UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMUNICAÇÃO E INFORMAÇÃO

Daiane Barrili dos Santos

REPOSITÓRIOS DE DADOS DE PESQUISA: IDENTIFICAÇÃO DOS CRITÉRIOS/REQUISITOS INTERNACIONAIS DE AVALIAÇÃO DA CONFIABILIDADE

Porto Alegre
2022

Daiane Barrili dos Santos

REPOSITÓRIOS DE DADOS DE PESQUISA: IDENTIFICAÇÃO DOS CRITÉRIOS/REQUISITOS INTERNACIONAIS DE AVALIAÇÃO DA CONFIABILIDADE

Tese apresentada ao Programa de Pós-Graduação em Comunicação e Informação da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do grau de Doutora em Comunicação e Informação.

Orientador: Profª Drª Samile Andrea de Souza Vanz

Porto Alegre

2022

CIP - Catalogação na Publicação

Barrili dos Santos, Daiane
REPOSITÓRIOS DE DADOS DE PESQUISA: IDENTIFICAÇÃO DOS CRITÉRIOS/REQUISITOS INTERNACIONAIS DE AVALIAÇÃO DA CONFIABILIDADE / Daiane Barrili dos Santos. -- 2022.
188 f.
Orientadora: Samile Andrea de Souza Vanz.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Biblioteconomia e Comunicação, Programa de Pós-Graduação em Ciência da Informação, Porto Alegre, BR-RS, 2022.

1. Repositórios de dados de pesquisa. 2. Avaliação de repositórios de dados de pesquisa. 3. Avaliação da confiabilidade. 4. Critérios/Requisitos. I. Andrea de Souza Vanz, Samile, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

PROGRAMA DE PÓS-GRADUAÇÃO EM COMUNICAÇÃO E INFORMAÇÃO
Faculdade de Biblioteconomia e comunicação – Campus Saúde
Rua Ramiro Barcelos, 2705
Porto Alegre – RS
CEP: 90035-007
Tel: (51) 3308 5116
E-mail: ppgcom@ufrgs.br

DAIANE BARRILI DOS SANTOS

REPOSITÓRIOS DE DADOS DE PESQUISA:
IDENTIFICAÇÃO DOS CRITÉRIOS/REQUISITOS INTERNACIONAIS DE AVALIAÇÃO DA CONFIABILIDADE

Tese apresentada ao Programa de Pós-graduação em Comunicação e Informação da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Doutora em Comunicação e Informação.

Aprovada em: _____.

BANCA EXAMINADORA:

Profa. Dra. Carolina Howard Felicíssimo
Rede Nacional de Ensino e Pesquisa (RNP)

Profa. Dra. Sônia Elisa Caregnato
Universidade Federal do Rio Grande do Sul

Prof. Dr. Rafael Port da Rocha
Universidade Federal do Rio Grande do Sul

Prof. Dr. Rene Faustino Gabriel Junior
Universidade Federal do Rio Grande do Sul

Profa. Dra. Caterina Marta Groposo Pavão (Suplente)
Universidade Federal do Rio Grande do Sul

DEDICATÓRIA

À Nossa Senhora do Trabalho.

AGRADECIMENTOS

À Universidade Federal do Rio Grande do Sul, pelo o ensino e experiências vividas.

À minha orientadora Profa. Dra. Samile Andrea de Souza Vanz, pelo apoio intelectual e conceitual através de críticas construtivas e sugestões e, principalmente, pela parceria, compreensão nos momentos de dificuldades e amizade.

À banca de qualificação do projeto de tese, Profa. Dra. Sônia Elisa Caregnato e Prof. Dr. Rene Faustino Gabriel Junior, pelas contribuições.

À banca de defesa, Dra. Carolina Howard Felicíssimo, Profa. Dra. Sônia Elisa Caregnato, Prof. Dr. Rafael Port da Rocha, Prof. Dr. Rene Faustino Gabriel Junior, e Prof. Dra. Caterina Marta Groposo Pavão (suplente), pela disponibilidade de avaliação, contribuindo com a qualidade da tese.

Aos colegas do Grupo de Pesquisa "Comunicação Científica" da UFRGS pela troca e compartilhamento de informações e conhecimentos nos encontros presenciais e online (durante a pandemia).

Às profas. Samile Andrea de Souza Vanz e Sônia Elisa Caregnato por terem me dado a oportunidade de cursar o doutorado, por terem me dado a chance de realizar um dos meus maiores sonhos (me tornar doutora antes dos meus 40 anos) e antes de me tornar mãe.

À minha família e amigos, pelo carinho, incentivo, palavras de estímulo e compreensão.

À Deus, por me dar forças e saúde em um momento tão difícil que o mundo está vivendo (pandemia da Covid-19), pelas graças alcançadas, e pela vida.

RESUMO

Baseado nos critérios/requisitos de avaliação da confiabilidade de repositórios de dados de pesquisa identificados nos instrumentos de avaliações e certificações conceituados internacionalmente, este estudo tem como objetivo elencar as características de um repositório confiável. Para que fosse possível verificar quais critérios/requisitos deveriam estar presentes em uma avaliação da confiabilidade de um repositório de dados de pesquisa brasileiro, foi necessário identificar na literatura referente à área e nos documentos institucionais, os principais critérios/requisitos utilizados em avaliações de organizações renomadas. Este estudo fundamenta-se em uma pesquisa bibliográfica e documental, a qual forneceu embasamento teórico e elementos coerentes para o entendimento sobre o estado atual em que se encontra o tema. O material bibliográfico foi levantado no âmbito da Ciência da Informação e serviu para dar suporte teórico ao desenvolvimento deste estudo. A principal técnica de pesquisa envolvida foi a pesquisa documental, que tem por objetivo colher evidências que apoiem o estudo, coletar informações que assegurem os resultados conforme os objetivos traçados para a pesquisa, utilizando como base os documentos existentes. Na pesquisa documental, houve a identificação dos elementos importantes para a proposição de um conjunto único de critérios/requisitos para serem utilizados por repositórios de dados de pesquisa brasileiros em autoavaliações de confiabilidade, foram utilizados documentos e relatórios sugeridos por especialistas. Utilizou-se o *CoreTrustSeal - Trustworthy Data Repositories Requirements* 2020–2022; o ACTDR - *Audit and Certification of Trustworthy Digital Repositories*; os Principles FAIR; e *The TRUST Principles for digital repositories*. Concomitante a análise dos instrumentos ocorreu a categorização dos critérios/requisitos para avaliação de confiabilidade. Após serem contrastadas as listas, os critérios/requisitos equivalentes e aqueles constatados importantes compuseram o conjunto de critérios/requisitos. O processo contínuo de autoavaliações proporciona diversos benefícios contribuindo para a construção de uma base organizacional sólida para submeter-se futuramente às certificações e auditorias. Esta pesquisa não teve como propósito a substituição das leituras dos documentos publicados pelas organizações, apenas atuará de modo complementar para facilitar a compreensão e instigar a reflexão auxiliando nos processos de planejamento e implementação tendo em vista o preparo para futura certificação, ou mesmo sem uma certificação formal, os requisitos contemplados no estudo podem ser usados como referência e para identificar as lacunas e áreas que precisam de maior atenção. Conclui-se que os critérios/requisitos selecionados neste estudo apresentam as características almejadas para um repositório de dados confiável e que seriam pertinentes ao contexto brasileiro, ou seja, devido ao estado incipiente em que se encontram os repositórios de dados de pesquisa brasileiros, deve-se considerar os requisitos que se encontram em instrumentos que abrangem a certificação básica. O conjunto único de critérios/requisitos, elaborado a partir dos instrumentos internacionais selecionados, poderá apoiar os repositórios de dados de pesquisa brasileiros nas autoavaliações e na melhoria geral, tornando-os confiáveis e direcionando-os para certificação.

Palavras-chave: Repositórios de dados de pesquisa. Avaliação de repositórios de dados de pesquisa. Avaliação da confiabilidade. Critérios/Requisitos.

ABSTRACT

Based on the criteria/requirements for evaluating the reliability of research data repositories identified in internationally renowned assessment and certification instruments, this study aims to list the characteristics of a reliable repository. In order to verify which criteria/requirements should be present in an assessment of the reliability of a Brazilian research data repository, it was necessary to identify in the literature referring to the area and in the institutional documents, the main criteria/requirements used in evaluations of renowned organizations . This study is based on a bibliographical and documentary research, which provided theoretical basis and coherent elements for the understanding of the current state of the theme. The bibliographic material was collected within the scope of Information Science and served to provide theoretical support for the development of this study. The main research technique involved was the documentary research, which aims to collect evidence to support the study, collect information that ensure the results according to the objectives outlined for the research, using existing documents as a basis. In the document analysis, where the important elements were identified for the proposition of a single set of criteria/requirements to be used by Brazilian research data repositories in self-reliability assessments, documents and reports suggested by experts were used. CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022 was used; the ACTDR - Audit and Certification of Trustworthy Digital Repositories; the FAIR Principles; and The TRUST Principles for digital repositories. Concomitantly with the analysis of the instruments, the categorization of reliability criteria took place. After the lists were contrasted, the equivalent criteria/requirements and those found to be important made up the set of criteria/requirements. The continuous process of self-assessments provides several benefits, contributing to the construction of a solid organizational base to be submitted to certifications and audits in the future. This research was not intended to replace the readings of documents published by organizations, it will only act in a complementary way to facilitate understanding and instigate reflection, helping in the planning and implementation processes with a view to preparing for future certification, or even without a certification. formally, the requirements contemplated in the study can be used as a reference and to identify gaps and areas that need further attention. It is concluded that the criteria/requirements selected in this study reflect the desirable characteristics of a reliable data repository that would be suitable for the Brazilian context, that is, due to the incipient state in which Brazilian research data repositories are, it should be considering the requirements found in instruments covering basic certification. The unique set of criteria/requirements, developed from selected international instruments, can support Brazilian research data repositories in self-assessments and general improvement, directing them to the path of certification, making them reliable.

Keywords: Research data repositories. Evaluation of research data repositories. Reliability assessment.Criteria/Requirements.

LISTA DE ILUSTRAÇÕES

LISTA DE ABREVIATURAS E SIGLAS

AADP	Acesso Aberto a Dados de Pesquisa
ABC	Academia Brasileira de Ciências
ABNT	Associação Brasileira de Normas Técnicas
ACTDR	Audit and Certification of Trustworthy Digital Repositories
AEI	Atributos da Encontrabilidade da Informação
AIP	Archival Information Package
BDC	Base de Dados Científicos
BDEP	Banco de Dados de Exploração e Produção
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
C3SL	Centro de Computação Científica e Software Livre
CCSDS	Consultative Committee for Space Data Systems
CMMI	Capability Maturity Model Integration
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
COAR	Confederation of Open Access Repositories
CRIS	Current Research Information Systems
DANS	Data Archiving and Networked Services
DCC	Data Curation Centre
DIN	Deutsches Institut für Normung
DSA	Data Seal of Approval
EI	Econtrabilidade da Informação
FAIR	Findable, Accessible, Interoperable and Reusable
FORCE11	The Future of Research Communications and e-Scholarship
FURG	Universidade Federal do Rio Grande
GLOBE	Global Collaboration Engine
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
ICPSR	Inter-Univertity Consortium for Political and Social Research
IODP	International Ocean Discovery Program
ISC	International Science Council
ISO	International Standards Organization
ITA	Instituto Tecnológico de Aeronáutica
LISA	Library and Information Science Abstract
MCTIC	Ministério da Ciência, Tecnologia, Inovações e Comunicações
MoU	Memorandum of Understanding
NDSA	National Digital Stewardship Alliance
NIST	National Institute of Standards and Technology
OAIS	Open Archival Information System
OCLC	Online Computer Library Center
OECD	Organization for Economic Co-operation and Development
OGC	Open Geospatial Consortium
OPENAIRE	Open Access Infrastructure for Research in Europe
OWL	Ontology Web Language
PDI	Preservation Description Information
RDA	Research Data Alliance

RDC	Research Data Canadá
RDF	Resource Description Framework
RENATA	Red Nacional Académica de Tecnologia Avanzada
RI	Repositórios Institucionais
RLG	Research Libraries Group
RNP	Rede Nacional de Ensino e Pesquisa
SAAI	Sistema Aberto de Arquivamento de Informação
SDI	Spatial Data Infrastructure
SEI/CMU	Software Engineering Institute/ Carnegie Mellon University
SiBi	Sistema de Bibliotecas
SIP	Submission Information Package
SKOS	Simple Knowledge Organization System
SNCTI	Sistema Nacional de Ciencia Tecnología e Innovación
SPARQL	Protocol and RDF Query Language
TDRs	Trusted Digital Repository
TRAC	Trustworthy Repository Audit & Certification
TRUST	Transparency, Responsibility, User focus, Sustainability and Technology
UFABC	Universidade Federal do ABC
UFPB	Universidade Federal da Paraíba
UFPR	Universidade Federal do Paraná
UFSCar	Universidade Federal de São Carlos
UNESP	Universidade Estadual Paulista
UNICAMP	Universidade Estadual de Campinas
UNIFESP	Universidade Federal de São Paulo
USP	Universidade de São Paulo
WDS	Word Data System

SUMÁRIO

1 INTRODUÇÃO

O advento das novas tecnologias da informação e comunicação vêm modificando o cenário atual, possibilitando o progresso da ciência e o desenvolvimento tecnológico. Novas concepções e práticas surgem e a informação é utilizada intensamente através de suportes tecnológicos que estão sendo desenvolvidos. Neste sentido, a ciência está se reconfigurando em relação aos métodos e novas formas de comunicar, utilizando-se de novas técnicas para compartilhamento e reuso de resultados de pesquisa. Isto posto, os dados de pesquisa, ou seja, os dados provenientes de pesquisas científicas, passaram a ser valorizados como ativos sendo reutilizados e disponibilizados gerando contribuições para o avanço científico e, com isso, novas competências associadas aos estudos com dados surgem para corroborar com a exploração de novos conhecimentos.

Nesta linha, as novas formas de publicação dão espaço aos repositórios digitais que, segundo Ribeiro e Vidotti (2009), são ferramentas que permitem a promoção científica e possibilitam disponibilizar os estudos científicos de maneira gratuita. Existem os repositórios temáticos ou especializados, os repositórios institucionais e, atualmente estão sendo desenvolvidos os repositórios de dados de pesquisa, para tornar acessíveis conjuntos de dados. De acordo com Misgar, Bhat e Wani (2020), a melhor maneira de disponibilizar dados de pesquisa é por meio de repositórios, que são serviços de banco de dados online que permitem gerenciar, armazenar e preservar a longo prazo. Dito isto, entende-se que um repositório de dados de pesquisa aumenta a transparência e a confiabilidade nos processos de pesquisa.

No nível mais básico, a definição de um repositório digital confiável deve começar com a missão de fornecer o acesso confiável e de longo prazo a recursos digitais gerenciados para a comunidade designada. Um repositório digital confiável compreenderá ameaças e riscos em seus sistemas. Monitoramento, planejamento e manutenção constantes, bem como ações e estratégias conscientes a implementação será exigida de repositórios para cumprir sua missão de preservação. Todos estes representam um empreendimento caro e complexo que os depositantes, partes interessadas, financiadores, a comunidade designada e outros repositórios digitais precisarão contar para ter um ambiente de preservação digital colaborativo. Comunicar os resultados da auditoria ao público irá gerar mais confiança e as auditorias objetivas adicionais, potencialmente levarão à certificação, promovendo ainda mais confiança no repositório e no sistema que o suporta. Finalmente, atingir o status de confiança não é uma conquista única, alcançada e esquecida. Para manter o status de confiança, um repositório

precisará realizar um ciclo regular de auditoria e/ou certificação (*CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011).

Como argumenta Pinheiro (2014, p. 159), os atuais avanços nas metodologias científicas, tendo os dados como elemento central, são resultados do aumento da complexidade das Ciências, "[...] com equipes mais numerosas trabalhando cada vez mais em colaboração, inclusive internacionalmente [...]. Portanto, não se trata apenas de um fenômeno tecnológico, mas científico e político, entre outras instâncias." A visibilidade que o acesso a dados vem obtendo, em todas as áreas, apresenta-se como novo desafio para áreas mais diretamente ligadas ao uso de Tecnologias da Informação e Comunicação e, neste cenário, à Ciência da Informação (CI). O conceito de dados aponta uma forte tendência de estudos e práticas profissionais que está se desenvolvendo no cerne das investigações realizadas por cientistas da informação. Assim, pode-se afirmar que as tendências de pesquisa em CI, no contexto da e-Science, mostram estudos sobre a proveniência, as melhores práticas, o planejamento, a curadoria entre outras atividades científicas relacionadas à teoria e à práxis com dados de pesquisa (SEMELER, 2017).

No contexto atual, percebe-se como as novas tecnologias estão mudando o futuro da ciência, fazendo com que a pesquisa seja mais colaborativa, transparente e aberta. Uma nova geração de experimentos e novos avanços científicos e tecnológicos estão sendo divulgados e publicados. Isso faz com que a pesquisa científica produza uma imensa quantidade de dados, o que torna a abertura dos dados uma necessidade.

Muitos países têm demonstrado interesse no que diz respeito a abertura de dados. Na Europa, por exemplo, estudos estão sendo desenvolvidos a fim de propor a construção de infraestruturas e políticas que contribuam para o compartilhamento. Pode-se citar o projeto intitulado '*Guidelines on Fair Data Management in Horizon 2020*' (EUROPEAN COMMISSION, 2020), que prevê uma ação piloto para o acesso a dados de pesquisa. Este documento ajuda os beneficiários do *Horizon 2020* a tornar seus dados de pesquisa localizáveis, acessíveis, interoperáveis e reutilizáveis, conforme os princípios FAIR - *Findable, Accessible, Interoperable and Reusable* - critérios para avaliar a qualidade dos conjuntos de dados e garantir que eles sejam gerenciados corretamente (EUROPEAN COMMISSION, 2014). O bom gerenciamento de dados de pesquisa não é um objetivo em si, mas o principal canal que leva à descoberta e inovação do conhecimento e à subsequente integração e reutilização de dados e conhecimento. Um órgão do Reino Unido para a

curadoria e preservação de dados de pesquisa, o *Data Curation Centre* (DCC)[1], lançado em 2004, ajuda a resolver os desafios de curadoria digital como também apoia os princípios FAIR. No website estão disponíveis documentos para pesquisadores britânicos referente a planos de gestão de dados e agências de fomento, entre outros.

Considerando a expansão das exigências quanto ao depósito, disponibilização e compartilhamento dos dados de pesquisa, se faz necessário uma infraestrutura que comporte os dados que estão sendo solicitados aos cientistas. Além disso, é imprescindível desenvolver políticas, regulamentar, instruir e treinar os pesquisadores para que façam escolhas mais assertivas sobre os repositórios confiáveis para o registro dos dados provenientes de suas pesquisas científicas.

Tendo em vista o reconhecimento dos benefícios, advindos do compartilhamento de dados, instituições governamentais passaram a estimular o Acesso Aberto a Dados de Pesquisa (AADP). Observa-se que a maior parte dessas inciativas são conduzidas por órgãos de fomento à atividade científica, desta forma, diversos órgãos financiadores passaram a solicitar de seus beneficiários adequações às práticas de AADP, que vão desde o requerimento de um plano de gestão de dados, associado ao projeto de pesquisa, até a definição do repositório no qual o pesquisador deve depositar seus dados (RAUEN, 2018).

Apesar de possuírem motivação comum, ações de AADP foram conduzidas de maneira distinta entre os países, expressando contornos mais centralizados em algumas experiências, como de países membros da União Europeia, e descentralizados, como no caso das políticas de AADP norte-americanas e australianas. As exigências quanto a disponibilização dos dados de pesquisa vêm sendo estabelecidas nesses países, até o momento, pelas agências de fomento.

A despeito disso, no Brasil, algumas inciativas já são observadas em instituições de pesquisa, universidades e agências de fomento como, por exemplo, na Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), que tem como requisito às solicitações de financiamento de projetos a obrigatoriedade do plano de gestão de dados. A Fundação Oswaldo Cruz (Fiocruz), já deu início a disponibilização de seus dados de forma aberta e definiu seu Plano de Dados Abertos (PDA), que apresenta as orientações para as ações de implementação e promoção da abertura de dados da Instituição. A Rede Nacional de Ensino e Pesquisa (RNP) também possui um Programa Nacional de AADP, uma iniciativa com o

[1] http://www.dcc.ac.uk/resources/data-management-plans

objetivo de promover e incentivar o compartilhamento de dados entre pesquisadores, conferindo maior eficiência na produção de conhecimento científico no Brasil (RNP, 2018).

É importante destacar que, em dezembro de 2019, aconteceu a 8ª Reunião de Acompanhamento do Compromisso 3, conhecido como "Compromisso pela Ciência Aberta", no Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC) cujo objetivo foi estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil. Na reunião, foi firmado o acordo para o repositório de dados Científicos do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), o LattesData, com a finalidade de disponibilizar os dados científicos de pesquisas financiadas pela instituição. Caracteriza-se por ser a expansão da Plataforma Lattes, com o propósito de armazenar e permitir acesso aos dados científicos provenientes dos projetos fomentados pelo CNPq, proporcionando o compartilhamento e reuso de dados pela comunidade científica (CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO, 2019).

Outro exemplo, é a Universidade Federal do Paraná (UFPR), que lançou, em janeiro de 2018, a Base de Dados Científicos (BDC) que visa reunir os dados utilizados em pesquisas publicadas pela comunidade da instituição. A iniciativa é resultado de uma parceria entre o Centro de Computação Científica e Software Livre (C3SL) e o Sistema de Bibliotecas (SiBi), ambos da UFPR.

Mesmo após a divulgação de diversas iniciativas, os pesquisadores questionam atualmente, onde e como os dados devem ser armazenados, para que seja possível o compartilhamento e o reuso. Curty (2018), relata que esta é uma pergunta frequente por parte dos pesquisadores quando se trata sobre a importância do compartilhamento de dados científicos para acesso aberto como forma de melhor aproveitar recursos públicos injetados em pesquisa, estimular transparência e a reprodutibilidade em ciência. No entanto, a escolha e opções de repositórios para depósito e armazenamento de dados de pesquisa não são claras, o que dificulta a decisão dos pesquisadores quanto ao destino dos dados.

De modo a atender às exigências impostas, muitos pesquisadores depositam seus dados em repositórios abertos, sem critérios para a escolha de tais plataformas. A seleção do repositório de dados, onde será possibilitada a reutilização futura dos dados, também contribui para o novo valor a ser alcançado. Nesse contexto, as características do repositório de dados como, por exemplo, disciplinas e comunidades atendidas, políticas de dados, custos de depósito e preservação, recursos de descoberta e acesso, serviços de suporte ao usuário, sustentabilidade e reputação, bem

como outros critérios de seleção, são importantes. Além disso, os depositantes de dados podem estar interessados na confiabilidade do repositório que será responsável por fornecer a administração e permitir o acesso contínuo aos dados que foram submetidos para uso futuro (DOWNS, 2021).

O repositório *Figshare*[2], por exemplo, é a alternativa para depósito por muitos pesquisadores. A plataforma possui recursos de compartilhamento e visibilidade, métricas e contagens de impacto e aceita muitos tipos de dados e arquivos (CURTY, 2018). O SciELO, por exemplo, em 2018 anunciou uma parceria com o Figshare, cobrindo os periódicos SciELO do Brasil. Já alguns especialistas sugerem que os dados sejam protegidos em repositórios temáticos porque, segundo eles, permitem o uso especializado de metadados e uma maior revisão e validação por especialistas no campo. No entanto, nem todas as áreas possuem repositórios de dados, o que explica as dificuldades para encontrar o local apropriado para armazenamento nos repositórios existentes, isto se deve a própria especificidade e peculiaridade dos dados (SILVA, 2019).

