, 2019.

NEWTON, Mark P.; MILLER, C. C.; BRACKE, Marianne Stowell. Librarian roles in institutional repository data set collecting: outcomes of a research library task force. **Collection Management**, [S. l.], v. 36, n. 1, p. 53-67, 2011.

PAMPPEL, Heinz; KINDLING, Maxi. Informationsinfrastrukturangebote für digitale Forschungsdaten. **Peter Schirmbacher sei Dank – E(hren)Journal**, Berlim, p. 15-33, 2017.

PYROUNAKIS, George; NIKOLAIDOU, Mara; HATZOPOULOS, Michael. Building digital collections using open source digital repository software: a comparative study. **International Journal of Digital Library Systems**, [S. l.], v. 4, n. 1, p. 10-24, jan./mar. 2014.

SANTOS, Paula Xavier; ALMEIDA, Bethânia de Araújo; ELIAS, Flávia; MOTTA, Márcia Luz da; GUANAES, Paulo; JORGE, Vanessa de Arruda; HENNING, Patrícia; OLIVEIRA, Gabriela. **Livro Verde – Ciência Aberta e dados abertos: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional**. Rio de Janeiro: Fiocruz, 2017.

SAYÃO, Luis Fernando; MARCONDES, Carlos Henrique. Software livre para repositórios institucionais: alguns subsídios para a seleção. *In: SAYÃO, Luis Fernando; TOUTAIN, Lúcia Brandão; ROSA, Flávia Garcia; MARCONDES, Carlos Henrique (org.). **Implantação e gestão de***

**repositórios institucionais:** políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p. 23-54.

SAYÃO, Luis Fernando; SALES, Luana Farias. Subsídios para a construção de um modelo de avaliação de sistemas de gestão de dados de pesquisa. **PontodeAcesso**, Salvador, v. 12, n. 3, p. 80-108, dez. 2018. Disponível em: <https://periodicos.ufba.br/index.php/revistaici/article/view/28965>. Acesso em: 28 jan. 2021.

WALTERS, Tyler. Assimilating digital repositories into the active research process. *In*: RAY, Joyce M. (ed.). **Research data management: practical strategies for information professionals**. West Lafayette: Purdue University Press, 2014.