Devido à falta de informações referentes aos repositórios apropriados para o compartilhamento de dados de pesquisa, as bibliotecas universitárias, os centros de pesquisa e as associações científicas estadunidenses começaram a qualificar os repositórios confiáveis. Nesse contexto, alguns foram aprimorados e outros criados, dentre os quais muitos com foco disciplinar, visando atender às necessidades específicas das disciplinas e campos científicos (CURTY, 2018).

Repositório confiável é aquele cuja missão é fornecer acesso de longo prazo a recursos digitais gerenciados. Um repositório com essas características aceita a responsabilidade pela manutenção a longo prazo dos recursos digitais, projeta seus sistemas de acordo com as convenções e os padrões comumente aceitos, estabelece metodologias para avaliação de sistemas que atendem às expectativas de confiabilidade da comunidade, cumpre suas responsabilidades com depositantes e usuários de forma aberta e explícita, e permite que sejam auditadas e medidas suas políticas, práticas e desempenho (JANTZ; GIARLO, 2006).

Um repositório digital confiável é mais do que uma organização encarregada de armazenar e administrar objetos digitais, ou seja, é aquele cuja missão é fornecer acesso confiável, por longo prazo, a recursos digitais administrados à sua comunidade-alvo, agora e no futuro (RLG/OCLC WORKING GROUP ON DIGITAL ARCHIVE ATTRIBUTES, 2002).

[2] https://figshare.com/

Sabe-se que confiança se desenvolve em diversos níveis. No caso de repositórios digitais confiáveis, no mínimo três níveis são aplicáveis: a confiança de que os produtores estão enviando as informações corretas; a confiança de que os consumidores estão recebendo as informações corretas; e a confiança de que os fornecedores estão prestando serviços adequados (THOMAZ, 2007).

A sustentabilidade dos repositórios levanta uma série de questões desafiadoras em diferentes setores: organizacional, técnica, financeira e jurídica. A avaliação dos repositórios pode ser uma contribuição importante para garantir a confiabilidade e a durabilidade dos repositórios de dados. Ao passar por uma avaliação, os repositórios podem demonstrar para seus usuários e financiadores que uma autoridade os avaliou e endossou sua confiabilidade (CORETRUSTSEAL, 2020a).

Neste contexto, avaliar o desenvolvimento desses repositórios é de suma importância, pois possibilita a evolução de novas práticas e maior domínio sobre este novo campo de atuação. Por este motivo, conhecer aspectos de gerenciamento de dados de pesquisa, padrões, convenções e metadados utilizados em repositórios de instituições de diversos países, se torna essencial, pois permite o desenvolvimento e melhorias. Dito isto, a realização de avaliações permite identificar os pontos fortes e fracos, trazendo mais confiabilidade ao repositório, como também ajudam as comunidades de dados - produtores, repositórios e consumidores - a melhorar a qualidade e a transparência de seus processos e a aumentar a conscientização e a conformidade com os padrões estabelecidos (CORETRUSTSEAL, 2020a).

Existem na literatura, estudos que abordam o tema auditoria, certificação e avaliação. Deste modo, há o propósito de se estabelecer a diferenciação conceitual. Barzelay (2014) enfatiza que a auditoria consiste em uma forma de investigação que tem o intuito de verificar o cumprimento de critérios e ações. No caso de instituições com repositórios de médio porte, as auditorias podem não aparentar ser importantes porque seu sistema encontra-se, muitas vezes, em estado incipiente. Nestes casos, as auditorias têm o papel de serem mecanismos que comprovem o nível de preparo da instituição para confrontar essa área (ARELLANO, 2017).

A certificação, segundo a Associação Brasileira de Normas Técnicas (2014), é um processo no qual uma entidade independente avalia se determinado produto atende às normas técnicas. O resultado satisfatório leva a concessão da certificação, garantindo a conformidade e a qualidade e conferindo maior credibilidade ao repositório perante os pesquisadores. A avaliação, por sua vez, consiste fundamentalmente em fazer um julgamento de valor a respeito de uma intervenção ou sobre qualquer um de seus componentes, com o objetivo de

ajudar na tomada de decisões, que pode ser resultado da aplicação de critérios e de normas (CONTANDRIOPOULOS *et al*., 1997).

Instituições brasileiras estão em busca do desenvolvimento e implementação de repositórios de dados de pesquisa. Desta forma, se faz necessário ampliar os estudos referentes à avaliação, pois trata-se de uma etapa crucial na implementação e acompanhamento desse tipo de sistema de informação. Internacionalmente, existem instituições que avaliam e oferecem certificações para repositórios de dados, no entanto, no Brasil, os estudos são incipientes, com falta de padronização e clareza na definição de critérios avaliativos.

Em suma, esta pesquisa pretende contribuir para o desenvolvimento de repositórios de dados de pesquisa de qualidade, que atendam a critérios estabelecidos para repositórios confiáveis, avançando no conhecimento sobre os critérios de avaliação. Seu objetivo consiste em fornecer uma síntese dos requisitos para avaliação em conformidade com os principais modelos internacionais a fim de facilitar a compreensão dos desenvolvedores frente ao desafio de implementar e manter repositórios de dados de pesquisa confiáveis e com garantia de acesso contínuo a longo prazo. Com esta pesquisa, espera-se alavancar as discussões referentes a avaliação como também fornecer subsídios para a compreensão do processo de avaliação e auditorias. Pretende-se realizar uma análise e comparação a partir dos instrumentos existentes, incluindo uma visão geral com vista a criar um conjunto único que apoie os repositórios nas autoavaliações e na melhoria geral, os direcionando para o caminho da certificação. A visão é desenvolver uma abordagem prática para os repositórios de dados de pesquisa autoavaliarem seus níveis de capacidade atuais e definir onde é necessário concentrar recursos.

Sendo assim, esta pesquisa não tem como propósito a substituição das leituras dos documentos publicados pelas organizações, apenas atuará de modo complementar para facilitar a compreensão e instigar a reflexão auxiliando nos processos de planejamento e implementação tendo em vista o preparo para futura certificação. Mesmo sem a pretensão de uma certificação formal, os requisitos contemplados no estudo podem ser usados como referência e para identificar as lacunas e áreas que precisam de maior atenção.

Considera-se que o domínio das tecnologias que envolvem as novas práticas em AADP são fundamentais para o progresso da pesquisa científica. Por este motivo, a revisão de literatura terá como objetivo trazer conceitos relacionados a dados de pesquisa, repositórios de dados de pesquisa confiáveis e avaliação de repositórios de dados de pesquisa, a fim de traçar

um histórico do desenvolvimento, como também identificar as iniciativas que vêm contribuindo para o progresso da Ciência, frente às atuais tecnologias de comunicação e informação.

As próximas seções apresentam a justificativa e o objetivo do estudo, além da fundamentação teórica. Os procedimentos metodológicos foram planejados em 3 etapas, conforme detalhado a seguir: na etapa 1 será realizado um levantamento bibliográfico e a seleção do material; na etapa 2 será executada a seleção, leitura e interpretação da documentação selecionada; na etapa 3, será realizada a análise, padronização e sistematização dos critérios/requisitos. A tese finaliza com a apresentação dos resultados e conclusões.

1.1 JUSTIFICATIVA

As considerações já mencionadas que justificam a relevância da tese apresentada, quais sejam, contribuir para o desenvolvimento de repositórios de dados de pesquisa de qualidade, avançando no conhecimento sobre os critérios para avaliação, afim de auxiliar as instituições a criarem e manterem repositórios adequados e confiáveis, soma-se ao fato de que a equipe responsável pelo o desenvolvimento do projeto ‘Rede de Dados de Pesquisa Brasileira (RDP Brasil)’, pertence a UFRGS, envolvendo o CEDAP, que caracteriza-se por ser um órgão auxiliar da Faculdade de Biblioteconomia e Comunicação (FABICO). O projeto RDP Brasil foi desenvolvido em parceria com a Universidade Federal do Rio Grande (FURG), sob coordenação executiva da Rede Nacional de Ensino e Pesquisa (RNP) e do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), com o objetivo de desenvolver pesquisas sobre a temática. Desta forma, pretende-se preencher a lacuna que existe na área da CI no que diz respeito a estudos sobre AADP, buscando-se ao mesmo tempo, um arcabouço conceitual sólido entre os pesquisadores para dar continuidade as pesquisas já desenvolvidas pelos grupos de pesquisas no Brasil.

É importante destacar que este trabalho se insere dentro da área de conhecimento Ciências Sociais Aplicadas/Ciência da Informação. A proposição da temática desta pesquisa foi inicialmente norteada pelo interesse do autor em assuntos relacionados a dados de pesquisa. Ainda é imprescindível ressaltar que, dentre a literatura existente, houve a percepção da falta de um estudo específico que aborde esses novos padrões na comunidade científica de forma teórico-prática, portanto, acredita-se que os resultados deste estudo possa ser um importante insumo para iniciativas relacionadas ao AADP.

Térmens e Leija (2017) caracterizam que o estudo de repositórios confiáveis é uma das linhas de pesquisa em preservação digital mais trabalhadas em nível internacional e que se busca definir metodologias e ferramentas de avaliação de conformidade com padrões. Souza e Aganette (2020) em seu estudo, puderam perceber que as publicações sobre o assunto "repositórios confiáveis" ainda são incipientes e que grande parte aborda aspectos conceituais e implementação da preservação digital. Citam a possibilidade de preservação realizada pelos repositórios, mas não exploram o conceito de repositórios digitais confiáveis. Acredita-se que isso acontece porque ainda há um número reduzido de repositórios que se enquadram nesta categoria, dada a exigência de cumprimento de normas e padrões e a realização de auditorias para a certificação da confiabilidade dos softwares, um processo dispendioso para muitas instituições gestoras de repositórios.

O debate desenvolvido nesta pesquisa busca retratar a atual situação no Brasil, ou seja, o estado incipiente em que se encontram os estudos referentes aos critérios/requisitos essenciais para o desenvolvimento de repositórios de dados de pesquisa confiáveis. O processo contínuo de autoavaliações proporciona diversos benefícios contribuindo para a construção de uma base organizacional sólida para submeter-se futuramente às certificações e auditorias. A partir das argumentações apresentadas, destaca-se como problema desta pesquisa a seguinte questão:

Quais critérios e requisitos são pertinentes para assegurar a confiabilidade de repositórios de dados de pesquisa?

1.2 OBJETIVOS

Os objetivos desta pesquisa estão divididos em objetivo geral e objetivos específicos da seguinte forma:

1.2.1 Objetivo geral

Reunir critérios e requisitos pertinentes para assegurar a confiabilidade de repositórios de dados de pesquisa.

1.2.2 Objetivos específicos

Para responder ao problema de pesquisa proposto e atingir o objetivo geral da pesquisa, os seguintes passos são necessários:

a) realizar um levantamento, em âmbito internacional, dos princípios e instrumentos de avaliação existentes para repositórios de dados de pesquisa confiáveis;
b) identificar e cotejar os critérios/requisitos encontrados nos documentos e estabelecer uma padronização;
c) criar um conjunto único de critérios/requisitos, a partir dos instrumentos internacionais selecionados, que apoie os repositórios de dados de pesquisa nas autoavaliações e na melhoria geral, os direcionando para o caminho da certificação, tornando-os confiáveis.

2 REVISÃO DE LITERATURA

São apresentadas nesta seção informações levantadas na literatura da área que serviram de embasamento teórico para a pesquisa. A revisão de literatura aborda os seguintes aspectos: dados de pesquisa; repositórios de dados de pesquisa confiáveis; e avaliação de repositórios de dados de pesquisa.

2.1 DADOS DE PESQUISA

O tema dados abertos de pesquisa vem sendo estudado desde 2004, quando os ministros de Ciência e Tecnologia dos países membros da *Organization for Economic Co-operation and Development* (OECD) se reuniram em Paris, para discutir a necessidade de um guia internacional voltado para o acesso aos dados de pesquisa (HENNING *et al*., 2019, p. 5). Esta reunião culminou na *Declaration on access to research data from public funding*[3], que estabelece objetivos e princípios relacionados com as ações de abertura dos dados tais como: transparência, conformidade legal, responsabilidade formal, profissionalismo, proteção da propriedade intelectual, interoperabilidade, qualidade e segurança, além da eficiência e prestação de contas para a sociedade (ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT, 2004). Em 2007 esta mesma instituição lança uma nova versão dessa declaração[4], incluindo quatro novos princípios: qualidade, segurança, flexibilidade e sustentabilidade. Nesta mesma linha, o relatório da *Royal Society*, lançado em 2012, analisa os desafios e as oportunidades trazidas pelas novas formas de reunir, armazenar, manipular e transmitir os dados e as informações sobre pesquisas científicas (ROYAL SOCIETY, 2012).

Os dados estão em todos os lugares. Existem dados sobre pessoas, empresas, governos e as novas tecnologias são instrumentos que estão impulsionando a inovação e a abertura desses dados, o que permite uma maior transparência e participação dos cidadãos nas soluções de problemas e avanços tecnológicos. Os dados abertos estão tendo alta visibilidade e são abordados na literatura em pelo menos dois cenários: referente às organizações, às empresas e aos governos que necessitam estar adequados à legislação e por pressão popular precisam

[3] https://legalinstruments.oecd.org/en/instruments/157
[4] https://www.oecd.org/sti/inno/38500813.pdf

abrir seus dados a sociedade; e os dados de pesquisa, que apresentam uma tendência de abertura, visando o aprimoramento da ciência como um todo. Um efeito destes processos é a existência de uma quantidade cada vez maior de instituições disponibilizando seus dados, tendo a necessidade de ter recursos para estruturar tais informações para melhorar a disseminação e a recuperação.

Dados abertos podem ser definidos como dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa, sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras (OPEN KNOWLEDGE FOUNDATION *et al*., 2009). Uma das tecnologias que fornece suporte a abertura de dados é a Web Semântica, que dá às pessoas a capacidade de criarem repositórios de dados na Web, construírem vocabulários e escreverem regras para interoperarem com esses dados. A linkagem de dados só é possível com tecnologias como *Resource Description Framework* (RDF), *SPARQL Protocol and RDF Query Language* (SPARQL), *Ontology Web Language* (OWL) e *Simple Knowledge Organization System* (SKOS) (W3C BRASIL, 2011). Conforme Berners-Lee (2006), a Web Semântica não consiste apenas em colocar dados na Web, mas sim criar links para que uma pessoa ou máquina possa explorar a web de dados.

De acordo com a OECD (2007, online), o termo ‘dado de pesquisa’ é definido como registros factuais (pontuações numéricas, registros textuais, imagens e sons) usados como fontes primárias para pesquisa científica, e que são comumente aceitos na comunidade científica como necessários para validar resultados da pesquisa. Um conjunto de dados de pesquisa constitui uma representação parcial e sistemática do assunto investigado (ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT, 2007). Também podem ser definidos como todas as informações que foram coletadas, observadas, geradas ou criadas para validar os resultados da pesquisa original, incluindo os formatos não digitais, como cadernos de laboratório e cadernos de esboços (INGRAM, 2016).

Sayão e Sales (2015) classificam os dados de pesquisa de acordo com sua natureza, origem ou de acordo com seu status no fluxo de trabalho da pesquisa. Segundo a sua natureza, os dados seriam classificados como “[...] números, imagens, vídeos ou áudio, software, algoritmos, equações, animações ou modelos e simulações.”. Quanto a sua origem, os dados podem ser classificados como observacionais, quando “[...] obtidos por meio de observações diretas [...]”, computacionais, quando “[...] resultados de execução de modelos computacionais ou de simulações” e dados experimentais, quando “[...] provenientes de situações controladas em bancadas de laboratório” (SAYÃO; SALES, 2015, p. 8).

O reconhecimento do potencial informacional dos dados, distribuídos em rede de computadores transforma a visão que caracterizava dados de pesquisa, registrados em mídia impressa ou mesmo em formatos digitais, como simples subprodutos dos processos de pesquisa. Os dados eram considerados somente na sua configuração final, sem considerar os seus ciclos de vida, versões e linhagens e eram descartados ou armazenados em mídias ou servidores sem a devida gestão quando os projetos eram concluídos (SAYÃO; SALES, 2014).

Embora sejam considerados o alicerce do conhecimento científico e tecnológico, dados de pesquisa não são fáceis de estruturar, organizar, descrever e disponibilizar, para que sejam compreensíveis agora e no futuro (DUDZIAK, 2016). Políticas e parâmetros estão sendo desenvolvidos há alguns anos e percebe-se a necessidade de aprimoramentos das infraestruturas voltadas ao compartilhamento e reuso desses dados.

Com a discussão no sentido de que os resultados de pesquisas científicas devem ser de amplo acesso a fim de garantir os atributos da ciência moderna, o campo dos dados de pesquisa evoluiu em três grupos: as preocupações com a disponibilização dos dados e sua permanência; a exploração dos dados; e os dados ligados a uma publicação como elementos de prova, ou seja, uma publicação ampliada. Em uma publicação ampliada, os dados de pesquisa devem estar intrinsecamente ligados às publicações científicas. Atualmente, os editores sugerem a disponibilização dos dados para fins de validação de uma pesquisa e para possibilitar a reprodução dos resultados. Interligar esses dados as suas respectivas publicações permite uma maior valorização dos artigos publicados como também aumentam as possibilidades de que outros pesquisadores acessem e citem esses dados.

Compartilhar publicamente dados de pesquisa com outros pesquisadores permite que esses recursos valiosos contribuam muito além de sua análise original. Além de serem usados para confirmar os resultados originais, os dados brutos podem ser usados para explorar hipóteses relacionadas ou novas, particularmente quando combinados com outros conjuntos de dados disponíveis publicamente. A comunidade científica também se beneficia, ou seja, o compartilhamento de dados incentiva perspectivas múltiplas, ajuda a identificar erros, desencoraja fraudes, é útil para treinar novos pesquisadores e aumenta o uso eficiente de recursos financeiros e populacionais de pacientes, evitando a coleta de dados duplicados. (PIWOWAR; DAY; FRIDSMA, 2007).

Organizações internacionais como, por exemplo, a *Research Data Alliance* (RDA) e a *The Future of Research Communications and e-Scholarship* (FORCE11) estão se empenhando para a consolidação de políticas globais que estimulem as melhores práticas de

compartilhamento e reuso dos dados de pesquisa. Da mesma forma, no Brasil, existem organizações como a RNP que financia grupos de pesquisa sobre AADP com o objetivo de executar atividades que contribuam para a identificação de práticas, mapeamento de requisitos e prototipação de sistemas que facilitem a disseminação de informações científicas (REDE NACIONAL DE ENSINO E PESQUISA, 2017).

Atualmente, no Brasil, existem poucos repositórios de dados de pesquisa abrangentes, o que atribui às instituições a responsabilidade de desenvolver formas de gerenciar, compartilhar e disponibilizar os dados de pesquisa e é neste cenário que as universidades e agências de fomento estão desenvolvendo políticas e serviços em AADP para apoiar o gerenciamento de dados de pesquisa. Rocha, Caregnato e Gabriel Junior (2018), mencionam, que o cenário demanda das universidades e de centros de pesquisa a oferta de novos serviços de apoio ao pesquisador, como os serviços de dados da pesquisa, que auxiliam pesquisadores no gerenciamento dos dados e promovem a preservação e o compartilhamento desses dados.

Na publicação *The FAIR guiding principles for scientific data management and stewardship*, Wilkinson *et al.* (2016) apresentam a necessidade urgente de melhorar a infraestrutura que suporta a reutilização de dados. No artigo, os autores descrevem que além da coleta, anotação e arquivamento adequados, a administração de dados inclui a noção de 'cuidado de longo prazo' de ativos digitais valiosos, com o objetivo de que eles sejam descobertos e reutilizados para investigações posteriores, isoladamente ou em combinação com dados gerados (WILKINSON *et al*., 2016).

Os autores salientam que os princípios FAIR se aplicam não apenas aos 'dados' no sentido convencional, mas também aos algoritmos, ferramentas e fluxos de trabalho que levaram a esses dados. Todos os objetos de pesquisa digitais se beneficiam da aplicação desses princípios, uma vez que todos os componentes do processo de pesquisa devem estar disponíveis para garantir transparência, reprodutibilidade e reutilização (WILKINSON *et al*., 2016).

A iniciativa GO FAIR foi lançada no final de 2017 pelos Governos holandês, alemão e francês como meio de promover e acelerar a convergência da comunidade, com base na visão da União Europeia. Seguindo os exemplos da Internet e da *World Wide Web* (WWW), a GO FAIR opera por meio da participação voluntária das partes interessadas, para atingir uma massa crítica de usuários comprometidos com o conjunto de especificações. A iniciativa trabalha com pesquisadores especializados, órgãos políticos, editores, repositórios e agências de financiamento (WILKINSON *et al*., 2016). A GO FAIR Brasil foi estabelecida no final de

2018 e em 2020 foi lançada a rede GO FAIR Brasil Saúde, que se propõe a tornar os dados interoperáveis com os mesmos padrões internacionais (GO FAIR BRASIL SAÚDE, 2020).

Como desdobramento e alinhamento às novas exigências, países ao redor do mundo têm traçado estratégias nacionais para sua inserção na abertura de dados, como a Alemanha, Austrália, Brasil, Canadá, Estados Unidos, Holanda, Portugal, Reino Unido e União Europeia. As iniciativas contemplam desde políticas até a criação de infraestruturas de repositórios de dados e formação de pesquisadores (SANTOS; ALMEIDA; HENNING, 2017).

Alguns repositórios de dados de pesquisa estão disponíveis aos pesquisadores, e políticas editoriais estão sendo definidas por periódicos para recomendar que todos os artigos submetidos sejam acompanhados do depósito de dados em repositórios de dados de pesquisa. A decisão baseou-se em um princípio inerente à publicação científica, isto é, qualquer pesquisador pode ser capaz de verificar, replicar ou reproduzir a pesquisa realizada, até mesmo utilizar os resultados de pesquisas para promover novos avanços no conhecimento científico e tecnológico (DUDZIAK, 2016).

Diversos editores de revistas internacionais passaram a fazer esta exigência como a Elsevier, a PLoS, a Nature, a Wiley, entre outros que possuem políticas de depósito de dados de pesquisa associado ao processo de submissão de artigos. Entre as revistas científicas brasileiras, um exemplo com política de publicação semelhante à PLoS é o da *Brazilian Political Science Review* (BPSR), vinculada à Associação Brasileira de Ciência Política. Caracteriza-se por ser uma revista de acesso aberto, publicada em inglês e que desde 2013 os autores dos artigos são solicitados a disponibilizar os dados que embasaram os *papers* (MARQUES, 2014).

Outro conceito importante abordado na literatura é o de *data publishing*, que se caracteriza como um modelo de publicação cientifica, que promove a publicação de conjuntos de dados; publicações acadêmicas e metadados que descrevem um conjunto de dados. Seu objetivo final é fornecer informações sobre o quê, onde, por que, como e quem é responsável pelos dados (PAVÃO; SILVA; SILVEIRA, 2020).

Os *data papers*, ou seja, os artigos de dados que se caracterizam por ser uma publicação em uma revista que possui como objetivo descrever dados em vez de informar sobre uma pesquisa ou conclusões também vêm aparecendo em estudos:

> Essa nova modalidade de publicação, que utiliza os data journals como periódicos para publicação da descrição consistente, visa desengavetar os

> dados de pesquisa e torná-los visíveis à comunidade acadêmica, tendo como motivador principal as potencialidades que o reuso de dados pode trazer para o desenvolvimento científico, ao mesmo tempo que oferecem aos pesquisadores a possibilidade de citação de uma maneira tão consistente quanto a de um artigo de pesquisa tradicional. (ROCHA, 2020, p. 160).

Alguns aspectos importantes são abordados na literatura como, a política editorial, que deve conter orientações sobre onde os autores irão depositar os dados provenientes de seus artigos. A inserção dos conjuntos de dados linkados com os artigos, seja por meio de repositório ou documento complementar, irá exigir da equipe de pareceristas de periódicos novos perfis de avaliadores. Da mesma forma que ocorre em periódicos convencionais, os periódicos que receberem os conjuntos de dados irão precisar desenvolver diretrizes, critérios e instruções para os revisores (PAVÃO; SILVA; SILVEIRA, 2020).

Santos (2020), em seu estudo, avaliou como os indexadores estão acompanhando os questionamentos sobre os dados abertos de pesquisa no processo de indexação de periódicos científicos. Estudos como este, são importantes pois os dados de pesquisa começaram a ser vistos como complemento essencial aos documentos convencionais, como os artigos, e terão que possuir certa qualidade de dados para que possam embarcar nos indexadores, assim como os artigos são avaliados para tal finalidade de indexação. Deste modo, os indexadores, estão vendo a questão de dados de pesquisa para o processo de indexação, e como isto reflete na pesquisa.

É importante pensar estrategicamente, pois ainda não contamos com uma infraestrutura nacional que comporte os dados que estão sendo solicitados aos cientistas. Segundo Aventurier (2017b), no Brasil a prática de compartilhar dados é limitada por três causas: falta de treinamento dos pesquisadores, falta de incentivos para compartilhamento dos dados de pesquisa, e questões ligadas à privacidade, propriedade intelectual e ética dentro dos estudos. Acredita-se que uma mudança na cultura científica é necessária onde pesquisadores são estimulados e recompensados por compartilhar e onde instituições fornecem políticas para o chamado *data sharing* incluído pré-requisitos (mandatos) (AVENTURIER, 2017b).

Em seu artigo, Presser e Silva (2018) abordam o gerenciamento de dados de pesquisa como *wicked problem*, ou seja, como um problema único e altamente complexo, de forma que é difícil aprender com problemas anteriores e por existir múltiplas restrições ideológicas, políticas ou econômicas sobre as possíveis soluções. Segundo Presser e Silva (2018), é necessário reconhecer e definir a complexidade do problema e criar condições para a responsabilidade coletiva ao tratá-lo, ao invés da tradicional expectativa de ir diretamente na

busca da solução. Entende-se que o gerenciamento de dados de pesquisa não pode ser compreendido como um problema técnico.

Nesta perspectiva, observa-se que a literatura brasileira vem tratando sobre questões de compartilhamento, reuso e avaliação de dados de pesquisa, mas mantendo os devidos cuidados referente à ética e questões legais de propriedade intelectual. Para disponibilizar os dados de pesquisa é preciso possuir alguns conhecimentos mínimos sobre direitos autorais, desta forma, as licenças escolhidas para acompanhar aos dados de pesquisa são um assunto fundamental. As licenças dos dados devem apresentar-se com a intenção de que os seres humanos, assim como as máquinas, possam ler o conteúdo delas. Para isso, deveria tornar-se clara a licença dos dados no periódico, no texto do arquivo e disponibilizar também os metadados do arquivo de dados (TORRADO; MOURA, 2020).

Segundo Henning e Moreira (2020), algumas ações e iniciativas baseadas no que outros países estão desenvolvendo já existem no Brasil e outras estão em fase de negociação e desenvolvimento, podendo citar:

a) a elaboração de um Plano Nacional de Ciência Aberta com ações e prazos bem definidos, detalhando todos os investimentos financeiros e de recursos humanos necessários;
b) a criação de uma Plataforma de Ciência Aberta com desdobramentos para cada uma das suas práticas envolvendo as instituições responsáveis e interessadas em cada temática, como tem sido praticado nos Países Baixos;
c) investimento na criação e na qualidade dos repositórios de dados para que sejam confiáveis, assim como em cursos de capacitação sobre Ciência Aberta de modo a estabelecer cultura sobre a temática e dar maior visibilidade ao tema de acesso aberto e gestão de dados;
d) incentivo às universidades a criarem as suas políticas de abertura dos dados de pesquisa e passar a cobrar um Plano de Gestão de Dados mínimo em todos os projetos de pesquisa da instituição;
e) a tentativa de se filiar às possíveis iniciativas visando aprender com elas ou tentar acompanhar o trabalho que elas vêm desenvolvendo a nível internacional;
f) a preparação de profissionais (administradores de dados) para dar apoio aos pesquisadores na gestão e compartilhamento, assim como orientá-los quanto à publicação dos seus artigos em acesso aberto.

Um estudo realizado por Torino, Martinez e Vidotti (2020) discute a disponibilização e a publicação de dados de pesquisa, com o objetivo de esclarecer as divergências e similaridades. Esclarece que a disponibilização de dados de pesquisa consiste em torná-los acessíveis publicamente, desconsiderando a relevância de aspectos que são prioritários no processo de publicação. O objetivo principal da disponibilização é que os dados possam ser acessíveis e reutilizáveis pela comunidade científica. Em contrapartida, a publicação de dados de pesquisa se faz de modo análogo à publicação de artigos, levando em consideração aspectos normativos e de padronização, cuidadosa explicitação metodológica que permitirá a interpretação e o reuso dos dados por terceiros, gerando com isso citações.

Contudo, para que os dados de pesquisa sejam recuperados, há de se atentar para o ambiente de disponibilização, que deve oferecer infraestrutura para que tais dados sejam FAIR, o que os tornará recuperáveis, passíveis de reuso e de citação. Torino, Martinez e Vidotti (2020) mencionam que, disponibilizar não é o suficiente para que os dados de pesquisa possam ser reutilizados e que, além disso, o processo de publicação é que dará a eles características de qualidade, confiabilidade e veracidade. Atualmente foram desenvolvidas algumas ferramentas online que auxilia os investigadores e gestores de dados a tornar os conjuntos de dados FAIR como, por exemplo, a ferramenta FairDataBR[5], desenvolvida pela Universidade Federal da Paraíba (UFPB).

Outro requisito de grande importância que aparece na literatura, diz respeito a curadoria de dados de pesquisa e a importância de se considerar os perfis da grande e pequena ciência (*big science e small science*) durante o planejamento de uma infraestrutura de plataformas de gestão de dados. Isso porque são dois universos de pesquisa que demandam infraestruturas distintas.

Enquanto na grande ciência há uma uniformização na geração dos dados e uma demanda imediata para o compartilhamento e grandes investimentos em infraestruturas; na pequena ciência os dados são gerados e coletados por pequenas equipes numa infinidade de laboratórios, são extremamente heterogêneos e raramente arquivados para o compartilhamento e reuso. Esses dados precisam de infraestruturas específicas que considerem seus fluxos de geração, metodologias, culturas de compartilhamento e de esquemas de recompensa e sustentabilidade (SAYÃO; SALES, 2019). O problema apresentado pelos autores refere-se a forma como a gestão de dados de pesquisa é abordada,

[5] https://wrco.ufpb.br/fair/

ou seja, a gestão de dados é geralmente colocada de forma uniforme para toda a ciência e os perfis que caracterizam a pequena e grande ciência são geralmente desconsiderados no planejamento e desenvolvimento das infraestruturas de pesquisa.

A partir das reflexões referentes as infraestruturas específicas, e para que haja o compartilhamento, uso e reuso dos dados de pesquisa, Medeiros (2015), deixa claro em sua tese que é necessário que os dados sejam compartilhados em repositórios próprios para este fim, buscando o máximo de alcance possível no seu reuso.

2.2 REPOSITÓRIOS DE DADOS DE PESQUISA CONFIÁVEIS

Nos anos 90 teve início a era dos repositórios de documentos digitais, que tinham como objetivo difundir a produção científica de certo nicho ou instituição. Mais recentemente, o desenvolvimento da tecnologia sobre objetos digitais foi agregada à proposta de Ciência Aberta, com a disponibilização não apenas de documentos científicos, mas também dos dados coletados durante as pesquisas (CAMPÊLO; BARRETO NETO, 2019). Nessa perspectiva, surgem os repositórios de dados de pesquisa, que "[...] têm como objetivo fundacional garantir o acesso contínuo e aberto - agora e no futuro - aos resultados de pesquisa que se manifestam na forma de dados, e que são considerados parte importante do patrimônio digital da humanidade." (SAYÃO; SALES, 2016, p. 96). Sales (2014), apresenta alguns dos benefícios mais perceptíveis dos repositórios de dados de pesquisa para a pesquisa científica:

a) visibilidade dos dados: amplia a visibilidade dos dados de pesquisa permitindo que eles sejam consultados e citados mais frequentemente;
b) compartilhamento de dados: pela sua capacidade de agregação e organização de recursos informacionais, os repositórios tornam-se um dispositivo importante de troca de experiências e compartilhamento de dados;
c) crédito ao autor dos dados: os repositórios de dados tornam possível identificar as coleções de dados e seus autores de forma unívoca e persistente, permitindo que os autores sejam reconhecidos, citados, avaliados e recompensados pelo trabalho intelectual de coleta, geração e organização dos dados;
d) preservação digital: oferece um ambiente tecnológico, gerencial e de padronização propício para a preservação de longo prazo dos dados de pesquisa de valor contínuo;

e) memória científica e transparência: contribui para a formação da memória científica das instituições no que diz respeito aos dados, complementando os repositórios institucionais que estão focados nas publicações acadêmicas;
f) segurança dos dados: oferece sistema de armazenamento seguro, esquemas de backup e segurança física que se contrapõem ao armazenamento informal em mídias portáteis e computadores pessoais frequentemente usados pelos pesquisadores;
g) disponibilidade: permite que os dados estejam disponíveis on-line para serem acessados, baixados, visualizados e processados por pessoas ou por sistemas;
h) curadoria digital: proporciona um ambiente apropriado para os processos de avaliação, de adição de valor, reformatação, agregação e recriação de dados promovidos pela curadoria digital;
i) serviços inovadores: abre possibilidades de criação de novos serviços de informação para pesquisadores, gestores e financiadores de pesquisa a partir da análise e integração dos dados arquivados com fontes internas e externas à instituição;
j) reuso dos dados: aumenta o grau de reuso e reinterpretação dos dados possibilitando a realização de novas pesquisas de caráter interdisciplinar; minimiza a duplicação de esforços e otimiza os investimentos na coleta e geração de dados;
k) redes de repositórios: permite por meio de protocolos de interoperabilidade, como o OAI-PMH, a formação de redes de repositório dados; abre a possibilidade de inserção dos repositórios de dados às redes interoperáveis definidas pelo padrão *Linked Data*;
l) indicador de qualidade e produtividade da instituição: as coleções de dados organizadas e arquivados no repositório são evidências da qualidade e da relevância das atividades de pesquisa da instituição, atestando a sua produtividade e seu valor acadêmico.

Segundo Costa (2017), repositórios de dados de pesquisa configuram-se como bases digitais onde são armazenados e disseminados os dados de pesquisa. A autora menciona que uma das principais fontes de informações sobre repositórios de dados de pesquisa está no *Registry of Research Data Repositories* (re3data.org), que se caracteriza por ser um registro global de repositórios de dados de pesquisa que abrange repositórios de diferentes disciplinas. Ele apresenta repositórios para armazenamento permanente e acesso a conjuntos de dados para pesquisadores, órgãos de financiamento, editores e instituições acadêmicas. Desta forma, o registro das instituições provedoras de dados de pesquisa é importante para promover uma

cultura de compartilhamento, maior acesso e melhor visibilidade dos dados de pesquisa no mundo.

O informe *The State of Open Data 2018*[6], que caracteriza-se por ser um relatório anual do Figshare que analisa as atitudes globais em relação aos dados abertos, entrevistou pesquisadores de todos os continentes sobre as motivações, hábitos, conhecimento e práticas de compartilhamento de dados. Os resultados, comparados aos informes de 2016 e 2017, trazem informação relevante sobre a evolução dos dados de pesquisa em todo o mundo além de como fortalecer esta prática na academia, para que atinja os resultados esperados. Desta forma, entende-se que os resultados da pesquisa mostram um progresso em relação aos dados abertos, ou seja, eles estão se tornando mais integrados à comunidade de pesquisa (DIGITAL SCIENCE REPORT, 2018).

O apoio ao compartilhamento dos dados de pesquisa é outro ponto importante nas ações da Fapesp, que fomenta a criação de infraestrutura por meio de ações coordenadas com entidades de ensino e pesquisa públicas localizadas no estado de São Paulo. Essas ações surgem com base na maior eficiência no uso de informações advindas da ciência (SHINTAKU; LANNE, 2020).

Em uma reportagem no site da FAPESP, Claudia Bauzer Medeiros, professora do Instituto de Computação da Universidade Estadual de Campinas (UNICAMP) e integrante da Coordenação Adjunta da FAPESP para o Programa de Pesquisa em *eScience* e *data Science* cita uma das iniciativas pioneiras na América Latina, ou seja, o desenvolvimento de um portal buscador de metadados (um metabuscador[7]). Este portal foi desenvolvido pela Universidade de São Paulo (USP) e lançado em 2019, envolvendo as seis universidades públicas do Estado de São Paulo: Universidade Estadual de Campinas (UNICAMP), USP, Universidade Estadual Paulista (UNESP), Universidade Federal de São Carlos (UFSCar), Universidade Federal do ABC (UFABC) e a Universidade Federal de São Paulo (UNIFESP). Além destas universidades, participam da rede, o Instituto Tecnológico de Aeronáutica (ITA); Embrapa Informática Agropecuária (CNPTIA/Embrapa) e FAPESP COVID 19- Data Sharing/BR.

Cada instituição desenvolveu seu próprio repositório, para que fosse possível a integração por um portal único, que diariamente busca informações sobre os dados de cada

---

[6] DIGITAL SCIENCE. The State of Open Data 2018. 2018. Disponível em: <https://www.digital-science.com/resources/portfolio-reports/state-open-data-2018/>. Acesso em: 4 maio. 2020.

[7] https://metabuscador.uspdigital.usp.br/

instituição e disponibiliza-as de forma integrada. Por meio deste portal é possível ter acesso aos dados gerados em pesquisas científicas, o que aumentará a visibilidade da pesquisa permitindo o compartilhamento e o reuso. Carlos Henrique de Brito Cruz, diretor científico da FAPESP, menciona que essas iniciativas que buscam facilitar a integração e a colaboração entre pesquisadores têm dois resultados principais: o melhor progresso da ciência e a maior eficiência no uso de recursos que custeiam a pesquisa. Da mesma forma que o metabuscador desenvolvido pela USP, existe o RECOLECTA[8], que é um agregador nacional de repositórios de acesso aberto que reúne todas as infraestruturas digitais espanholas nos quais os resultados da investigação são publicados em acesso aberto, com foco em publicações científicas.

No Brasil, a RNP realizou estudos em 2018, com pesquisadores brasileiros para mapear práticas e percepções sobre o acesso aberto a dados de pesquisa no país. O projeto RDP Brasil teve como objetivo explorar o cenário nacional e internacional e apresentar o planejamento de uma investigação que busca uma solução tecnológica para efetivar o AADP dentro de uma perspectiva nacional. Na pesquisa realizada observou-se que 31,71% dos respondentes afirmaram utilizar algum repositório para acessar dados de pesquisa no entanto, ao serem solicitados a especificar o nome do repositório utilizado, os respondentes, na sua grande maioria, informaram bases de dados, portais, redes sociais acadêmicas, repositórios institucionais e revistas científicas, o que indica o desconhecimento por parte dos pesquisadores brasileiros do que seja um repositório de dados de pesquisa (VANZ *et al*., 2018). Deste modo, percebe-se que há uma grande parcela de pesquisadores brasileiros que não sabem exatamente o que é um repositório de dados de pesquisa, como também não conhecem critérios para identificar se são confiáveis ou não.

A RNP iniciou sua participação no 4º Plano de Ação Nacional, com o compromisso de implantar uma infraestrutura federada piloto de repositórios de dados em julho de 2020, juntamente com o IBICT e o CNPq que auxiliaram no processo de execução. O 4º Plano de Ação Nacional é composto por 11 compromissos, os quais foram cocriados com o envolvimento de 105 pessoas, representantes de 88 instituições, sendo 39 organizações da sociedade civil, 39 órgãos da Administração Pública Federal e 10 órgãos das Administrações Públicas Estaduais e Municipais. A criação da infraestrutura federada envolveu três grandes questões: a existência de um portal agregador dos dados, a disponibilidade dos próprios repositórios, e protocolos de interoperabilidade entre o portal agregador e os repositórios. O

[8] https://recolecta.fecyt.es/

Portal Brasileiro de Publicações Científicas em Acesso Aberto (oasisbr[9]) foi escolhido como o mecanismo agregador dos dados e já está disponibilizando os primeiros dados de pesquisa agregados (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2020).

O desenvolvimento de uma infraestrutura e manutenção de repositórios de dados de pesquisa está presente em diversos países e apresenta-se como um desafio tanto em termos de gestão quanto da representação e disponibilização dos conjuntos de dados que estão contidos nesses sistemas. Ainda não há um consenso sobre o que os repositórios devem oferecer para apoiar a publicação de dados, no entanto ao analisar algumas soluções existentes, verificou-se alguns exemplos como, os "generalistas" e "específicos". Silva (2019) menciona que os repositórios generalistas são desenvolvidos especialmente para apoiar a publicação de conjuntos de dados produzidos nos contextos científicos de "cauda longa" devido à heterogeneidade dos possíveis resultados. Já, os repositórios específicos servem como fonte de referência e parâmetro do desenvolvimento científico em determinadas áreas do conhecimento.

Os repositórios de dados de pesquisa são geralmente aplicados no âmbito dos domínios científicos que produzem grandes quantidades de dados, identificadas nos segmentos das *Big Sciences,* tais como Astronomia, Física de Altas Energias, Genética e Ciências Ambientais. Porém, mais recentemente a demanda por repositórios de dados está emergindo no contexto da cauda longa da ciência, ou seja, nos domínios disciplinares em que atividades são desenvolvidas num grande número de laboratórios relativamente pequenos e por pesquisadores individuais que coletivamente produzem a maioria dos resultados (SAYÃO; SALES, 2019).

Sales e Sayão (2020) mencionam em um dos seus artigos exemplos de repositórios recomendados por periódicos, ou seja, apropriados para a curadoria dos dados associados aos itens de literatura científica como, por exemplo, o Repositório Digital *Dryad*, que operacionaliza um repositório digital e que se caracteriza como um repositório multipropósito voltado para uma grande diversidade de tipos de dados de pesquisa. A partir desse ponto, entende-se a necessidade de se implantar repositórios multidisciplinares que viabilizem a interconexão de dados de pesquisa e publicações acadêmicas, expandindo as fronteiras da editoração científica na direção de cobrir integralmente a diversidade dos produtos de

[9] https://oasisbr.ibict.br/vufind/

pesquisa (SALES; SAYÃO, 2020). Nesse cenário, o compartilhamento de dados se torna uma parte essencial da publicação dos resultados de pesquisa. No entanto há o desafio para os editores científicos de oferecer opções de infraestruturas para a publicação de dados que os tornem acessíveis, citáveis e conectados às publicações, além de apoiar a revisão por pares, tendo como referência fluxos de pesquisa mais fidedignos, reprodutíveis e abertos (SALES; SAYÃO, 2020).

É importante caracterizar os repositórios de dados de pesquisa: eles oferecem serviços disciplinares e um espaço de colaboração para os pesquisadores e, "[...] se configuram como sistemas de informação que apoiam os pesquisadores na publicação, preservação e disseminação de suas coleções de dados; e, ao mesmo tempo, são ferramentas críticas para o descobrimento e o acesso às coleções [...]" (SAYÃO; SALES, 2019, p. 80).

Existem alguns critérios para se escolher um repositório conforme as ações relacionadas. Conforme Aventurier e Alencar (2016, p. 12), para dados depositados em conjunto com um artigo de periódico, pode-se escolher um repositório multidisciplinar indicado pelo periódico; ou um repositório exigido pela fundação de financiamento à pesquisa; ou um repositório da sua instituição, explicando aos editores do periódico a escolha ou exigência da instituição afiliada; um repositório da sua temática explicando aos editores do periódico a escolha; ou um repositório de menor custo. Se o conjunto de dados não estiver associado a uma publicação, a melhor prática é depositá-lo em um repositório e publicar um *data paper* descrevendo os dados em um periódico específico para isso denominado *data journal*.

De acordo Tartarotti, Dal'Evedove e Fujita (2019), outra questão de grande importância diz respeito ao gerenciamento de acesso. Ele contempla as decisões a serem tomadas pelo bibliotecário de dados quanto aos tipos de acesso em um repositório de dados de pesquisa, ou seja, pode ter o acesso aberto (qualquer usuário com acesso à Internet pode acessar os dados de pesquisa, que podem ter termos de uso específicos ou indicar o uso apropriado ou impróprio através de uma licença padrão, por exemplo, o *Creative Commons*). Normalmente, é fornecido acesso anônimo aos dados, mas em alguns casos, um nome e endereço de e-mail podem ser solicitados antes que o acesso seja concedido; acesso gerenciado (regras podem ser aplicadas ao uso dos dados). Por exemplo, os usuários podem não apenas precisar se registrar, mas também serem aprovados antes que o acesso seja concedido. A aprovação pode depender do status do usuário, por exemplo, membro da instituição acadêmica, ou de suas respostas a determinadas perguntas, por exemplo, sobre seu

objetivo da pesquisa; ou de acesso seguro (os dados são liberados somente através de mecanismos seguros. Isso pode envolver o acesso a um servidor remoto para executar análises, em vez de baixar dados diretamente; ter a saída verificada pela equipe do repositório de dados de pesquisa para garantir a não divulgação).

Os repositórios precisam operar em infraestruturas confiáveis e estáveis que maximizam os serviços e a disponibilidade. Dawei Lin, em um seminário on-line promovido pela FAPESP em agosto de 2020, menciona que os repositórios devem ganhar a confiança das comunidades que pretendem servir e demonstrar que são confiáveis e capazes de gerenciar adequadamente os dados que custodiam. Além disso, o hardware e o software usados devem ser relevantes e apropriados para a comunidade designada e às funções que um repositório cumpre (FAPESP, 2020). O autor cita o modelo de referência *Reference Model for an Open Archival Information System* (OAIS[10]), que especifica as funções de um repositório para atender às necessidades do usuário. O OAIS aumenta o grau de compreensão dos conceitos para o arquivamento de objetos digitais e promove a aplicação de estratégias para a preservação digital. Além disto, propõe o direcionamento do repositório para os aspectos ligados à manutenção do acesso à informação digital em longo prazo (RLG/NARA, 2007). Publicado inicialmente em 2002, na forma de um padrão recomendado, o OAIS foi, posteriormente aprimorado e atualmente é representado pela norma *Internacional Organization for Standardization* (ISO) 14721:2012 (SANTOS, FLORES, 2019).

Diante do propósito de atribuir confiabilidade aos repositórios, os critérios de Auditoria e Certificação na parte da infraestrutura organizacional, determinados pelo OAIS, são os atributos organizacionais que afetam o desempenho, a prestação de contas e a sustentabilidade. Eles são caracterizados como indicadores do planejamento abrangente do repositório digital, prontidão, capacidade de abordar suas responsabilidades e confiabilidades. Esta parte inclui, mas não restringe aos elementos de governança; estrutura organizacional; mandato ou finalidade; escopo; papéis e responsabilidades; estrutura de políticas; sistema de financiamento; questões financeiras, incluindo ativos; contratos, licenças e passivos e transparência (RLG/NARA, 2007).

Esse modelo fornece uma estrutura e terminologia comum para a preservação e disseminação de ativos digitais e foi base para a construção dos principais documentos de certificação abordados nesta pesquisa. Segundo Dawei Lin *et al*.(2020), a conformidade com

[10] https://public.ccsds.org/Pubs/650x0m2.pdf

o modelo de referência OAIS não garante confiabilidade. A fim de avaliar a confiabilidade, elementos adicionais do repositório precisam ser tratados, incluindo governança, recursos e segurança apropriados. É importante a identificação dos padrões utilizados e como são implementados.

Outros elementos são essenciais para o desenvolvimento de um repositório de dados de pesquisa como, por exemplo: um plano de desenvolvimento de infraestrutura, inventário de software, documentação com as características de conectividade, backup, detalhes sobre armazenamento, entre outros. Além destes elementos destacados, o repositório deve analisar ameaças potenciais, avaliar riscos e criar uma segurança consistente do sistema. Deve descrever cenários de danos com base em ações maliciosas, erro humano ou falhas técnicas que representam uma ameaça para o repositório e seus dados, produtos, serviços e usuários. Dito isto, torna-se importante medir a probabilidade e o impacto de tais cenários, decidir quais níveis de risco são aceitáveis, e determinar quais medidas devem ser tomadas para combater as ameaças ao repositório (CORETRUSTSEAL, 2020b). Os repositórios devem assumir a responsabilidade pela administração de objetos digitais e garantir que os materiais sejam mantidos no ambiente apropriado por períodos. Ou seja, a preservação e o acesso contínuo aos dados são funções explícitas do repositório, portanto deve estar descrito na missão da organização informações referentes a preservação e sobre o acesso aos dados.

Outro requisito importante refere-se aos regulamentos e licenças aplicáveis definidas pelo próprio repositório de dados, bem como quaisquer códigos de condutas geralmente aceitas no setor, relevante para a troca e uso adequado de conhecimento e informação. Além disso, o repositório deve garantir, na medida do possível, que os dados sejam criados, acessados, e usados em conformidade com as normas disciplinares e éticas, ou seja, conforme às disposições éticas e de privacidade que afetam a criação, curadoria e uso dos dados (CORETRUSTSEAL, 2020b).

Os repositórios de dados precisam de financiamento e uma equipe competente que tenha expertise em arquivamento de dados. No entanto, também se entende que a continuidade do financiamento raramente é garantida, e isso deve ser equilibrado com a necessidade de estabilidade. Para isto, deve-se questionar: o repositório é hospedado por uma instituição reconhecida (garantindo estabilidade a longo prazo e sustentabilidade) apropriado para sua comunidade designada? O repositório tem um orçamento suficiente, incluindo recursos de equipe e recursos de TI? O repositório garante que sua equipe tenha acesso a treinamento contínuo e profissional? Esses, entre outros aspectos fazem parte da infraestrutura

organizacional, e que devem manter a comunicação com consultores internos ou um comitê consultivo externo que pode ser preenchido com especialistas técnicos, de curadoria e da ciência de dados (CORETRUSTSEAL, 2020b).

No que se refere a integridade e autenticidade dos dados, o repositório deve fornecer evidências para mostrar que opera uma gestão de dados e metadados, com um sistema adequado. Este requisito cobre todo o ciclo de vida dos dados dentro do repositório, isto significa que, quaisquer alterações intencionais nos dados e metadados devem ser documentadas, incluindo a justificativa e o originador da mudança, ou seja, deve haver o controle de versões, assim como as ações relevantes para a preservação, que devem estar especificadas no plano de preservação (CORETRUSTSEAL, 2020b).

O sistema de descoberta e identificação de dados é a chave para o compartilhamento de dados. Uma vez descobertos, os conjuntos de dados devem ser referenciáveis por meio de citações completas, incluindo identificadores persistentes para garantir que os dados possam ser acessados no futuro. Para este requisito, as respostas devem incluir evidências relacionadas às seguintes questões: O repositório oferece recursos de pesquisa? O repositório mantém um catálogo de metadados pesquisável apropriado de acordo com os padrões (internacionalmente acordados)? Quais sistemas de identificadores persistentes o repositório usa? O repositório facilita a coleta automática dos metadados? O repositório oferece a recomendação de citações de dados? (CORETRUSTSEAL, 2020b).

Logo, pode-se dizer que a confiança é adquirida com o passar do tempo, no caso dos documentos digitais, será necessário comprovar a eficácia do sistema informatizado em questão. A preservação digital exigirá um sistema robusto para satisfazer os requisitos de integridade, autenticidade, atingindo assim a confiabilidade desejada. Assim sendo, entende-se, que não se pode tratar a confiabilidade como um status de “confiável” e “não confiável”, e sim como uma variável que depende do contexto tecnológico onde está situado o acervo. A dúvida em relação à confiabilidade dos repositórios digitais está em saber o que é preciso para atingir esta confiança. Além de definir as políticas institucionais, escolher as estratégias de preservação e implementar um repositório digital em conformidade com o modelo OAIS, é preciso adicionar confiabilidade as ações de preservação digital (SANTOS; FLORES, 2015).

A confiabilidade deve ser avaliada nas medidas de segurança, desde a construção dos repositórios digitais, a fim de garantir que os materiais armazenados permanecerão autênticos em logo prazo (MÁRDERO ARELLANO, 2008). Nesse sentido, a função de avaliação é crítica para verificar se os dados atendem a todos os critérios de seleção e para assegurar um

manejo adequado para sua preservação. A avaliação garante que os dados permanecem relevantes e compreensíveis para a comunidade designada. Frente a esse contexto, em que cada vez mais se implantam novos repositórios de dados de pesquisa, surge também a necessidade de garantia da qualidade. Dito isto, a etapa de avaliação se torna de suma importância desde a fase de planejamento.

2.3 MODELOS DE AVALIAÇÃO PARA REPOSITÓRIOS DE DADOS DE PESQUISA

Na literatura são encontrados diversos estudos relacionados aos Repositórios Institucionais (RI) com a abordagem de alguns fatores essenciais para o sucesso e a amplitude, dentre eles, a avaliação contínua, objetivando identificar pontos frágeis que necessitem de uma intervenção, bem como pontos fortes que possam ser ampliados, socializados e compartilhados com comunidades afins. Analisando estudos realizados em repositórios de publicações científicas, deduz-se que é possível utilizar a experiência no uso de repositórios digitais de publicações científicas e estendê-la para repositórios de dados de pesquisa, pois existe similaridade nas principais funcionalidades para tratar esses objetos documentais (RIBEIRO, 2019).

A avaliação de repositórios é uma etapa fundamental na construção e no desenvolvimento de sistemas de informação, pois permite quantificar a eficiência do sistema de modo que ele acompanhe o dinamismo do objeto de trabalho e dos usuários. Laguardia, Portela e Vasconcellos (2007, p.516) relatam que,

> A avaliação pode ser definida como a aplicação sistemática de procedimentos metodológicos para determinar, a partir dos objetivos propostos e com base em critérios internos e/ou externos, a relevância, a efetividade e o impacto de determinadas atividades com a finalidade de tomada de decisão. Em comum, as definições de avaliação as vêem como um julgamento de valor a respeito de uma intervenção ou sobre qualquer um de seus componentes, tomando como referência um padrão estipulado e cujo propósito é auxiliar os processos decisórios.

Direcionando o tema avaliação aos repositórios de dados de pesquisa, Ribeiro (2019) propõe em seu artigo, um modelo de maturidade que trata de forma integrada os repositórios confiáveis de documentos e os repositórios de dados. Expõe o estudo sobre modelos de maturidade como forma de avaliação da qualidade para repositórios, apresentando um estudo sobre repositórios confiáveis e critérios de avaliação, finalizando a proposta com a inclusão

dos princípios FAIR. O autor faz uma contextualização teórica sobre o surgimento dos modelos de maturidade e demonstra exemplos de modelos mais utilizados e suas respectivas características. Utiliza em seu estudo, o modelo *Capability Maturity Model Integration* (CMMI), cuja estrutura repousa nos conceitos de maturidade e capacidade, como forma de diagnosticar o processo de projeto e construção de sistemas aplicativos nas instituições.

O Modelo CMMI foi desenvolvido pelo *Software Engineering Institute/ Carnegie Mellon University* (SEI/CMU) dos Estados Unidos. A proposta deste modelo é habilitar a transição tecnológica e o aumento da qualidade no processo de desenvolvimento de software aplicativo por meio da adoção das melhores práticas de trabalho. Sua estrutura repousa nos conceitos de maturidade e capacidade, como forma de diagnosticar o processo de projeto e construção de sistemas aplicativos. O modelo conta com duas abordagens similares para melhoria: a capacidade – relacionada à situação dos processos de trabalho. E a maturidade – que perpassa diversas áreas de processo da instituição (RIBEIRO, 2019). Portanto, a possibilidade de avaliar funcionalidades específicas organizadas em degraus, conhecidos como níveis de maturidade, estrutura o caminho para a realização de diagnóstico e as ações para a melhoria da qualidade (GUEDES *et al.*, 2014).

Um estudo realizado em 2017, pelo *4TU.Centre for Research Data*[11] e pelo *TU Delf Research Data Services*[12], mostra que àquela época, poucos repositórios estavam seguindo os princípios FAIR. Os resultados também mostram que alguns princípios FAIR são fáceis para avaliar e outros são mais imprecisos (AVENTURIER, 2017a). A partir dos seguintes aspectos citados por Assis (2019) que estão relacionados a cada princípio FAIR é possível detectar, através de uma avaliação, se os princípios estão sendo aplicados nos repositórios de dados de pesquisa:

a) o repositório deve utilizar um identificador único persistente tanto para o conjunto de dados quanto para os metadados (ex: DOI, ARK, RRID, PID);
b) o conjunto de dados deve ser descrito por metadados ricos o suficiente para que, uma vez indexados em um mecanismo de busca possam ser encontrados mesmo sem o seu identificador único persistente;

[11] https://data.4tu.nl/info/en/about/organisation/
[12] https://www.tudelft.nl/en/library/research-data-management/

c) como não se pode prever que os dados e seus metadados estejam sempre juntos, a associação entre eles deve ocorrer pela inclusão do identificador persistente dos dados nos metadados;
d) para que os dados sejam encontrados, seus metadados devem ser indexados em mecanismos de busca (*search engine*), que possibilitem aos computadores e usuários encontrá-los com facilidade;
e) com o identificador persistente do conjunto de dados e/ou de seus metadados, o usuário deverá recuperá-los mais facilmente por meio de protocolos de comunicação padronizados (ex: HTTP ou FTP);
f) independente de licenciamento dos dados e dos metadados, o protocolo de comunicação usado para dar acesso a eles deve ser aberto, gratuito e passível de ser implementado por qualquer interessado (ex: HTTP ou FTP);
g) dependendo das restrições de acesso aos dados e/ou metadados, um mecanismo de autenticação e autorização para o acesso deve ser liberado pelo protocolo de comunicação (ex: os repositórios confiáveis oferecem essa opção);
h) é preciso existir um conjunto de estratégias de preservação para dados e metadados. Os metadados devem ser sempre acessíveis, possibilitando a criação de índices para o conjunto de dados atuais vigentes e aqueles não mais disponíveis;
i) para que se possa representar dados e metadados devem ser adotadas linguagens de representação do conhecimento que sejam padronizadas, acessíveis e amplamente aplicáveis (ex: RDF, XML, DICOM, etc.);
j) dados e metadados devem possuir referências a vocabulários e/ou ontologias que os descrevem. Deve-se garantir que esses também sigam os princípios FAIR;
k) é necessário referenciar o conjunto de dados, possibilitando que aqueles gerados a partir de outros conjuntos, sejam interligados, assegurando a ligação semântica entre eles;
l) prover metadados descritos com alto nível de detalhes que permita ao pesquisador avaliar a possibilidade do seu reuso bem como adequação às suas necessidades;
m) é fundamental que o responsável pelos dados e metadados defina explicitamente quem pode ter acesso a eles, com que finalidade e sob quais condições. Essas informações são definidas por meio de licenças de uso;
n) especificar a proveniência dos dados é importante não só para que o pesquisador possa avaliar a utilidade dos dados ou metadados, mas também para que possa atribuir o devido crédito a quem produziu, manteve ou editou esses dados;

o) além de atender aos padrões específicos da área de cada comunidade deve-se dar atenção as boas práticas de arquivamento e compartilhamento específicos da área de pesquisa.

No artigo publicado por Kim (2018), é relatado que poucos estudos referentes aos requisitos para repositórios de dados foram encontrados. Percebe-se que ainda são incipientes estudos que abordem a avaliação desse tipo de repositório, no entanto, para que se configurem como sistemas de informação que possibilite a disponibilização, o compartilhamento e o reuso de dados de pesquisa, se faz necessário que os repositórios de dados sejam ambientes multifacetados, com condições bem definidas para uma gestão de dados compatível com as complexidades impostas pelos dados de pesquisa e que disponham de serviços que os tornem espaços de colaboração e interação entre a comunidade de pesquisadores (SAYÃO; SALES, 2019).

Frente ao protagonismo dos dados de pesquisa, Sayão e Sales (2018) mencionam que é necessário refletir sobre quais parâmetros podem dimensionar os requisitos de uma plataforma de gestão de dados. Dito isto, apresentam em seu artigo intitulado “Subsídios para a construção de um modelo de avaliação de sistemas de gestão de dados de pesquisa”, os elementos essenciais para a composição de um modelo de avaliação de sistemas de informação acadêmica voltados à gestão de dados de pesquisa, em atenção aos parâmetros técnicos, gerenciais e organizacionais requeridos. A intenção do estudo realizado por Sayão e Sales (2018) foi alinhar alguns parâmetros necessários ao ajuste dos modelos de avaliação dos sistemas de informação às exigências do protagonismo dos dados de pesquisa, conceituando os itens mais importantes, reconhecidos pela literatura, como peças fundamentais para compor os sistemas colaborativos de gestão de dados de pesquisa. Os autores argumentam que existem duas áreas de requisitos na gestão dados científicos.

> A primeira está relacionada com as infraestruturas tecnológicas – sistemas, normas e protocolos - necessárias para assegurar a coleta, preservação e acesso aos dados, e ainda a disponibilidade de serviços de amplo espectro. Além do mais, para as infraestruturas informacionais subjacentes às plataformas, a aplicação de padrões técnicos e semânticos, como ontologias, é algo crítico. Mesmo a noção de qualidade é afetada pela proveniência, integridade, autenticidade e propósito dos dados que, por sua vez, têm uma dependência essencial aos esquemas de representação – metadados e documentação. A segunda área de requisitos considera os aspectos políticos, legais e éticos decorrentes do acesso e reuso dos dados além do contexto

> inicial para que foram gerados, posto que a reutilização é fortemente condicionada por arcabouços legais e éticos e por código de conduta específicos da cultura de cada disciplina, como comissão de ética, políticas de consentimento, propriedade intelectual e tradição de compartilhamento e políticas de incentivo e de financiamento (SAYÃO; SALES, 2018, p. 84).

Os autores explicitaram os principais conceitos que formam o quadro 1 de requisitos que podem contribuir para a composição de diferentes modelos de avaliação para plataformas de gestão de dados de pesquisa.

Quadro 1 – Requisitos de uma plataforma de gestão de dados

Infraestrutura	
Tipo de plataforma	- Disciplinar -Multidisciplinar
Serviços	-Serviços informacionais - Serviço de referência de dados e consultoria - Aquisição/desenvolvimento de coleções de dados - Competência informacional para pesquisadores - Identificações persistentes - Citação padronizada dos dados - Controle de versões - Publicações dos dados - Formato de arquivos - Tamanho do arquivo - Link com outros recursos -Disponibilidade dos dados para revisão por pares - Depósito em outras plataformas - Informação de representação: (metadados e documentação) - Serviços computacionais - Reformatação e limpeza dos dados - Segurança dos dados - Análise dos dados - Visualização dos dados - Análise exploratória - Mineração de dados - Modelagem por computador - Simulação interativa e realidade virtual - Workflow científico - Serviços realizados pelo pesquisador - Revisão por pares - Qualidade dos dados - Gestão dos dados no laboratório
Interoperabilidade	-Integração dos sistemas de repositório com os sistemas de publicação - Coleta automática via protocolo OAI-PMH - Submissão de dados para múltiplos sistemas via o protocolo SWORD - Empacotamento via padrão Bagit - Acesso aos conteúdos por meio de APIs

	- Integração da plataforma de dados com os sistemas de arquivamento confiável -Exportação de metadados
Requisitos conjunturais, políticos e administrativos	- Política do repositório - Institucionalização - Reconhecimento pela comunidade científica -Estabilidade e persistência - Visibilidade - Presença nos diretórios, agregadores e dispositivos de busca - Licenças - Licenças de depósito - Termos de uso -Licença de acesso - Tempo de embargo - Licença do material associado ao dataset - Certificação
Curadoria de dados de pesquisa	-Adição às coleções de dados de metadados, versionamento, identificação persistente, arquivamento. - envolvem links semânticos com outros materiais publicados, anotações estruturadas baseados em ontologias, entre outros.
Preservação digital de longo prazo	- Dispor de uma política explicita de preservação digital, que considere parâmetros arquivísticos tais como proveniência e autenticidade dos dados que não podem ser regerados e estejam conectadas à sistemas confiáveis de preservação fundamentados no modelo de referência ISO/OAIS.
Custo	- A operação dos repositórios de dados de pesquisa pressupõe um custo considerável - tanto monetário quanto custos de outra natureza - para as instituições que abrigam estas plataformas.

Fonte: Adaptado de Sayão e Sales, 2018.

Os autores propõem de forma sistemática as possibilidades tecnológicas, padrões e práticas que se desdobram em unidades conceituais, que podem se ajustar às diferentes demandas de acesso, arquivamento, preservação, curadoria e demais serviços, e aos diferentes tipos de dados. Contudo, no estudo desenvolvido por Kim (2018), os requisitos funcionais para repositórios de dados de pesquisa podem ser amplamente agrupados em 13 categorias: metadados; identificadores; gerenciamento de autenticação e permissão; acesso a dados; suporte de política; publicação; envio/ingestão/gerenciamento; configuração de dados; localização; integração; preservação e sustentabilidade; interface do usuário; dados e qualidade do produto.

Nesta mesma linha, Sanchez, Vechiato e Vidotti (2019), desenvolveram uma pesquisa cujo objetivo é apresentar as vantagens de utilizar os critérios inseridos nos conceitos de Encontrabilidade da Informação (EI) e da Arquitetura da Informação, mediante o processo de projeto, implementação e/ou avaliação de repositórios de dados de pesquisa. A EI é um termo

introduzido pelo autor Morville (2005) em um estudo, que visa atender as necessidades informacionais dos sujeitos em ambientes informacionais, contribuindo para que as informações disponíveis sejam encontradas rapidamente com o mínimo de esforço possível.

Para que a apropriação da informação aconteça, é importante que o conteúdo esteja efetivamente representado através da descrição de recursos informacionais com o uso de metadados, que são um dos principais Atributos da Encontrabilidade da Informação (AEI) (SANCHEZ; VECHIATO; VIDOTTI, 2019), e essenciais para a padronização e para a interoperabilidade, que como critério de avaliação, pode estar relacionada à capacidade do modelo do repositório de dados de trocar informações com outros sistemas de forma padronizada, tendo como objetivo mais perceptível o aumento no nível de encontrabilidade dos conteúdos na medida em que eles se tornam disponíveis através de múltiplas rotas (SAYÃO; SALES, 2019). Morville (2005) conceitua encontrabilidade como a qualidade de ser localizável ou navegável; o grau no qual um determinado objeto é facilmente descoberto ou localizado e; o grau no qual um sistema ou ambiente suporta a navegação e recuperação.

O conceito de encontrabilidade apresentado por Morville (2005) está relacionado a uma abordagem mais técnica que científica, o que faz necessário reconfigurá-lo para ser discutido e incorporado na Ciência da Informação. Desta forma, entende-se que a ‘Encontrabilidade da informação’ sustenta-se fundamentalmente nas funcionalidades de um ambiente informacional e nas características dos sujeitos psico-sociais. Relacionada aos processos que compõem o fluxo infocomunicacional, desde a produção até a apropriação da informação, a encontrabilidade da informação deriva-se dos princípios da Arquitetura da Informação e da Mediação Infocomunicacional e tem como elemento fundamental a Intencionalidade dos sujeitos nas ações informacionais empreendidas durante o processo de comunicação que subsidiam a elaboração de técnicas e de tecnologias para a organização, representação da informação e recuperação da informação (VECHIATO; VIDOTTI, 2014).

Na obra de Vechiato e Vidotti (2014) são apresentados os treze (13) atributos de AEI que são caracterizados como aspectos práticos do conceito e que devem ser levados em consideração no projeto, implementação ou avaliação de um ambiente e são entendidos como características que potencializam as possibilidades de encontro da informação pelos sujeitos num sistema informacional. De acordo com Sanchez, Vechiato e Vidotti (2019), a EI traz contribuições significativas para o projeto e para a implementação de Repositórios de Dados, e a partir de avaliações é possível detectar problemas relacionados a EI.

Assim sendo, a avaliação de repositórios permite mensurar a eficiência, garantindo que se acompanhe o dinamismo dos objetivos de trabalho e dos usuários. Na literatura foram propostos vários modelos de avaliação, no entanto, a falta de padronização e clareza na definição de critérios avaliativos são fatores que dificultam a comparação dos modelos e tornam os resultados e as análises muitas vezes subjetivos (LAMEIRA, 2016). A partir dos modelos de avaliação e dos critérios e atributos abordados por diversos autores se faz necessária a realização de um levantamento e perpassar por construções teóricas para, então sim, obter uma padronização de critérios e requisitos.

Em 1996, um grupo internacional de representantes de arquivos e bibliotecas nacionais, universidades, indústria, escritórios de publicação e outras organizações governamentais e do setor privado pela primeira vez articulou a necessidade de Certificação de Repositórios Digitais Confiáveis- *Trusted Digital Repository* (TDRs). Daí em diante, vários padrões para TDRs foram desenvolvidos em todo o mundo, no entanto não há o conhecimento se os repositórios digitais são realmente melhores em preservar informações digitais após a certificação do que eram antes (DONALDSON, 2020).

Como exemplo de repositório que passou por processos de avaliações, pode-se citar o *Inter-Univertity Consortium for Political and Social Research* (ICPSR), que foi um dos primeiros seis repositórios de dados a receber o Selo *Data Seal of Approval* (DSA) em 2011, que caracteriza-se por ser um dos padrões mais amplamente usados no ano citado. Os desenvolvedores deste padrão articularam os sete principais benefícios da aquisição do DSA: confiança das partes interessadas; melhorias em comunicação; melhoria nos processos; transparência; diferenciação entre os repositórios; aumento da conscientização sobre preservação digital; e menos trabalho e tempo intensivo (DONALDSON *et al*., 2017). O estudo realizado por Donaldson *et al*. (2017) examinou os benefícios de adquirir DSAs a partir do ponto de vista de quem os possui e constatou-se que os participantes relataram que o uso de informações sobre o DSA em pedidos de financiamento foi utilizado para mostrar aos financiadores que trata-se de um repositório confiável, oferecendo credibilidade atestando a qualidade do repositório.

O ICPSR também conquistou a certificação *Word Data System* (WDS) em 2013. O DSA foi desenvolvido em 2008 pela *Data Archiving and Networked Services* (DANS) em resposta ao requisito de seus financiadores, as duas organizações científicas holandesas KNAW e NWO, para criar um Selo de Aprovação de Dados que ajuda a garantir que os dados arquivados possam ser encontrados e usados no futuro (CORETRUSTSEAL, 2020a). O WDS

é um órgão interdisciplinar do *International Science Council* (ISC) e sua missão é promover a administração a longo prazo, o acesso universal aos dados científicos e serviços de dados (WORLD DATA SYSTEM, 2020). Os critérios para o Selo de Aprovação de Dados foram alinhados com as diretrizes nacionais e internacionais para arquivamento de dados digitais como o Nestor, a TRAC, e o DRAMBORA.

O ICPSR, em 2006, foi um caso de teste para Auditoria e Certificação de Repositórios Confiáveis – *Trustworthy Repository Audit & Certification* (TRAC), que apresentava um conjunto de critérios usados como referência para a certificação de repositórios digitais em conformidade com o OAIS. No entanto, em virtude de sua descontinuidade, em 2012 tornou-se a ISO 16363 (INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH, 2020).

O referencial normativo preconizado pela ISO 16363 (2012) define uma prática recomendada para avaliar a confiabilidade de toda a gama de repositórios digitais. Estabelece a documentação necessária para a realização de um processo de auditoria, requisitos mínimos para os auditores e, deste modo, delinear o processo de certificação. Assim sendo, estabelece as metodologias apropriadas para determinar a robustez e a sustentabilidade de um repositório digital (CARVALHO *et al*., 2014).

O ICPSR caracteriza-se também por ser um repositório certificado pela *CoreTrustSeal* que se constitui por ser uma organização internacional, comunitária, não governamental e sem fins lucrativos, que promove infraestruturas de dados sustentáveis e confiáveis. Oferece a qualquer repositório de dados uma certificação de nível básico com base no catálogo e nos procedimentos dos Requisitos de Repositórios de Dados Confiáveis DSA e WDS. Esse catálogo universal de requisitos reflete as principais características dos repositórios de dados confiáveis e é o culminar de um esforço cooperativo entre a DSA e o WDS amparado pela *Research Data Alliance (*RDA*)* para mesclar suas certificações de repositórios de dados (CORETRUSTSEAL, 2020a).

Atualmente a *CoreTrustSeal* substitui a certificação DSA e a certificação WDS, ou seja, ambas foram incorporadas a *CoreTrustSeal*. É cada vez mais provável que os financiadores nacionais e internacionais determinem políticas abertas de gerenciamento de dados e comecem a exigir o armazenamento e acessibilidade de dados a longo prazo. Se queremos compartilhar dados, precisamos armazená-los em um repositório de dados confiável. A adoção dos Requisitos de Repositórios de Dados Confiáveis *CoreTrustSeal* por muitos repositórios de dados serve como um exemplo das melhorias feitas para garantir que

seus recursos atinjam as propriedades dos Princípios TRUST (*Transparency, Responsibility, User focus, Sustainability and Technology*).

Desta forma, quando repositórios de dados, financiadores e criadores de dados adotam os Princípios FAIR e implementam os Princípios TRUST, os usuários do repositório se beneficiam diretamente por meio de recursos contínuos e aprimorados para o uso eficiente e eficaz dos dados. Os Princípios TRUST se constituem como um meio para agilizar a comunicação com todas as partes interessadas e serve como um guia de orientação para que os repositórios possam demonstrar sua transparência, responsabilidade, foco no usuário, sustentabilidade e tecnologia (DAWEI LIN *et al*., 2020). Em um estudo realizado por Silva *et al.* (2021) investigou-se de que forma os itens constitutivos dos Princípios TRUST podem ser contemplados em repositórios de dados e concluiu-se que podem ser empregados como uma ferramenta útil de certificação, visto que direcionam o seu foco para aspectos bastante específicos dos repositórios. Dando continuidade a essa proposta e fomentando a discussão sobre o assunto, realizou-se em 7 julho de 2020 o *Trust Principles Mini Symposium*, evento ocorrido na modalidade não presencial e transmitido pela plataforma Zoom, no qual os palestrantes forneceram uma introdução aos princípios TRUST, juntamente com uma visão geral dos conceitos e implementações (RDC, 2020).

A obtenção da certificação e a conclusão de auditorias por muitos repositórios digitais demonstra o desejo de que os repositórios sejam considerados confiáveis. Dito isto, a certificação realizada pela *CoreTrustSeal* (2020a), envolve um processo minimamente intensivo, pelo qual os repositórios de dados fornecem evidências de que são sustentáveis e confiáveis. Os repositórios investem em esforços para a realização da certificação pois há o aumento da confiança das partes interessadas no repositório como, os financiadores, organizações, editores entre outros; há o aumento da consciência sobre a preservação digital; melhora a comunicação dentro do repositório; melhora os processos; garante a transparência e diferencia o repositório de outros (DILLO, 2018). A missão do *CoreTrustSeal* continua sendo a de oferecer uma certificação básica/principal que possui uma estrutura escalonada de melhorias para as organizações que buscam um caminho mais rigoroso (L'HOURS; KLEEMOLA; LEEUW, 2019). O CoreTrustSeal entende que a revisão contínua das políticas, processos e procedimentos é necessária para garantir uma eficaz administração dos objetos digitais, portanto os requisitos são revisados e atualizados, se necessário, a cada três anos. Desta forma, no início de julho de 2022, foi realizada uma atualização dos requisitos e

publicados os requisitos preliminares do CoreTrustSeal 2023-2025 (CORETRUSTSEAL, 2022).

Um repositório primeiro realiza uma autoavaliação interna, que é então revisada pelos colegas da comunidade. Essa abordagem da comunidade garante uma atmosfera inclusiva na qual o repositório candidato e os revisores interagem estreitamente. Além dos benefícios externos, como o fortalecimento da confiança das partes interessadas, o aprimoramento da reputação do repositório e a demonstração de que o repositório está seguindo as boas práticas, a certificação principal fornece vários benefícios internos ao repositório. Um exemplo de caso (entre outros), no qual utilizou-se o *CoreTrustSeal*, foi o *National Institute of Standards and Technology* (NIST). Em fevereiro de 2020 foi realizada uma autoauditoria/autoavaliação em preparação para a apresentação de um pedido de certificação. Os resultados foram animadores, ou seja, dos 16 requisitos, o repositório teve níveis de conformidade 4 para 11 dos requisitos; os 5 requisitos restantes foram classificados em um nível 3 (MEDINA-SMITH, 2021).

Especificamente, a certificação principal oferece uma referência para comparação e ajuda a determinar os pontos fortes e fracos de um repositório. A conclusão de uma autoavaliação é muito útil, mesmo que um repositório não deseje solicitar a certificação principal, pois permite uma avaliação dos procedimentos internos do repositório, que podem ser examinados com relação aos critérios relevantes e atualizados quando necessário. O status atual do repositório é, portanto, tornado aparente e pode servir para credenciamento. Ao enviar o pedido de revisão, os procedimentos e a documentação do repositório são avaliados adicionalmente por profissionais externos, levando em consideração os objetivos e o contexto específicos; assim, o repositório obtém insights independentes sobre como ele pode evoluir e amadurecer para aumentar ainda mais sua confiabilidade. Finalmente, a certificação principal oferece uma base sólida para o repositório solicitar uma certificação de nível superior no futuro (CORETRUSTSEAL, 2020a).

Mendeley data; Edinburgh data share; e *DataverseNO* são alguns exemplos de repositórios reconhecidos com a certificação *CoreTrustSeal*, entre os 126 repositórios listados no site[13]. Também se encontra disponível no site um documento para download com requisitos para repositórios de dados confiáveis da *CoreTrustSea*l. Todos os requisitos são

[13] https://www.coretrustseal.org/why-certification/certified-repositories/. Levantamento realizado em jun. de 2022.

obrigatórios e avaliados como itens independentes. Cada requisito é acompanhado por um texto de orientação que descreve as informações e evidências que os candidatos devem fornecer para permitir uma revisão objetiva.

A partir dos exemplos encontrados na literatura percebe-se a importância da certificação de repositórios. Mesmo que o instrumento *CoreTrustSeal* esteja sendo usado em muitos repositórios de dados, existe uma diversidade de repositórios e as diferenças entre eles exigem a aplicação de abordagens individuais para atingir a conformidade com os requisitos de confiabilidade o que irá exigir a evolução desses instrumentos (DOWNS, 2021).

Neste cenário, esta questão foi tratada também pelo grupo de trabalho da *Research Libraries Group* (RLG) e *Online Computer Library Center* (OCLC) em 2002, quando iniciaram uma colaboração para estabelecer atributos de um repositório digital para organizações de pesquisa, construindo e incorporando o padrão internacional do OAIS. (THOMAZ, 2007). Thomaz (2007, p. 84) menciona que "[...] a certificação se tornou um componente-chave para repositórios digitais contemporâneos". Assim, é a partir desse ponto que um repositório digital poderá, ao longo do tempo, obter a confiança dos pesquisadores que o acessam.

Desta forma, o uso de um ferramental para auxílio ao processo de avaliação também deve ser considerado. Com o objetivo de incentivar a abertura e compartilhamento de dados de pesquisa e conferir maior credibilidade perante os pesquisadores, algumas organizações estão investindo na elaboração de instrumentos que permitam avaliar e contribuir para o desenvolvimento de repositórios de dados de pesquisa confiáveis e de qualidade. Por meio das avaliações é possível identificar aspectos deficientes e as características de sucesso.

Segundo Austin *et al.* (2015), o Comitê de Padrões e Interoperabilidade da *Research Data Canadá* (RDC), que se caracteriza por ser uma organização não governamental colaborativa que promove o acesso e a preservação de dados de pesquisa canadenses, pesquisou 32 plataformas de dados on-line. Neste estudo, foi desenvolvido uma lista de verificação para comparar os critérios e recursos entre as plataformas. A pesquisa revelou uma heterogeneidade de recursos e serviços entre as plataformas; o uso não padronizado de termos; conformidade desigual com os padrões relevantes; e uma escassez de repositórios certificados. Foi identificado que apenas 20% (6 de 32 repositórios) foram certificados ou avaliados.

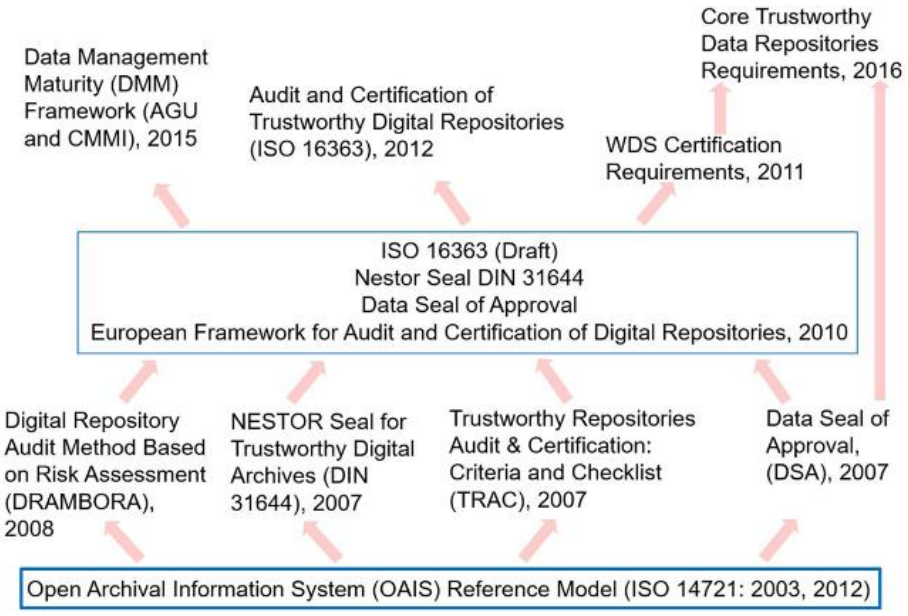
Constata-se a partir da literatura que vários instrumentos surgiram nas últimas décadas para avaliar a confiabilidade dos repositórios de dados. Desde o desenvolvimento inicial do Modelo de

Referência OAIS pelo *Consultative Committee for Space Data Systems* (CCSDS) em 2002, e subsequente publicação pela *International Standards Organization* (ISO), os repositórios de dados têm se esforçado para atender aos requisitos para se tornarem confiáveis. Após a publicação inicial da ISO e do Modelo de Referência OAIS, vários instrumentos foram desenvolvidos para medir a confiabilidade dos repositórios de dados, uma vez que o Modelo de Referência OAIS é uma estrutura e não um instrumento para medir ou avaliar a confiabilidade do repositório de dados. Para sanar tal questão, surgiram padrões que abordam a auditoria do ambiente OAIS. O modelo OAIS tornou-se a norma *International Organization for Standardization* (ISO) 14721:2012, então traduzida para o Brasil pela Associação Brasileira de Normas Técnicas (ABNT), como Norma Brasileira Recomendada (NBR), sendo assim, ABNT/NBR 15472:2007, Sistema Aberto de Arquivamento de Informação (SAAI).

Os instrumentos de avaliação inicial desenvolvidos incluem o DSA; o Selo Nestor para Arquivos Digitais Confiáveis, publicado como o padrão *Deutsches Institut für Normung* (DIN); TRAC, e DRAMBORA, o Método de Auditoria do Repositório Digital com Base na Avaliação de Risco, bem como outros (DOWNS, 2021).

Enquanto o CCSDS e a ISO conduziam as revisões do projeto da norma ISO 16363, Auditoria e Certificação de Repositórios Digitais Confiáveis, que também emanava do Modelo de Referência OAIS, a *European Framework for Audit and Certification of Digital Repositories* foi assinada em 2010 por representantes da DSA, DIN e ISO 16363 para oferecer orientação para repositórios de dados em busca de certificação. Depois que muitos repositórios foram certificados em conformidade com os requisitos WDS e com os requisitos DSA, esses dois instrumentos de avaliação foram mesclados para criar os requisitos de repositórios de dados confiáveis que são oferecidos pelo *CoreTrustSeal*, conforme já mencionado neste estudo. A figura 1 mostra a evolução de diversos instrumentos desenvolvidos (DOWNS, 2021).

Figura 1 – Evolução de diversos instrumentos desenvolvidos e utilizados, internacionalmente, para avaliação de repositórios de dados confiáveis.



Fonte: Downs (2019).

Entende-se que as avaliações realizadas tanto em repositórios de dados de pesquisa em funcionamento quanto em processo de implementação, podem auxiliar no desenvolvimento de melhores práticas, proporcionando qualidade a estes repositórios. Dito isto, desenvolvedores/implementadores de repositórios de dados de pesquisas brasileiros, após realizar avaliações internas, poderão desenvolver uma gestão de dados sólida. Com o conjunto de critérios/requisitos proposto neste estudo, será possível realizar autoavaliações baseadas nos instrumentos de avaliações a nível internacional evitando uma heterogeneidade de recursos e a conformidade desigual com os padrões relevantes.

Para dar embasamento e auxiliar no desenvolvimento de um conjunto de critérios/requisitos para avaliação, foram identificadas na literatura e sites de organizações, algumas iniciativas de avaliação de repositórios já consolidados. Foram incluídas as iniciativas de repositórios digitais de publicações científicas, como também de RIs. Como já foi mencionado neste estudo, algumas características desse tipo de repositório podem ser adaptadas e utilizadas nos repositórios de dados de pesquisa.

Referente as pesquisas iniciais foi possível identificar que as iniciativas possuem critérios similares e complementares. Para o desenvolvimento de um conjunto de

critérios/requisitos será necessário um aprofundamento nos estudos e nas ferramentas existentes para que seja possível coletar e estudar os padrões internacionais, identificar semelhanças e diferenças entre os instrumentos abordados na literatura e estabelecer os critérios que serão utilizados para avaliar os repositórios de dados de pesquisa.

No site do Re3data.org é possível filtrar a pesquisa de repositórios de dados utilizando-se o recurso "*Certificates*", ou seja, o diretório recupera os repositórios que possuem algum tipo de certificação. Com esta filtragem é possível observar que os 14 repositórios brasileiros listados no diretório re3data.org não possuem certificações[14].

Algumas iniciativas voltadas à certificação de repositórios de dados científicos foram apresentadas no *Second Latin American and Caribbean Scientific Data Management Workshop*[15], que aconteceu em fevereiro de 2021, promovido virtualmente pela Academia Brasileira de Ciências (ABC), a Fapesp, o WDS e a RDA, onde estavam presentes pesquisadores como, Rorie Edmunds, Dawei Lin, John Crabtree, Ingrid Dillo, Mustapha Mokrane, e Maja Dolinar, membros da WSA, RDA, DANS entre outras instituições. O workshop foi concentrado em discussões sobre as melhores práticas de gerenciamento de dados para repositórios e sobre novas tendências e perspectivas para sistemas de dados científicos como, por exemplo, a criação e manutenção de repositórios de dados confiáveis. Conforme o palestrante, Rorie Edmunds – diretor executivo do WDS (2021), "A certificação promove a confiança, que está no cerne do armazenamento e compartilhamento dos dados". Portanto, a fim de avaliar a confiabilidade dos repositórios de dados, foram criados nos últimos anos os serviços de certificação (ALISSON, 2021).

Um memorando – *Memorandum of Understanding* (MoU) de 2010, foi assinado em colaboração entre três grupos: o *Data Seal of Approval*, o *Repository Audit and Certification Working Group do CCSDS* e o *DIN Work Group Trustworthy Archives Certification*, para apoiar o trabalho sobre padrões para repositórios digitais confiáveis. Esses grupos de trabalho possuíam como objetivo a certificação de repositórios digitais, implementando mecanismos para garantir a colaboração na criação de uma estrutura integrada para auditoria e certificação. O MoU foi assinado como parte de uma série de iniciativas patrocinadas pela Comissão Europeia sobre auditoria e certificação de repositórios digitais confiáveis. O *European Framework for Audit and Certification* (Marco Europeu para Auditoria e Certificação) garante que repositórios digitais possam receber diferentes certificados, permitindo que os

[14] Consultado em 11 maio de 2022.
[15] https://fapesp.br/eventos/wds

repositórios mostrem um dos três símbolos (a serem aprovados em Certificação Básica, Certificação Estendida e Certificação Formal) (TRUSTED DIGITAL REPOSITORY.EU (20-?) O primeiro nível (Certificação Básica, concedida pelo *CoreTrustSeal*) requer alguns dias para a aplicação. Os dois últimos níveis, Certificação Estendida, (que exige a certificação *CoreTrustSeal* e uma auto auditoria revisada por membro externo baseada na ISO 16363 ou Nestor Seal) e Certificação Formal (e concedida a repositórios que, além da certificação básica, obtém auditoria externa completa e certificação com base na ISO 16363 ou equivalente a DIN 31644) apresenta padrões de auditoria para repositórios digitais confiáveis e exige vários meses para coletar informações de forma detalhada.

Os instrumentos selecionados neste estudo refletem as características desejáveis de um repositório de dados confiável e que seriam adequados ao contexto brasileiro, ou seja, devido ao estado incipiente em que se encontram os repositórios de dados de pesquisa brasileiros, deve-se considerar os requisitos que se encontram em instrumentos que abrangem a certificação básica ou estendida.

Embora alguns modelos publicados na literatura sugerem que a implementação de um Repositório Digital Confiável requer a conformidade com os requisitos do modelo OAIS – ISO 14721 e que sejam devidamente auditados e certificados com o padrão *Audit and Certification of Trustworthy Digital Repositories* (ACTDR) – ISO 16363 (SANTOS, 2018), deve-se considerar as peculiaridades de um repositório de dados de pesquisa. Cabe ressaltar que o ACTDR se firma como principal padrão para auditoria externa, tornando-se a ISO 16363:2012.

Por possuir uma terminologia baseada no modelo OAIS, e possuir os princípios FAIR implícitos nos requisitos (que estabelece condições favoráveis ao compartilhamento e reuso de dados científicos, a partir do tratamento especial voltado para os dados e metadados), utilizou-se neste estudo o *CoreTrustSeal*, que oferece a qualquer repositório de dados uma certificação de nível básico e reflete as principais características dos repositórios de dados confiáveis. Conforme já mencionado, a adoção dos Requisitos de Repositórios de Dados Confiáveis *CoreTrustSeal* serve para garantir que os recursos atinjam as propriedades dos Princípios TRUST. O conjunto de critérios/requisitos que será elaborado também terá como base o ACTDR por ser considerado um padrão para auditoria e certificação em conformidade com os requisitos do modelo OAIS.

Percebe-se, através da literatura e pelo levantamento realizado neste estudo, que caracterizam-se por ser documentos conceituados e mais utilizados para a realização de avaliações relacionadas a confiabilidade e preparo para desejável certificação. Além disso,

nota-se dificuldades na interpretação devido ao idioma inglês. Dito isto, optou-se em elaborar um conjunto de critérios/requisitos para avaliação de repositórios de dados de pesquisa baseado no *CoreTrustSeal* e ACTDR e nos princípios FAIR e TRUST para colaborar no desenvolvimento de repositórios de dados de pesquisa brasileiros confiáveis.

Na seção de procedimentos metodológicos serão demonstradas as técnicas e métodos para o desenvolvimento da pesquisa.

3 PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa constitui-se num levantamento documental cujo objeto de estudo são os critérios/requisitos para repositórios confiáveis de dados de pesquisa. Na literatura consultada, foi realizada uma primeira aproximação para compreensão do cenário que envolve a avaliação de repositórios de dados de pesquisa no contexto nacional e internacional.

Quanto a abordagem, esta pesquisa caracteriza-se como qualitativa, para compreensão e aprofundamento de uma determinada questão; e configura-se como básica, que objetiva gerar novos conhecimentos para o avanço da ciência, isto é, busca traçar um panorama de uma determinada realidade. Desta forma, a partir da coleta de documentos na literatura, em sites de organizações e instituições foi possível observar, analisar e comparar os modelos de avaliações existentes para que fosse possível criar um conjunto único de critérios/requisitos, a partir dos instrumentos internacionais selecionados, que apoie os repositórios de dados de pesquisa brasileiros nas autoavaliações e na melhoria geral, os direcionando para o caminho da certificação, tornando-os confiáveis.

Para a realização dos procedimentos metodológicos, as atividades foram planejadas e divididas em 3 etapas: a pesquisa bibliográfica e documental constitui-se na primeira etapa, sendo efetuada em nível nacional e internacional. Este levantamento permitiu a identificação de autores que exploraram, a partir de suas publicações, o tema avaliação de repositórios.

A segunda etapa, equivale a seleção, leitura e interpretação da documentação que resultará na construção da base teórica para a discussão e comparação dos diferentes instrumentos para avaliação identificados na literatura e nos sites de organizações.

Durante a terceira etapa, foi executada a análise, padronização e sistematização dos critérios/requisitos de cada instrumento selecionado neste estudo. As seções seguintes apresentam o detalhamento de cada etapa.

3.1 PRIMEIRA ETAPA

Foi realizada uma pesquisa bibliográfica que forneceu embasamento teórico e elementos coerentes para o entendimento sobre o estado atual em que se encontra o tema. Primeiramente, o material bibliográfico foi levantado no âmbito da Ciência da Informação, abrangendo o período de 1972 a 2018 e serviu para dar suporte teórico ao desenvolvimento deste estudo. Todo o encaminhamento desta pesquisa foi embasado em conceitos e estudos

anteriores sobre os assuntos em questão e para verificar a inovação do tema na área de aplicação. Para o levantamento das informações, foram utilizadas fontes bibliográficas referenciais como a *Library and Information Science Abstract* (LISA); repositórios de teses e dissertações, como a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), a Biblioteca Digital de Teses e Dissertações da USP; portal de periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); Google Scholar; bibliotecas virtuais de universidades nacionais e estrangeiras; e sites de normas técnicas como a Associação Brasileira de Normas Técnicas (ABNT) e a *International Organization for Standardization* (ISO). Sobre as fontes recuperadas, pode-se citar: livros, normas técnicas, teses, dissertações e artigos científicos. A busca foi realizada nos campos palavras-chave e utilizou-se as seguintes expressões: dados de pesquisa; repositórios de dados de pesquisa; avaliação de repositórios; avaliação de repositórios de dados de pesquisa; repositórios confiáveis; *research data repositories; evaluation of research data repositories* e *trusted digital repositories.*

Durante o desenvolvimento da tese, a partir das publicações encontradas, foi realizada a leitura com vistas a conhecer os documentos coletados. Por conseguinte, foram lidas em profundidade as publicações selecionadas, categorizando os principais temas e conceitos com a finalidade de fundamentar a redação do referencial teórico que orientaram os demais passos da pesquisa e a execução dos seus objetivos.

Assim, a escolha dos artigos que fizeram parte da revisão bibliográfica, que se encontram relatados ao longo desta tese, foi realizada buscando sempre utilizar fontes de responsabilidade dos principais autores da área. Dito isto, o levantamento bibliográfico teve a finalidade de verificar a existência ou não de instrumento para avaliação semelhante ao conjunto de critérios/requisitos proposto por este estudo. A pesquisa bibliográfica identificou 4 documentos importantes para confiabilidade de repositórios: o *CoreTrustSeal*, o ACTDR, os Princípios FAIR e TRUST. Na próxima seção será explicado como eles foram analisados.

3.2 SEGUNDA ETAPA

Para que fosse possível verificar quais critérios/requisitos são importantes para garantir a confiabilidade de um repositório de dados de pesquisa brasileiro, foi necessário identificar na literatura referente à área e nos documentos institucionais, os principais critérios/requisitos utilizados em avaliações de organizações renomadas. A pesquisa documental requer fontes condizentes portanto, o universo para a aplicação dessa técnica de

pesquisa no presente estudo se restringiu aos documentos disponíveis na web, em fontes de informação confiáveis como sites de organizações e instituições internacionais. Desta forma, um conjunto único de critérios/requisitos, foi elaborado de acordo com os critérios/requisitos essenciais para um repositório de dados de pesquisa com base nos instrumentos já abordados neste estudo. A seleção dos 4 documentos, citados a seguir, guiou-se pelas seguintes características: referem-se ao principal assunto da pesquisa (avaliação de repositórios); possuem um reconhecimento internacional e as metodologias a eles aplicadas foram tratadas com aprofundamento podendo ser relacionadas a avaliação de confiabilidade de repositórios de dados de pesquisa.

a) *CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022*[16]*;*
b) *ACTDR - Audit and Certification of Trustworthy Digital Repositories;*
c) *Principles FAIR*[17]*;*
d) *The TRUST Principles for digital repositories*[18]*.*

O modelo de análise deste estudo foi a partir da averiguação dos instrumentos de avaliação internacionais que serviram de base para a definição e que compuseram o conjunto de critérios/requisitos. De acordo com Teixeira (2003), a fase do tratamento do material leva o pesquisador à teorização sobre os dados, produzindo o confronto entre a abordagem teórica anterior e o que a investigação de campo aporta de singular como contribuição. Posteriormente à coleta de dados, a fase seguinte da pesquisa é a de análise e interpretação. Estes dois processos, apesar de conceitualmente distintos, aparecem sempre estreitamente relacionados. A análise tem como objetivo organizar e sumariar os dados de forma que possibilitem o fornecimento de respostas ao problema proposto para investigação.

No entanto, a interpretação, tem como objetivo a procura do sentido mais amplo das respostas, o que é feito mediante sua ligação a outros conhecimentos anteriormente obtidos (GIL, 1999). Segundo Minayo (1994), a fase de análise de dados na pesquisa social reúne três finalidades: estabelecer uma compreensão dos dados coletados, confirmar ou não os

[16]https://www.coretrustseal.org/why-certification/requirements/
[17]https://www.go-fair.org/fair-principles/
[18]https://www.nature.com/articles/s41597-020-0486-7.pdf

pressupostos da pesquisa e/ou responder às questões formuladas e ampliar o conhecimento sobre o assunto pesquisado, articulando-o ao contexto cultural do qual faz parte.

3.3 TERCEIRA ETAPA

Na terceira etapa aconteceu a análise, padronização e sistematização dos critérios/requisitos de cada instrumento selecionado neste estudo. Concomitante a análise dos instrumentos realizou-se a categorização dos critérios. Como resultado desta seleção e análise dos critérios/requisitos inseridos nestes instrumentos desenvolvidos pelas instituições/organizações, foi gerado um checklist/lista única de verificações com todos os critérios/requisitos dos instrumentos analisados, cuja finalidade é ter um padrão único que permita a comparação de analogias e diferenças das quatro fontes estudadas, ou seja, este instrumento de controle teve como objetivo verificar uma lista de requisitos, coletar dados de forma ordenada e conduzir uma observação sistemática otimizando o processo.

Com o preenchimento do checklist foi possível elaborar anotações analíticas para refletir sobre os dados coletados, identificar pontos a serem checados, identificar relações e elaborar comparações entre os critérios/requisitos. Para cada instrumento selecionado foi elaborada uma lista e procurou-se descrever detalhes sobre os critérios/requisitos e as áreas abrangentes. Após, sentiu-se a necessidade de agrupar os requisitos semelhantes para possibilitar análises adicionais e procurou-se observar as áreas similares e seus respectivos critérios/requisitos. Após serem contrastadas as listas, os critérios/requisitos equivalentes e aqueles constatados importantes compuseram o conjunto de critérios/requisitos.

O resultado das análises comparativas entre os 4 documentos está apresentado na seção 4.2. A síntese em uma lista única de critérios está na seção 4.3, apresentada a seguir.

4 ANÁLISE E INTERPRETAÇÃO DOS DADOS

Nesta seção são relatadas as interpretações dos dados obtidos a partir das análises dos instrumentos selecionados.

4.1 ANÁLISE DOCUMENTAL

Nesta seção, foram apresentados os documentos na íntegra mas, em português, tendo em vista que os originais foram publicados em inglês. O foco desta pesquisa foi fundamentado sob a análise dos documentos contemplados no quadro 2.

Quadro 2 – Informações sobre os documentos selecionados

Documento/Título	Autoria/Responsabilidade	Data de publicação
CoreTrustSeal Trustworthy Data Repositories Requirements: extended guidance 2020–2022	*CoreTrustSeal Standards and Certification Board* Líder do projeto: Jonas Recker	2019
ACTDR - Audit and Certification of Trustworthy Digital Repositories	*CCSDS – The Consultative* Committee *for Space Data Systems*	2011
Principles FAIR	Wilkinson, Mark D., *et al.*	2016
The TRUST Principles for digital repositories	Dawei Lin *et a*l.	2020

Fonte: elaborado pela autora.

A seguir são apresentadas as análises referentes a cada um dos documentos:

4.1.1 Coretrustseal Trustworthy Data Repositories Requirements: extended guidance 2020–2022

O documento em análise é dividido em seções contemplando uma parte introdutória, algumas orientações gerais e, posteriormente, os requisitos com suas respectivas subdivisões.

Quadro 3 – Critérios/ Requisitos *CoreTrustSeal*

Infraestrutura organizacional **1. Missão/escopo** **2. Licenças** **3. Continuidade de acesso** **4. Confidencialidade/Ética** **5. Infraestrutura organizacional** **6. Orientação de especialista**
Gerenciamento de objetos digitais **7. Integridade e autenticidade dos dados** **8. Avaliação** **9. Procedimentos de armazenamento documentados** **10. Plano de preservação** **11. Qualidade dos dados** **12. Fluxos de trabalho** **13. Descoberta e identificação de dados** **14. Reutilização de dados**
Tecnologia **15 Infraestrutura técnica** **16 Segurança**

Fonte: elaborado pela autora.

Cada um dos requisitos descrito no instrumento é acompanhado por um texto de orientação que descreve as informações que os candidatos a certificação devem fornecer para permitir uma revisão objetiva. O instrumento solicita o nível de conformidade para cada um dos requisitos:

0 - Não se aplica;
1 - O repositório ainda não considerou;
2 - O repositório possui um conceito teórico;
3 - O repositório está em fase de implementação;
4 - O requisito foi totalmente implementado no repositório.

Os níveis de conformidade são indicadores informados pelo candidato, no entanto, os revisores julgam a conformidade com base nas declarações. Ou seja, no momento de indicar o nível de conformidade é necessário justificar em detalhes as características no local onde solicita-se a resposta. Abaixo foram descritos os requisitos contemplados no instrumento com suas definições e comentários.

REQUISITOS:

Contexto

R0. Fornecer o contexto do repositório:

a) tipo de repositório (este item ajudará os revisores a entender qual a função o repositório. Esta explicação pode fazer referência a coleções de dados relevantes, tipos de dados, formatos e disciplinas com as quais o repositório trabalha): repositório institucional, repositório nacional (governamental), repositório de publicação, biblioteca, museu, arquivo, repositório de projetos de pesquisa, entre outros;
b) breve descrição do repositório (este item deve fornecer uma visão geral do repositório);
c) breve descrição da comunidade designada (a definição clara da comunidade demonstra que o candidato a certificação compreende o escopo e as metodologias – formatos preferidos da comunidade de usuários que possuem como alvo. Para servir bem a comunidade designada, o repositório deve ter um conhecimento profundo da composição, habilidades e necessidades);
d) nível de curadoria (este item tem como objetivo verificar se o repositório distribui seu conteúdo para consumidores de dados sem nenhuma alteração mantendo a integridade dos dados. Além de indicar níveis de curadoria, um repositório deve demonstrar que garante acessibilidade a longo prazo): conteúdo distribuído conforme depositado, curadoria básica/adição de metadados básicos, curadoria aprimorada/conversão para novos formatos e aprimoramento da documentação, curadoria de nível de dados/de forma aprimorada mas com adição adicional dos dados.

Orientação: deve-se listar os parceiros, informar o resumo de mudanças significativas e outras informações relevantes como, por exemplo, referir seu registro no re3data, número de funcionários, tamanho da coleção, número médio de downloads, modelo de negócio ou financiamento entre outras informações.

Infraestrutura organizacional

1. Missão/escopo

R1. O repositório tem a missão explícita de fornecer acesso e preservar os dados em seu domínio.

Orientação: o repositório é responsável pela administração dos objetos digitais e por garantir que os materiais sejam mantidos no ambiente apropriado por períodos de tempo. A preservação e o acesso contínuo aos dados é uma função explícita do repositório. Para este requisito é necessário descrever a missão da organização em preservar e fornecer acesso aos dados.

2. Licenças

R2. O repositório mantém todas as licenças aplicáveis cobrindo o acesso e uso de dados e monitora a conformidade.

Orientação: este requisito refere-se aos regulamentos de acesso e licenças aplicáveis estabelecidas pelo próprio repositório de dados, bem como quaisquer códigos de conduta que são geralmente aceitos no setor relevante para o intercâmbio e uso adequado de conhecimento e informação. Para este requisito deve-se descrever os contratos de licenças, condições de uso (direitos de propriedade intelectual, uso pretendido, proteção de dados confidenciais, etc). Deve-se considerar as consequências se o não cumprimento for detectado (no caso de divulgação de dados pessoais confidências, pode haver penalidades legais severas que afetam tanto o usuário quanto o repositório, ou seja, os repositórios devem ter uma política pública em vigor para o descumprimento.

3. Continuidade de acesso

R3. O repositório tem um plano de continuidade para garantir o acesso contínuo e a preservação de seus acervos.

Orientação: este requisito cobre a governança relacionada à operação contínua do repositório ao longo do tempo e durante desastres, bem como evidências em relação ao planejamento de sucessão, designadamente, as medidas em vigor para garantir o acesso e a disponibilidade dos acervos de dados, atualmente e no futuro. Para este requisito deve-se descrever: o nível de responsabilidade assumido por acervos de dados, incluindo quaisquer períodos de preservação garantidos; planos de médio prazo (três a cinco anos) e longo prazo (cinco anos) em vigor para garantir a disponibilidade e acessibilidade contínua dos dados. Ou seja, tanto a resposta a mudanças rápidas de circunstância quanto o planejamento de longo prazo devem ser descritos. Neste requisito é necessário descrever o nível de responsabilidade assumido pelos dados e o nível de risco para a organização atual, ou seja, é importante descrever se o depositante compartilha a responsabilidade pelo futuro dos dados, se o repositório fornece acesso, preservação e/ou armazenamento de dados a algum nível mínimo de qualidade por um período mínimo de tempo. Essas informações possibilitam o julgamento em relação a sustentabilidade do repositório em termos de finanças e processos, que pode não estar sob responsabilidade do repositório em si, e sim por um host superior ou uma organização principal. Sendo assim, se faz necessário um acordo formal para garantir quem assumirá a responsabilidade em caso de descontinuidade do serviço.

4. Confidencialidade/Ética

R4. O repositório garante, na medida do possível, que os dados são criados, com curadoria, acessados e usados em conformidade com as normas disciplinares e éticas.

Orientação: este requisito se refere às disposições éticas e de privacidade que afetam a criação, curadoria e uso dos dados. O repositório deve demonstrar que tem boas práticas para dados que possuem riscos de divulgação e deve incluir orientações para os depositantes e usuários. Esse procedimento é necessário para manter a confiança daqueles que concordam em ter dados pessoais/confidenciais armazenados no repositório. Para este requisito deve-se

responder como o repositório cumpre as normas disciplinares aplicáveis, como o repositório solicita confirmação de que a coleta ou criação de dados foi realizada de acordo com os critérios legais e éticos, como os dados com risco de divulgação são gerenciados de maneira adequada para limitar acesso, se os funcionários são treinados em relação ao gerenciamento de dados com risco de divulgação, se existem medidas em vigor caso as condições não forem cumpridas. Ou seja, espera-se que as organizações responsáveis pelos dados possuam o dever ético de gerenciá-los no nível esperado pela prática científica da comunidade designada. Para repositórios que que contêm dados sobre indivíduos, organizações ou áreas e espécies protegidas, existem expectativas legais e éticas adicionais de que os direitos dos titulares dos dados serão protegidos. Ou seja, a divulgação desses dados pode representar um risco de dano pessoal, uma violação da confidencialidade comercial ou a divulgação de informações críticas. Se houver algum risco de que esses dados sejam depositados, por exemplo, por acidente, o repositório deve tomar as medidas adequadas para lidar com esses dados e que eles sejam tratados de acordo com os regulamentos legais. O repositório deve apresentar evidências de que possui os procedimentos documentados em vigor para garantir a conformidade.

5. Infraestrutura organizacional

R5. O repositório tem financiamento adequado e número suficiente de funcionários qualificados gerenciados por meio de um sistema claro de governança para realizar a missão com eficácia.

Orientação: os repositórios precisam de financiamento para cumprir suas responsabilidades, junto com uma equipe competente com experiência em arquivamento de dados. No entanto, entende-se que a continuidade do financiamento raramente é garantida. Para este requisito deve-se responder se o repositório é hospedado por uma instituição reconhecida para a garantia da estabilidade e sustentabilidade de longo prazo; se o repositório tem financiamento suficiente incluindo recursos de equipe, TI; se o repositório garante que sua equipe tenha acesso a treinamento contínuo e desenvolvimento profissional. Pode-se fornecer descrições completas das tarefas executadas pelo repositório e as habilidades necessárias para executá-las, no entanto tais descrições não são obrigatórias. A resposta para este requisito deve conter evidências que descrevam os processos de tomada de decisão/gestão da organização e as

entidades envolvidas. A equipe deve ter treinamento apropriado em gerenciamento de dados para garantir padrões de qualidade consistentes.

6. Orientação de especialista

R6. O repositório adota mecanismo(s) para garantir orientação e feedback contínuos de especialistas.

Orientação: um repositório eficaz se esforça para realizar evoluções e adotar novas tecnologias mais eficazes a fim de permanecer valioso para sua comunidade designada. Devido ao ritmo rápido de mudanças é aconselhável que o repositório possua um aconselhamento e feedback de especialistas para garantir sua relevância e melhorias contínuas. Para este requisito deve-se responder se o repositório possui consultores internos ou um comitê consultivo externo que podem incluir especialistas técnicos, de curadoria e ciência dos dados; deve-se responder como o repositório se comunica com os especialistas para aconselhamento; e como o repositório se comunica com a comunidade designada para feedback. O repositório deve dar evidências de que está ligado a uma rede mais ampla de experiência, a fim de demonstrar acesso a aconselhamento e orientação para suas atividades e monitoramento de novos desafios potenciais.

Gerenciamento de objetos digitais

7. Integridade e autenticidade dos dados

R7. O repositório garante a integridade e autenticidade dos dados.

Orientação: o repositório deve fornecer evidências para mostrar que opera um sistema de gerenciamento de dados e metadados adequado para garantir integridade e autenticidade durante processos de armazenamento e acesso a dados. Este requisito cobre todo o ciclo de vida dos dados dentro do repositório. Para proteger a integridade dos dados e metadados, quaisquer alterações intencionais devem ser documentadas, incluindo a justificativa e o originador da mudança. Convém que medidas sejam implementadas para garantir que mudanças não intencionais ou não autorizadas possam ser detectadas e versões corretas de

dados e metadados recuperadas. A autenticidade cobre o grau de confiabilidade dos dados originais depositados e sua procedência, incluindo a relação entre os dados originais e os divulgados e se as relações existentes entre os conjuntos de dados e/ou metadados são mantidas ou não. Para este requisito deve-se incluir evidências sobre a integridade dos dados, ou seja, descrever a documentação da integridade dos dados e metadados; descrever em detalhes como todas as alterações nos dados e metadados são registradas; a descrição estratégica de controle de versão; deve-se utilizar o uso de padrões em convenções internacionais apropriados. Para incluir evidências sobre gerenciamento de autenticidade deve-se responder se o repositório tem uma estratégia para alterações de dados; se mantém links para metadados e outros conjuntos de dados; e se é verificado as identidades dos depositantes.

8. Avaliação

R8. O repositório aceita dados e metadados com base em critérios definidos para garantir relevância e compreensibilidade para os usuários de dados.

Orientação: a função da avaliação é fundamental para avaliar se os dados atendem a todos os critérios de seleção e para garantir o manejo adequado para a sua preservação. A avaliação e a reavaliação ao longo do tempo garantem que os dados permaneçam relevantes e compreensíveis para a comunidade designada. Para este requisito deve-se incluir evidências como, por exemplo, se o repositório possui uma política de desenvolvimento de coleção para orientar a seleção de dados; evidências sobre qual abordagem é usada para dados que não se enquadram no perfil de missão/coleta; se o repositório possui procedimentos em vigor para determinar se são fornecidos os metadados necessários para interpretar e usar os dados; se existe alguma avaliação automatizada da aderência dos metadados aos esquemas relevantes; qual é a abordagem do repositório e se os metadados fornecidos são insuficientes para a preservação a longo prazo; se o repositório publica uma lista de formatos preferidos; se existem verificações para garantir que os produtores de dados sigam formatos preferidos; e qual é o processo de remoção de itens de sua coleção tendo em mente o impacto nos identificadores persistentes. Para que a coleção permaneça relevante e utilizável pela comunidade designada, especialmente à luz das mudanças na tecnologia, cultura e legislação (como por exemplo, proteção de dados ou direitos de propriedade intelectual), os critérios de

seleção podem ter que ser revisados ao longo do tempo e os ativos digitais reavaliados adequadamente.

9. Procedimentos de armazenamento documentados

R9. O repositório aplica processos e procedimentos documentados durante o gerenciamento do armazenamento de arquivos de dados.

Orientação: os repositórios precisam armazenar dados e metadados desde o ponto de depósito até o ponto de acesso. Repositórios que realizam preservação digital devem oferecer armazenamento de arquivo de acordo com o OAIS. Para este requisito, o repositório deve oferecer evidências relacionadas as seguintes questões: como os processos e procedimentos são documentados e gerenciados; se o repositório tem uma estratégia para várias cópias; quais verificações existem para garantir a consistência entre as cópias de arquivo; e como a deterioração da mídia de armazenamento é tratada e monitorada. Os procedimentos são documentados e padronizados de forma que diferentes gerenciadores de dados, embora realizem as mesmas tarefas separadamente, cheguem ao mesmo resultado.

10. Plano de preservação

R10. O repositório assume a responsabilidade pela preservação a longo prazo e gerencia essa função de forma planejada e documentada.

Orientação: o repositório, os depositantes de dados e a comunidade designada precisam entender o nível de responsabilidade assumido para cada item depositado no repositório. O repositório deve ter direitos para assumir essas responsabilidades, ou seja, os procedimentos devem ser documentados e sua conclusão assegurada. Para este requisito, as respostas devem fornecer informações referentes as seguintes questões: se o repositório possui uma abordagem documentada para preservação; se o nível de responsabilidade pela preservação de cada item é compreendido e como isso é definido; se existem planos relacionados a migrações futuras ou medidas semelhantes para lidar com a ameaça de obsolescência; se o contrato entre o depositante e o repositório prevê todas as ações necessárias para cumprir as responsabilidades; se o repositório tem direitos para copiar, transformar e armazenar os itens,

bem como fornecer acesso a eles; as ações relevantes para a preservação estão especificadas na documentação, incluindo transferência de custódia, padrões de informações de envio e padrões de informações de arquivo. Desta forma, o repositório deve informar se possui uma documentação clara para garantir uma abordagem organizada para preservação de longo prazo, acesso contínuo para tipos de dados apesar das mudanças de formatos e se há documentação suficiente para apoiar a usabilidade pela comunidade designada. Deve-se abordar se o repositório tem níveis de preservação definidos e, em caso afirmativo, como eles são aplicados. O plano de preservação deve ser gerenciado para garantir que as mudanças na tecnologia de dados e nos requisitos do usuário sejam tratadas de maneira estável e oportuna.

11. Qualidade dos dados

R11. O repositório tem experiência apropriada para lidar com dados técnicos e qualidade de metadados e garante que informações suficientes estejam disponíveis para os usuários finais fazerem avaliações relacionadas à qualidade.

Orientação: os repositórios devem garantir que haja informações suficientes sobre os dados para que a comunidade designada avalie a qualidade dos dados. A avaliação da qualidade torna-se cada vez mais relevante quando a comunidade designada é multidisciplinar, onde os usuários podem não ter experiência pessoal para fazer uma avaliação da qualidade apenas a partir dos dados. Os repositórios devem ser capazes de avaliar a integridade e qualidade dos dados e metadados. Para este requisito é necessário descrever a abordagem da qualidade dos dados e metadados feita pelo repositório; deve-se descrever se o repositório possui verificações de controle de qualidade para garantir a integridade e a compreensibilidade dos dados depositados (em caso afirmativo, deve-se fornecer as referências aos padrões de controle de qualidade e mecanismos de relatório aceitos pela comunidade de prática relevante e incluir detalhes de como quaisquer problemas são resolvidos). Também é importante descrever a capacidade da comunidade de comentar e/ou classificar dados e metadados e se as citações de trabalhos relacionados ou links para índices de citações são fornecidos.

12. Fluxos de trabalho

R12. O arquivamento ocorre de acordo com fluxos de trabalho definidos, desde o armazenamento até a disseminação.

Orientação: para garantir a consistência das práticas entre conjunto de dados e serviços, os fluxos de trabalho devem ser definidos de acordo com as atividades do repositório e claramente documentados. O modelo de referência OAIS pode ajudar a especificar as funções de fluxo de trabalho de um repositório. Para este requisito deve-se descrever: os fluxos de trabalho (descrições de processos de negócios); a comunicação clara para depositantes e usuários sobre o manuseio de dados; os níveis de segurança e impacto nos fluxos de trabalho (proteção da privacidade dos sujeitos); a verificação qualitativa dos resultados; os tipos de dados gerenciados e qualquer impacto no fluxo de trabalho; e o gerenciamento de mudanças de fluxos de trabalho. Este requisito confirma que todos os fluxos de trabalho estão documentados, ou seja, o repositório deve adotar uma abordagem consistente, rigorosa e documentada para gerenciar todas as atividades em seus processos e que as alterações nesses processos são implementadas, avaliadas, registradas e administradas de maneira adequada. O requisito não exige descrições detalhadas dos fluxos de trabalho, mas busca evidências de como e onde esses fluxos de trabalho são documentados.

13. Descoberta e identificação de dados

R13. O repositório permite que os usuários descubram os dados e os consultem de forma persistente por meio de citações adequadas.

Orientação: a descoberta de dados eficaz é a chave para o compartilhamento de dados. Uma vez descobertos, os conjuntos de dados devem ser referenciados por meio de citações completas, incluindo identificadores persistentes para ajudar a garantir que os dados possam ser acessados no futuro. Para este requisito, o repositório deve incluir evidências referentes aos seguintes aspectos: se o repositório oferece recursos de pesquisa; se o repositório mantém um catálogo de metadados pesquisável para padrões apropriados; quais sistemas de identificadores persistentes o repositório usa; se o repositório facilita a coleta automática dos metadados; e se o repositório oferece recomendação de citação de dados. Ou seja, o

repositório deve dar evidências de que toda a curadoria de dados e metadados apoia a descoberta de objetos digitais claramente definidos e identificados e permite sua vinculação com objetos digitais relacionados de acordo com os padrões de domínio. Deve ficar claro para a comunidade como os dados são citados, de modo que o crédito e a atribuição apropriados sejam dados aos indivíduos/ organizações que contribuíram para sua criação.

14. Reutilização de dados

R14. O repositório permite a reutilização dos dados ao longo do tempo, garantindo que os metadados apropriados estejam disponíveis para apoiar a compreensão e o uso dos dados.

Orientação: os repositórios devem garantir que os dados continuem a ser compreendidos e usados de forma eficaz no futuro, apesar das mudanças na tecnologia e na base de conhecimento da comunidade designada. Este requisito avalia as medidas tomadas para garantir que os dados sejam reutilizáveis. Para este requisito deve-se dar evidências em relação: a quais metadados são fornecidos pelo repositório quando os dados são acessados; como o repositório garante a compreensão contínua dos dados; se os dados são fornecidos em formatos usados pela comunidade designada e quais são esses formatos; e se são tomadas medidas para levar em consideração a possível evolução dos formatos. Para atender a este requisito o repositório deve demonstrar um conhecimento profundo dos cenários de reutilização e das necessidades da comunidade em termos de suas práticas, ambiente técnico e adesão aos padrões aplicáveis. Mudanças na tecnologia e nas metodologias e normas empregadas podem levar a necessidade de reconsiderar o formato em que os dados são divulgados. Da mesma forma, metadados apropriados de alta qualidade em conformidade com um esquema generalizado e/ou disciplinar específico desempenham um papel essencial e devem ser mencionados nas evidências fornecidas.

Tecnologia

15. Infraestrutura técnica

R15. O repositório funciona em sistemas operacionais bem suportados e em outro software de infraestrutura central, e está usando tecnologias de hardware e software apropriadas para os serviços que fornece à sua comunidade designada.

Orientação: os repositórios precisam operar em infraestruturas centrais confiáveis e estáveis que maximizam a disponibilidade do serviço. Ademais, o hardware e o software usados devem ser relevantes e apropriados para a comunidade e para as funções que o repositório desempenha. Para este requisito, deve-se responder as seguintes questões: quais padrões o repositório usa para referência (se são padrões internacionais ou comunitários) e com que frequência são revisados; de que forma esses padrões são implementados; se o repositório possui um plano de desenvolvimento de infraestrutura; se existe um inventário de software e se a documentação do sistema está disponível; se a disponibilidade e conectividade são suficientes para atender as necessidades da comunidade; e se existem procedimentos e arranjos para fornecer recuperação rápida ou backup de serviços essenciais no caso de uma interrupção. Assim sendo, os fluxos de trabalho e os atores humanos que fornecem serviços de repositório devem ser suportados por uma infraestrutura tecnológica adequada que atenda às necessidades da comunidade e permita que o repositório se recupere de desastres de curto prazo. O repositório deve demonstrar que compreende o ecossistema mais amplo de padrões, ferramentas e tecnologias disponíveis para gerenciamento e curadoria de dados de pesquisa e selecionou opções que se alinham com requisitos locais. Se possível deve-se comprovar utilizando-se um modelo de referência como, por exemplo, *Spatial Data Infrastructure* (SDI), *Open Geospatial Consortium* (OGC), W3C ou padrões ISO.

16. Segurança

R16. A infraestrutura técnica do repositório oferece proteção para as instalações e seus dados, produtos, serviços e usuários.

Orientação: o repositório deve analisar ameaças potenciais, avaliar riscos e criar um sistema de segurança consistente. Deve descrever cenários de danos com base em ações maliciosas, erro humano ou falha técnica que representam uma ameaça ao repositório e seus dados, produtos, serviços e usuários. Deve-se medir a probabilidade e o impacto de tais cenários, decidir quais níveis de risco são aceitáveis e determinar quais medidas devem ser tomadas

para combater as ameaças ao repositório e sua comunidade designada. Este processo deve ser contínuo. Para este requisito deve ser descrito: o sistema de segurança de TI, funcionários com funções relacionadas à segurança e quaisquer ferramentas de análise de risco utilizados; os níveis de segurança que são exigidos e como eles são suportados; deve-se descrever quaisquer procedimentos de autenticação e autorização empregados para gerenciar com segurança o acesso aos sistemas em uso. O repositório deve demonstrar que compreende todos os riscos técnicos aplicáveis ao serviço prestado bem como ao ambiente físico. Além disso, deve-se demonstrar que possui mecanismos para prevenir, detectar e responder a um incidente de segurança, ou seja, deve demonstrar de que forma a segurança da infraestrutura técnica é controlada pelo repositório e por sua instituição hospedeira/terceirizada e quem está no comando. Também deve demonstrar se os procedimentos de autenticação e autorização são suficientes para garantir a segurança dos acervos de dados em cada estágio do fluxo de trabalho e mostrar quais políticas de segurança da empresa estão em vigor para controlar segurança de todos os sistemas, incluindo segurança de rede, verificações de intrusos, segurança de instalações físicas e política de senhas.

Assim sendo, os requisitos do *CoreTrustSeal* aqui apresentados podem auxiliar sobre muitos aspectos referentes a melhoria e qualidade dos repositórios de dados de pesquisa. A próxima seção apresenta os requisitos ACTDR.

4.1.2 ACTDR - Audit and Certification of Trustworthy Digital Repositories

O principal propósito deste documento é definir a Prática Recomendada CCSDS com base no processo de auditoria e certificação para avaliar a confiabilidade de repositórios digitais. Destina-se principalmente aos responsáveis pela auditoria de repositórios digitais e também para aqueles que trabalham ou são responsáveis por repositórios digitais e possuem o objetivo de medir da confiabilidade de seu repositório. Algumas instituições também podem optar por usar essas métricas durante um processo de design ou redesenho do seu repositório.

Este documento está dividido em seções informativas, normativas e anexos. As seções são informativas e fornecem uma visão da justificativa, algumas questões importantes de design e uma introdução a terminologia e conceitos. Essas seções fornecem métricas agrupadas da seguinte forma: cobre a infraestrutura organizacional; cobre o gerenciamento de objetos digitais; e cobre Infraestrutura e Gerenciamento de Riscos de Segurança. Cada seção

agrupa métricas em uma ou mais subseções. No quadro 4 são apresentadas as divisões do instrumento analisado:

Quadro 4 – Critérios/ Requisitos *ACTDR*

Infraestrutura organizacional a) governança e viabilidade organizacional b) estrutura organizacional e pessoal c) responsabilidade processual e política de preservação d) sustentabilidade financeira e) contratos, licenças e responsabilidades
Gestão de objetos digitais a) *ingest*: aquisição de conteúdo b) *ingest*: criação do *Archival Information Package* (AIP) c) planejamento de preservação d) preservação de AIP e) gestão de informações f) gestão de acesso
Gestão de risco de infraestrutura e segurança a) gestão de risco de infraestrutura técnica b) gestão de risco de segurança

Fonte: elaborado pela autora.

A seguir, cada uma das áreas são detalhadas.

INFRAESTRUTURA ORGANIZACIONAL

a) governança e viabilidade organizacional

- o repositório deve ter uma declaração de missão que reflita um compromisso com a preservação, retenção de longo prazo e gerenciamento de acesso: isso é necessário para garantir o compromisso com a preservação, retenção, gestão e acesso no nível administrativo

mais alto do repositório. Exemplos e maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito é publicar a declaração de missão ou estatuto do repositório ou de sua organização onde deve abordar explicitamente a preservação. Se a preservação não estiver entre os objetivos principais de uma organização que abriga um repositório digital, então a preservação pode não ser essencial para a organização.

- **o repositório deve ter um Plano Estratégico de Preservação que define a abordagem que o repositório terá no suporte de longo prazo:** esse critério é importante para ajudar o repositório a tomar decisões administrativas, formar políticas e alocar recursos para preservar com sucesso seus acervos. O repositório pode demonstrar que está atendendo a esse requisito através de um Plano Estratégico de Preservação; atas de reuniões; e documentação de decisões administrativas. O plano estratégico deve ser baseado na missão estabelecida da organização, e em seu valores, visão e objetivos definidos. Os planos estratégicos geralmente cobrem um determinado período, normalmente na faixa de 3-5 anos.

> - o repositório deve ter um plano de sucessão adequado, planos de contingência, e/ou acordos de custódia em vigor no caso de o repositório deixar de operar ou a instituição governamental ou de financiamento alterar substancialmente seu escopo. Isto é necessário para preservar o conteúdo da informação confiada ao repositório por transferí-lo para outro custodiante no caso de o repositório deixar de operar. Exemplos e maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito é a Sucessão escrita e plano (s) de contingência; declaração explícita e específica com a intenção de garantir a continuidade do repositório; software e metadados suficientes para permitir a reconstituição do repositório e seu conteúdo em caso de falha do repositório. A falha de um repositório ameaça a sustentabilidade a longo prazo das informações. Não é suficiente para o repositório ter um plano informal ou política a respeito para onde vão seus dados caso ocorra uma falha. Um plano formal com procedimentos identificados é necessário.
>
> - O repositório deve monitorar seu ambiente organizacional para determinar quando executar seu plano de sucessão, planos de contingência e/ou acordos de custódia para garantir que o repositório possa reconhecer quando é necessário executar esses planos. Pode-se demonstrar que está atendendo a esse requisito através de políticas administrativas, procedimentos, protocolos, requisitos; orçamentos e análise

financeira; calendários fiscais; Planos de negócio; qualquer evidência de monitoramento ativo e preparação. A gestão de um repositório deve ter procedimentos formais para periodicamente verificar a viabilidade do repositório. Esta verificação periódica deve ser usada para determinar se, ou quando, para executar o plano de sucessão formal do repositório, planos de contingência e/ou arranjos de custódia.

- **o repositório deve ter uma Política de Coleta ou outro documento que especifique o tipo de informação que irá preservar, reter, gerenciar e fornecer acesso:** este requisito é importante para que o repositório tenha orientações sobre aquisição de conteúdo digital e como irá preservar, reter, gerenciar e fornecer acesso. O repositório pode demonstrar que está atendendo a esse requisito por meio da política de cobrança e documentos comprobatórios; política de preservação, missão, objetivos e visão do repositório. A política de coleta pode ser usada para entender o que o repositório mantém, o que ele não mantém, e por quê. A política de coleta apoia a missão mais ampla do repositório e sem tal política, o repositório provavelmente coletará de maneira aleatória ou armazenará grandes quantidades de conteúdo digital de baixo valor. A política de coleta ajuda a organização a identificar que conteúdo digital irá aceitar ou não, para a gestão. Em uma organização com uma ampla missão, a política de coleção ajuda a definir o papel do repositório dentro do contexto organizacional mais amplo.

b) estrutura organizacional e pessoal

- **o repositório deve identificar e estabelecer as funções de que necessita e deve nomear pessoal com habilidades e experiência adequadas para cumprir essas funções:** a equipe do repositório deve ser composta por pessoal com o treinamento e as habilidades necessárias para transportar as atividades do repositório. O repositório deve ser capaz de documentar por meio de planos de desenvolvimento, organogramas, descrições de cargos e políticas relacionadas e procedimentos que o repositório está definindo e mantendo as habilidades e funções que são necessárias para a operação sustentada pelo repositório.

    - O repositório deve ter identificado e estabelecido as funções de que necessita executar: isso é necessário para garantir que o repositório possa completar todas as tarefas associadas com a preservação e gerenciamento de longo prazo dos objetos de dados. Exemplos e maneiras pelas quais o repositório pode demonstrar que está

atendendo a esse requisito será através de um plano de pessoal; definições de competência; descrições de cargos; desenvolvimento profissional da equipe, planos; certificados de treinamento e credenciamento; além de evidências de que o repositório analisa e mantém esses documentos conforme os requisitos evoluem. A preservação depende de uma série de atividades, desde a manutenção de hardware e software até migração de conteúdo e mídia de armazenamento para negociação de acordos de direitos de propriedade intelectual. A fim de garantir a sustentabilidade a longo prazo, um repositório deve estar ciente de todas as atividades necessárias e demonstrar que pode concluí-las com sucesso.

- O repositório deve ter o número adequado de funcionários para dar suporte a todas funções e serviços: isso é necessário para garantir que os níveis de pessoal do repositório sejam adequados para preservar o conteúdo digital e fornecer um repositório seguro e de qualidade. Exemplos e maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito será por meio de organogramas; definições de funções e responsabilidades; comparação dos níveis de pessoal para *benchmarks* e padrões da indústria. O repositório deve determinar o número apropriado e o nível de pessoal que corresponde aos requisitos e compromissos. O repositório também deve demonstrar como avalia a eficácia e adequação do pessoal para apoiar as suas funções e serviços.

- o repositório deve ter implementado um programa ativo de desenvolvimento profissional que fornece aos funcionários oportunidades de desenvolvimento de habilidades e conhecimentos para garantir que os conjuntos de habilidades da equipe evoluam conforme a tecnologia do repositório e alteração dos procedimentos de preservação. Formas pelas quais o repositório pode demonstrar que está atendendo a esse requisito é por meio de planos e relatórios de desenvolvimento profissional; requisitos de treinamento e orçamentos de treinamento, documentação de gastos com treinamento (valor por equipe); metas de desempenho e documentação das atribuições e realizações do pessoal, e cópias dos certificados atribuídos. A tecnologia e as práticas gerais para preservação digital continuarão a mudar, assim como os requisitos de sua comunidade designada, então o repositório deve garantir que a habilidade de sua equipe evoluam.

c) responsabilidade processual e política de preservação: A documentação garante às partes interessadas (consumidores, produtores e colaboradores do conteúdo digital) que o repositório está atendendo aos seus requisitos e desempenhando plenamente sua função como um repositório digital confiável. Um repositório deve criar documentação que reflita sua declaração de missão. Isso envolve documentar todos os processos do repositório, tomada de decisão e definição de metas. Ele garante que as políticas e procedimentos de repositório sejam realizados em aprovação, de forma consistente, resultando na preservação em longo prazo e no acesso ao conteúdo digital.

- o repositório deve ter definido sua comunidade designada e base associada de conhecimento e devem ter essas definições apropriadamente acessíveis: isso é necessário para que seja possível testar se o repositório atende às necessidades de sua comunidade designada. O repositório pode demonstrar que está atendendo a esse requisito através de uma definição escrita da comunidade designada. A comunidade designada é definida como "um grupo identificado de consumidores potenciais que deve ser capaz de compreender um determinado conjunto de informações" e pode ser composta de várias comunidades de usuários.

- o repositório deve ter Políticas de Preservação em vigor para garantir que o seu Plano Estratégico de Preservação seja cumprido: é necessário para garantir que o repositório possa cumprir a parte de sua missão relacionada à preservação. Formas pelas quais o repositório pode demonstrar que está atendendo a esse requisito será através de políticas de preservação e declaração da missão do repositório. As políticas de repositório mostram como o repositório atende aos requisitos do plano estratégico de preservação. Por exemplo, um plano estratégico de preservação pode conter uma exigência de que o repositório "cumpra os padrões de preservação atuais". A política de preservação pode então exigir que o repositório "monitore os padrões de preservação atuais para garantir a conformidade do repositório com os padrões".

- O repositório deve ter mecanismos para revisão, atualização e desenvolvimento de suas políticas de preservação à medida que o repositório cresce e à medida que a tecnologia e a prática da comunidade evolui. O repositório pode demonstrar que está atendendo a esse requisito apresentando documentação escrita atual e anterior na

forma de Políticas de Preservação, Planos Estratégicos de Preservação, Planos de Implementação de Preservação, procedimentos, protocolos e fluxos de trabalho; especificações de ciclos de revisão para documentação; detalhamento de documentação, avaliações, pesquisas e feedback. As Políticas de Preservação capturam os compromissos organizacionais e as intenções de pessoal, segurança e outras questões relacionadas à preservação. O repositório pode achar benéfico manter todas as versões das políticas de preservação (por exemplo, versões desatualizadas são claramente identificadas e mantidas de alguma forma organizada), a fim de documentar os resultados de monitoramento de novos desenvolvimentos, mostrando a capacidade de resposta do repositório aos padrões e práticas, requisitos emergentes e padrões que são específicos para o domínio, se apropriado, e desenvolvimentos semelhantes. A equipe qualificada e colegas são uma parte importante do processo de revisão, pois auxiliam na atualização e ampliação desses documentos. As políticas devem ser compreensíveis pela equipe do repositório para que possam realizar seu trabalho e os procedimentos de preservação devem ser demonstrados como compreensíveis e implementáveis.

- o repositório deve ter um histórico documentado das mudanças em suas operações, procedimentos, software e hardware: é necessário para fornecer uma "trilha de auditoria" através da qual as partes interessadas possam identificar e rastrear as decisões tomadas pelo repositório. Deve-se apresentar estoques de bens de capital; documentação da aquisição, implementação, atualização, e retirada de software e hardware críticos do repositório; retenção e descarte de arquivos, cronogramas e políticas, cópias de versões anteriores de políticas e procedimentos. Esta documentação pode incluir decisões sobre a organização e a documentação ou entrevistas com a equipe apropriada.

- o repositório deve se comprometer com a transparência e responsabilidade em todas as ações apoiando a operação e gestão do repositório que afetam a preservação de conteúdo digital ao longo do tempo: é necessário porque a transparência, no sentido de estar à disposição de quem deseja saber, é a melhor garantia de que o repositório opera de acordo com padrões e práticas. Exemplos pelos quais o repositório pode demonstrar que está atendendo a esse requisito é através de relatórios de auditorias, certificações financeiras e técnicas; divulgação de governança de documentos, revisões de programas independentes,

contratos e acordos com fornecedores de financiamento e serviços essenciais. Se o repositório usa software para capturar informações sobre seu histórico, ele deve ser capaz de demonstrar essas ferramentas de rastreamento. Quando apropriado, o histórico está vinculado a estratégias de preservação e descreve os efeitos potenciais na preservação de conteúdo digital. Este requisito mostra que a organização se compromete a divulgar seus métodos para preservar o conteúdo digital, pelo menos para a comunidade designada ou outra parte interessada, a fim de demonstrar que está atendendo a todas as partes e requisitos atuais de preservação.

- o repositório deve definir, coletar, rastrear e fornecer adequadamente suas medições de integridade da informação: é necessário para fornecer a documentação que desenvolveu ou adaptou medidas adequadas para garantir a integridade da sua exploração. O repositório pode demonstrar que está atendendo a esse requisito a partir da definição ou especificação escrita das medidas de integridade do repositório (por exemplo, soma de verificação calculada ou função *hash*); documentação dos procedimentos e mecanismos para monitorar medições de integridade e responder aos resultados das medições de integridade que indicam que o conteúdo digital está em risco; um processo de auditoria para coletar, rastrear e apresentar medições de integridade; política de preservação e documentação do fluxo de trabalho. Os mecanismos para medir a integridade irão evoluir conforme a tecnologia evolui e o repositório pode fornecer documentação que desenvolveu ou se adaptou medidas adequadas para garantir a integridade de suas participações. Se os protocolos, regras e mecanismos estão incorporados no software de repositório, deve haver alguma maneira de demonstrar a implementação de medidas de integridade.

- o repositório deve se comprometer com um cronograma regular de autoavaliação e certificação externa: é necessário para garantir que o repositório continue sendo confiável e que não haja ameaça ao seu conteúdo. Exemplos e maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito será através de listas de verificação preenchidas e datadas de autoavaliações e/ou auditorias de terceiros; certificados concedidos em conformidade com os padrões ISO; cronogramas e evidências de alocações de orçamento para certificação futura. Uma verificação única de confiabilidade não é adequada porque muitas coisas vão mudar com o tempo. Um compromisso de longo prazo deve ser demonstrado.

d) sustentabilidade financeira

- o repositório deve ter processos de planejamento de negócios de curto e longo prazo para sustentar o repositório ao longo do tempo: isso auxilia na garantia da viabilidade do repositório ao longo do período de tempo que prometeu fornecer acesso a seu conteúdo para sua comunidade designada. Pode-se demonstrar que está atendendo a esse requisito através de planos estratégicos, operacionais e/ou de negócios atualizados e plurianuais; previsões financeiras com vários cenários de orçamento; e planos de contingência. Um processo de planejamento de negócios anual é comumente aceito como o padrão para a maioria organizações.

- o repositório deve ter práticas e procedimentos financeiros que sejam transparentes, compatíveis com os padrões e práticas contábeis relevantes e auditado por terceiros de acordo com os requisitos legais: esse requisito é necessário a fim de proteger contra má conduta ou outra atividade desagradável que possa ameaçar a viabilidade econômica do repositório. O repositório pode demonstrar que está atendendo a esse requisito através dos requisitos de disseminação demonstrados no planejamento e práticas de negócios; normas e práticas de contabilidade e auditoria; e demonstrações financeiras anuais. O repositório não pode simplesmente reivindicar transparência, mas deve mostrar que ajusta seus negócios para mantê-los transparentes, compatíveis e auditáveis. Requisitos de confidencialidade podem proibir a divulgação de informações sobre as finanças do repositório, mas o repositório deve ser capaz de demonstrar que está satisfazendo as necessidades de sua comunidade designada.

- o repositório deve ter um compromisso contínuo de analisar e relatar sobre risco financeiro, benefício, investimento e despesas: isto é necessário para demonstrar que o repositório identificou e documentou essas categorias e as gerencia ativamente, incluindo a identificação e resposta aos riscos, descrevendo e aproveitando os benefícios, especificando e equilibrando os investimentos e antecipando a preparação para despesas. Exemplos e maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito é elaborando documentos de gestão de risco que identificam ameaças percebidas e potenciais (registro de risco); planejamento de investimento em infraestrutura e tecnologia; análises de custo/benefício; documentos de investimentos financeiros; requisitos e exemplos de licenças, contratos e gerenciamento de ativos; e evidência de revisão com base no risco. O repositório

deve ter o objetivo de manter um equilíbrio adequado entre risco e benefícios, investimento e retorno.

e) contratos, licenças e responsabilidades

- o repositório deve ter e manter contratos ou depósitos apropriados para os materiais digitais que gerencia, preserva e/ou para os quais fornece acesso: este requisito é importante para garantir que o repositório tenha os direitos e autorizações necessários para permitir que ele colete e preserve conteúdo digital ao longo do tempo, faça com que essas informações estejam disponíveis para sua comunidade designada, e para defender esses direitos quando questionados. O repositório pode mostrar que está atendendo a esse requisito por meio de contratos de depósito e licenças devidamente assinados e executados de acordo com as normas locais, leis e regulamentos nacionais e internacionais; políticas sobre acordos de depósito de terceiros; definições de níveis de serviço e usos permitidos; políticas de repositório sobre o tratamento de 'Obras órfãs' e resolução de disputas de direitos autorais; relatórios de avaliações de risco; procedimentos para revisar e manter regularmente acordos, contratos e licenças. Os repositórios podem precisar mostrar evidências de que seus contratos estão sendo cumpridos e isto é especialmente importante para aqueles com acordos de depósito de terceiros. Contratos e acordos de depósito formais devem ser legítimos, ou seja, eles precisam ser assinados e atuais. Repositórios envolvidos na colheita da Web podem achar difícil de cumprir este requisito devido à maneira como as informações baseadas na Web são colhidas/capturadas para preservação de longo prazo e, portanto, contratos ou acordos de depósito raramente são necessários. Alguns repositórios capturam, gerenciam e preservam o acesso a este material sem permissão por escrito dos criadores de conteúdo e outros passam pelo processo demorado e caro de entrar em contato com os proprietários do conteúdo antes de capturar e processar informações. Idealmente, acordos são rastreados, vinculados, gerenciados e disponibilizados em um banco de dados de contratos.

> - o repositório deve ter contratos ou acordos de depósito que especifiquem e transfiram todos os direitos de preservação necessários, e esses direitos transferidos serão documentados: este item é necessário para ter controle suficiente das informações para preservação e limitar a exposição do repositório a responsabilidades ou danos jurídicos e financeiros. Formas pelas quais o repositório

pode demonstrar que está atendendo a esse requisito é por meio de contratos, acordos de depósito; especificações de direitos transferidos para diferentes tipos de conteúdo digital (se aplicável); e declarações de política sobre os direitos de preservação necessários. O direito de mudar ou alterar as informações digitais é muitas vezes restrito por lei ao criador, portanto é importante que os contratos e acordos do repositório digital atendam à necessidade de ser capaz de trabalhar e potencialmente modificar objetos digitais para mantê-los acessíveis. Acordos com depositantes devem especificar e/ou transferir para o repositório certos direitos possibilitando ações de preservação adequadas e necessárias para os objetos digitais dentro do repositório. As negociações jurídicas podem levar tempo, retardando a gestão de objetos digitais em risco. É aceitável para um repositório digital receber ou aceitar objetos digitais, mesmo com direitos de preservação mínimos.

- o repositório deve ter especificado todos os aspectos apropriados de aquisição, manutenção, acesso e retirada em acordos escritos com depositantes e outras partes relevantes: isso é necessário para garantir que as respectivas funções de repositório, produtores e contribuintes no depósito de conteúdo digital e transferência de responsabilidade pela preservação sejam compreendidos e aceitos por todas as partes. Será possível demonstrar por meio de contratos de submissão, contratos de depósito e atos de doação executados adequadamente. O contrato de depósito especifica todos os aspectos dessas questões que são necessários para o repositório cumprir a sua função e pode haver um único acordo cobrindo todos os depósitos, ou acordos específicos para cada depósito, ou um acordo padrão complementado por condições para alguns depósitos. Estas condições especiais podem ser adicionadas ao contrato padrão ou substituir alguns aspectos do acordo padrão. Os acordos podem precisar cobrir restrições sobre o acesso e todos os direitos de propriedade sobre os objetos digitais. Acordos podem colocar responsabilidades sobre os depositantes, como garantir que os pacotes de informações de envio estejam em conformidade com alguns padrões preacordados e podem permitir que os repositórios recusem os pacotes de informações que não atendem a esses padrões. Outros repositórios podem assumir a responsabilidade de corrigir erros e a divisão de responsabilidades deve ser sempre clara. Acordos por escrito podem não ser necessários, assim, o ônus da prova recai sobre o repositório para demonstrar que não precisa de tais acordos. Um acordo

deve incluir, no mínimo, direitos de propriedade, direitos de acesso, condições de retirada, nível de segurança, nível de encontrar ajudas, tempo, volume e conteúdo das transferências.

- o repositório deve ter políticas escritas que indiquem quando ele aceita a responsabilidade pela preservação do conteúdo de cada conjunto de objetos de dados enviados. É necessário para evitar mal-entendidos entre o repositório e produtor/depositante sobre quando e como a transferência de responsabilidade pelo conteúdo digital ocorre. Exemplos pelas quais o repositório pode demonstrar que está atendendo a esse requisito será através de contratos de submissão, contratos de depósito e atos de doação executados adequadamente; e recibo de confirmação enviado de volta ao produtor/depositante. Se este requisito não for cumprido, existe o risco de, por exemplo, o original ser apagado antes do repositório assumir a responsabilidade pelos objetos de dados enviados. Sendo entendido que o repositório já assumiu a responsabilidade de preservação do pacote de dados, existe o risco de que o produtor/depositante possa fazer alterações nos dados e estes poderiam não ser preservados adequadamente, pois já foram geridos pelo repositório.

- o repositório deve ter políticas em vigor para lidar com responsabilidades e desafios para propriedade/direitos: isso é necessário para minimizar a responsabilidade potencial e os desafios aos direitos do repositório. O repositório pode demonstrar que está atendendo a esse requisito elaborando uma definição de direitos, licenças e permissões a serem obtidos dos produtores e contribuidores de conteúdo digital; citações de leis e regulamentos relevantes; histórico documentado para responder aos desafios de maneira que não inibe a preservação; e registros de pareceres jurídicos relevantes. As Políticas de Preservação e Planos de Implementação de Preservação do repositório e mecanismos devem ser examinados por autoridades institucionais apropriadas e/ou especialistas jurídicos para garantir que as respostas aos desafios cumpram as leis e requisitos relevantes.

- **o repositório deve rastrear e gerenciar os direitos de propriedade intelectual e restrições ao uso do conteúdo do repositório, conforme exigido pelo acordo de depósito, contrato ou licença:** isso é necessário para permitir que o repositório rastreie, atue e verifique os direitos e restrições relacionadas ao uso dos objetos digitais dentro do repositório. Pode-se

demonstrar que está atendendo a esse requisito através de uma declaração de Política de Preservação que define e especifica os requisitos do repositório e processo de gestão de direitos de propriedade intelectual; acordos de depositantes; amostras de acordos e outros documentos que especificam e tratam dos direitos de propriedade intelectual; documentação de monitoramento ao longo do tempo de mudanças no status de propriedade intelectual em conteúdo digital detido pelo repositório; e resultados do monitoramento e metadados que capturam informações de direitos. O repositório deve ter um mecanismo de rastreamento de licenças e contratos aos quais é obrigado e qualquer que seja o formato do sistema de rastreamento, ele deve ser suficiente para a instituição rastrear, agir e verificar os direitos e restrições relacionados ao uso dos objetos digitais dentro do repositório.

GESTÃO DE OBJETOS DIGITAIS

a) *ingest* (gestão/armazenamento de conteúdo): aquisição de conteúdo

- o repositório deve identificar as informações de conteúdo e as informações de propriedades que preservará: este requisito é necessário para deixar claro para os financiadores, depositantes e usuários quais responsabilidades que o repositório está assumindo e quais aspectos são excluídos. Também é uma etapa importante na definição das informações que são necessárias dos produtores de informações ou depositantes. O repositório pode demonstrar que está atendendo a esse requisito por meio da declaração de missão; acordos de submissão; acordos de depósito; títulos de doação; fluxo de trabalho e documentos da política de preservação, incluindo a definição escrita das propriedades conforme acordado no contrato de depósito ou escritura de doação; procedimentos de processamento escritos; e documentação de propriedades a serem preservadas. Este processo começa em geral com a declaração de missão do repositório e pode ser mais especificado em acordos de pré-adesão com produtores ou depositantes. Pode-se desejar preservar a aparência e o layout exatos dos documentos textuais, enquanto outros podem escolher manter as unidades de medição dos campos de dados e normalizar os dados durante o processo de gestão. Se os identificadores exclusivos forem associados a objetos digitais antes de gerir, eles também podem ser propriedades que precisam ser preservadas.

- o repositório deve ter procedimento(s) para identificar as informações de propriedades que ele preservará para estabelecer um entendimento claro com depositantes, financiadores e comunidades designadas do repositório. Estes procedimentos serão necessários para confirmar a autenticidade ou para identificar reivindicações errôneas de autenticidade do registro digital preservado. O repositório pode demonstrar que está atendendo a esse requisito por meio das definições das propriedades da informação que devem ser preservadas; submissão de acordos de depósito, políticas de preservação, procedimentos de processamento por escrito, e documentação do fluxo de trabalho. Estes procedimentos documentam os métodos e fatores que um repositório usa para determinar os aspectos de diferentes tipos de informações de conteúdo para os quais ele aceita a responsabilidade para com as comunidades designadas.

- o repositório deve ter um registro das informações de conteúdo e de propriedade da informação que irá preservar para identificar por escrito as informações de conteúdo dos registros. Pode ser realizado por meio de políticas de preservação, manuais de processamento, inventários ou pesquisas de coleção, e registros de conteúdo. O repositório deve demonstrar que estabelece e mantém uma compreensão de suas coleções digitais para realizar a preservação necessária e manter as propriedades com os quais se comprometeu. O repositório pode usar essas informações para determinar a eficácia de suas atividades de preservação ao longo do tempo.

- o repositório deve especificar claramente as informações que precisam ser associadas com informações de conteúdo específicas no momento de seu depósito: isso é necessário para que haja um entendimento claro do que precisa ser adquirido do produtor. Maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito será através dos requisitos de transferência; acordos produtor-arquivo; e planos de fluxo de trabalho para produzir o AIP. Para a maioria dos tipos de objetos digitais a serem ingeridos, o repositório deve ter critérios escritos, preparado pelo repositório por conta própria ou em conjunto com outras partes, que especificam exatamente quais objetos digitais são transferidos, qual documentação está associada ao(s) objeto(s) e quaisquer restrições de acesso, sejam técnicas, regulamentares ou impostas por doadores. Esses critérios documentam quais informações o repositório e suas comunidades designadas podem esperar dos objeto(s) digital(is) após o

depósito. O nível de precisão nessas especificações irá variar com a natureza da política de coleta do repositório e sua relação com os criadores.

- o repositório deve ter especificações adequadas que permitam o reconhecimento e análise dos *Submission Information Package* (SIP): este item é necessário para ter certeza de que o repositório é capaz de extrair informações do SIPs, ou seja, o repositório pode demonstrar que está atendendo a esse requisito por meio de pacotes de informações para os SIPs; informações de representação para os dados de conteúdo SIP, incluindo especificações de formato de arquivo documentado; padrões de dados publicados; e documentação de construção de objeto válido. O repositório deve ser capaz de determinar quais são os conteúdos de um SIP em relação a construção técnica de seus componentes.

- o repositório deve ter mecanismos para verificar adequadamente a identidade do produtor de todos os materiais: esse procedimento é necessário para evitar o fornecimento de proveniência errônea das informações que são preservadas. Maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito é por meio de acordos de submissão; acordos de depósito; títulos de doação juridicamente vinculativos; evidências de medidas tecnológicas apropriadas; e logs de procedimentos e autenticações. Os procedimentos operacionais padrão escritos do repositório e as práticas reais devem garantir que objetos digitais são obtidos do depositante esperado.

- o repositório deve ter um processo de *ingest* que verifica cada SIP para integridade e correção: isso é necessário para detectar e corrigir erros no SIP quando criado e potencial erros de transmissão entre o depositante e o repositório. O repositório pode demonstrar que está atendendo a esse requisito através de política de preservação apropriada, documentos do plano de implementação de preservação e arquivos de log do sistema executando procedimento(s) de *ingest*; logs ou registros de arquivos recebidos durante o processo de transferência; documentação de operação padrão, procedimentos detalhados e/ou fluxos de trabalho; registros de formatos; definições de integridade e correção. As informações coletadas durante o processo de *ingest* devem ser comparadas com as informações de alguma outra fonte para verificar a exatidão do processo de transferência de dados. Incluirão metadados técnicos e descritivos obtidos antes do *ingest* e também podem incluir as expectativas definidas pelo depositante, o produtor do objeto, um registro de formato ou as

próprias expectativas do repositório. Até que ponto um repositório pode determinar a exatidão vai depender do que sabe sobre o SIP e quais ferramentas estão disponíveis para verificar a correção.

- o repositório deve obter controle suficiente sobre os objetos digitais para preservá-los: este requisito é necessário para garantir que a preservação possa ser realizada com controle legal e o repositório poderá demonstrar que está atendendo a esse requisito utilizando documentos que mostram o nível de controle físico que o repositório realmente possui. Um catálogo de banco de dados/metadados listando todos os objetos digitais no repositório e metadados suficientes para validar a integridade desses objetos (tamanho do arquivo, soma de verificação, hash, localização, número de cópias, etc.). É necessário controle físico e legal suficiente para que os arquivos façam quaisquer mudanças exigidas por seu Plano de Implementação de Preservação para esses dados e distribuí-los para seus consumidores. Por exemplo, nos casos em que os SIPs fazem referência apenas a objetos digitais, o repositório também deve referenciar os objetos digitais ou preservá-los se o repositório atual não for comprometido com tal preservação.

- o repositório deve fornecer ao produtor/depositante as respostas adequadas em pontos acordados durante os processos de *ingest*: isso é necessário para garantir que o produtor possa verificar se não há falhas na comunicação que, de outra forma, poderiam permitir a perda de SIPs. Exemplos de maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito será através de acordos de envio; acordos de depósito; títulos de doação; documentação de fluxo de trabalho; procedimentos operacionais padrão; e evidências de 'reportar' com relatórios, correspondência, memorandos, ou e-mails. Com base no plano de processamento inicial e no acordo entre o repositório e o produtor/depositante, o repositório deve fornecer relatórios de progresso em pontos acordados ao longo do processo de *ingest*. Os produtores/depositantes podem solicitar mais informações *ad hoc* quando os relatórios previamente acordados forem insuficientes.

- o repositório deve ter registros contemporâneos de ações e processos de administração que são relevantes para a aquisição de conteúdo: isso é necessário para garantir que essa documentação, que pode ser necessária em uma auditoria, seja precisa e autêntica. Exemplos de maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito:

através da documentação escrita das decisões e/ou ações tomadas; metadados de preservação registrados, armazenados e vinculados a objetos digitais pertinentes; e recibos de confirmação enviados de volta aos fornecedores. Esses registros devem ser criados no momento ou próximo às ações a que se referem e são relacionadas às ações realizadas durante o processo de *ingest*. Os registros podem ser automatizados ou podem ser escritos por indivíduos, dependendo da natureza das ações descritas.

b) *ingest*: criação do *Archival Information Package* (AIP)

- o repositório deve ter para cada AIP ou classe de AIPs preservada pelo repositório uma definição associada que é adequada para analisar o AIP e adequada para a necessidades de preservação de longo prazo: isso é necessário para garantir que o AIP e sua definição associada seja inclusa. As informações de embalagem sempre podem ser encontradas, processadas e gerenciadas dentro do arquivo.

- o repositório deve ser capaz de identificar qual definição se aplica a qual AIP: esse requisito é necessário para garantir que a definição apropriada seja usada ao analisar/interpretar um AIP. Exemplos pelas quais o repositório pode demonstrar que está atendendo a esse requisito será utilizando documentação que vincula claramente cada AIP, ou classe de AIPs, à sua definição. O repositório pode usar qualquer método para associar as definições e os AIPs que fornece para a ligação contínua das duas entidades.
- o repositório deve ter uma definição de cada AIP que seja adequada para preservação de longo prazo, permitindo a identificação e análise de todos os componentes: isso é necessário para mostrar explicitamente que os AIPs são adequados para a finalidade pretendida, que cada componente foi adequadamente concebido e executado e os planos para a manutenção estão em vigor. Maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito será através da demonstração do uso das definições para extrair informações de conteúdo e Informações de Descrição de Preservação/*Preservation Description Information* (PDI) de AIPs. Deve-se notar que a proveniência de um objeto digital, por exemplo, pode ser estendida ao longo das ações de preservação adicionais e a documentação deve identificar cada classe de AIP e descrever como cada uma é implementada

dentro do repositório. As implementações podem, por exemplo, envolver alguma combinação de arquivos, bases de dados e/ou documentos. Ainda assim, o repositório deve identificar claramente quando novas versões de AIPs precisam ser criadas para mantê-los adequados para o propósito.

- o repositório deve ter uma descrição de como os AIPs são construídos a partir dos SIPs: isso é necessário para garantir que o(s) AIP(s) representem adequadamente as informações em SIP(s). Formas pelas quais o repositório pode demonstrar que está atendendo a esse requisito será com a utilização de documentos de descrição do processo; documentação da relação SIP-AIP; e documentar como os AIPs são derivados dos SIPs. Em alguns casos, o AIP e o SIP serão quase idênticos, exceto pela embalagem e localização, e o repositório precisa apenas declarar isso. Em outros casos, transformações complexas podem ser aplicadas a objetos durante o processo de *ingest*, e uma descrição precisa dessas ações podem ser necessárias para refletir como o AIP foi adequadamente transformado das informações no SIP.

- o repositório deve documentar a disposição final de todos os SIPs. Em particular, o seguinte aspecto deve ser verificado:

- o repositório deve seguir procedimentos documentados se um SIP não for incorporado em um AIP ou descartado e deve indicar porque o SIP não foi incorporado ou descartado: isso é necessário para garantir que os SIPs recebidos foram tratados de forma adequada, e, em particular, não foram perdidos acidentalmente. O repositório pode demonstrar que está atendendo a esses requisitos através dos arquivos de processamento do sistema; registros de descarte; acordos/escrituras de doadores ou depositantes; sistema de rastreamento de proveniência; arquivos de log do sistema; documentos de descrição do processo; documentação da relação SIP com AIP; documentação clara de como os AIPs são derivados dos SIPs; documentação do padrão/processo contra o qual ocorre a normalização; documentação de resultado da normalização e como o AIP resultante é diferente do(s) SIP(s). A adesão de procedimentos e o processamento interno e logs de auditoria devem manter registros de todos as transformações de SIPs para demonstrar que eles se tornam AIPs (ou parte de AIPs) ou são eliminados. As

informações descritivas apropriadas também devem documentar a proveniência de todos objetos digitais.

- o repositório deve ter e usar uma convenção que gere identificadores persistentes/únicos para todos os AIPs: em particular, os seguintes aspectos devem ser verificados:

- o repositório deve identificar exclusivamente cada AIP e ter identificadores únicos; deve atribuir e manter identificadores persistentes do AIP e seus componentes de forma a serem únicos no contexto; a documentação deve descrever quaisquer processos usados para mudanças em tais identificadores; e o repositório deve ser capaz de fornecer uma lista completa de todos esses identificadores e fazer verificações pontuais de duplicações. Maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito será utilizando documentação que descreve a convenção de nomenclatura e a evidência física de sua aplicação, por exemplo, Histórico.

- o repositório deve ter um sistema de serviços de ligação/resolução confiável a fim de encontrar o objeto identificado com exclusividade, independentemente de sua localização física: esse requisito é necessário para que as ações relacionadas aos AIPs possam ser rastreadas ao longo do tempo, e ao longo das alterações de armazenamento. O repositório pode demonstrar que está atendendo a esse requisito utilizando documentação que descreve a convenção de nomenclatura e a evidência física de sua aplicação (por exemplo, o Histórico). Um repositório precisa garantir que haja uma convenção de nomenclatura padrão aceita que identifica seus materiais de forma única e persistente para uso dentro e fora do repositório. O requisito de "visibilidade" aqui significa "visível" para os gerentes de repositório e auditores. Isso não significa que esses identificadores exclusivos precisam estar visíveis para os usuários finais ou que eles servem como o principal meio de acesso a objetos digitais. Os componentes de um AIP devem ser adequadamente vinculados e identificados para gestão de longo prazo, mas não impõe restrições sobre como os AIPs são identificados com os arquivos. Assim, no caso geral, um AIP pode ser distribuído por vários arquivos, ou um único arquivo pode conter mais de um AIP.

Portanto, identificadores e nomes de arquivos podem não corresponder necessariamente entre si e a documentação deve representar esses relacionamentos.

- o repositório deve ter acesso às ferramentas e recursos necessários para fornecer informações de representação para todos os objetos digitais que ele contém. Em particular, os seguintes aspectos devem ser verificados:

- o repositório deve ter ferramentas ou métodos para identificar o tipo de arquivo de todos objetos de dados enviados.

-o repositório deve ter ferramentas ou métodos para determinar qual representação da informação é necessária para tornar cada objeto de dados compreensível para a designada comunidade.

- o repositório deve ter acesso às informações de representação necessárias.

- o repositório deve ter ferramentas ou métodos para garantir que as informações de representação são persistentemente associadas aos objetos de dados relevantes: este requisito é necessário para garantir que os objetos digitais do repositório sejam compreensíveis para comunidade designada. O repositório pode demonstrar que está atendendo a este requisito através de assinatura ou acesso a registros de informações de representação; registros visíveis (com links persistentes para objetos digitais); registros de banco de dados que incluem informações de representação e um link persistente para objetos digitais. Essas ferramentas e recursos podem ser mantidos internamente ou podem ser compartilhados por meio, por exemplo, de um confiável conjunto de registros. No entanto, este requisito não exige que cada repositório tenha tais ferramentas e recursos, apenas que tenha acesso a elas.

- o repositório deve ter processos documentados para a aquisição de *Preservation Description Information* (PDI) para suas informações de conteúdo associadas e adquirir PDI de acordo com os processos documentados. Em particular, os seguintes aspectos devem ser verificados:

- o repositório deve ter processos documentados para aquisição de PDI.

- o repositório deve executar seus processos documentados para aquisição de PDI.

- o repositório deve garantir que o PDI esteja persistentemente associado as informações de conteúdo relevantes: isso é necessário para garantir que uma trilha auditável para apoiar as reivindicações de autenticidade esteja disponível, que alterações não autorizadas nos acervos digitais possam ser detectadas, e que os objetos digitais possam ser identificados e colocados em seu contexto apropriado. O repositório pode demonstrar que está atendendo a esses requisitos através de procedimentos operacionais padrão; manuais que descrevem os procedimentos de *ingest*; documentação visível sobre como o repositório adquire e gerencia a descrição de preservação O PDI é necessário não apenas para o repositório, mas para ajudar a garantir que as informações de conteúdo não sejam corrompidas (*Fixity*). O PDI deve estar permanentemente associado às informações de conteúdo.

- o repositório deve garantir que as informações de conteúdo dos AIPs sejam compreensíveis para a sua comunidade designada no momento da criação do AIP. Em particular, os seguintes aspectos devem ser verificados:

- o repositório deve ter um processo documentado para testar a compreensibilidade para suas comunidades designadas das informações de conteúdo dos AIPs no momento de sua criação.

- o repositório deve executar o processo de teste para cada classe de conteúdo e informações dos AIPs.

- o repositório deve trazer as informações de conteúdo do AIP até o nível necessário de compreensibilidade, o que é necessário para garantir que os acervos digitais sejam compreensíveis por sua comunidade designada. O repositório pode demonstrar que está atendendo a esses requisitos através de procedimentos de teste a serem executados contra os acervos digitais para garantir sua compreensibilidade para a comunidade designada definida; registros de tais testes sendo realizados e avaliados; evidências de coleta ou identificação de informações de representação para preencher quaisquer lacunas de inteligibilidade que foram encontrados; e retenção de indivíduos com experiência em disciplina. Este requisito está relacionado com a

compreensibilidade do AIP. Se o material *ingest* não for compreensível, o repositório precisa gerenciar ou disponibilizar informações adicionais para se certificar de que os AIPs são compreensíveis para a comunidade designada.

- o repositório deve verificar cada AIP quanto à integridade e exatidão: exemplos pelos quais o repositório pode demonstrar que está atendendo a esse requisito será através da descrição do procedimento que verifica se os AIPs estão completos e corretos. O repositório deve ter certeza de que os AIPs que ele cria são como se espera que sejam. Se o repositório tiver um processo padrão para verificar os SIPs quanto à integridade e exatidão e comprovadamente um processo correto para transformar SIPs em AIPs, então ele simplesmente precisa demonstrar que as verificações iniciais foram realizadas com sucesso e que o processo de transformação foi realizado sem indicar erros. Por outro lado, os repositórios que devem criar processos para muitos de seus AIPs também precisarão gerar métodos exclusivos para validar a integridade e exatidão. Isso pode incluir a realização de testes de algum tipo no conteúdo do AIP que pode ser comparado com os testes do SIP. Esses testes podem ser simples (contando o número de registros em um arquivo, ou realizando alguma medida estatística simples), mas eles também podem ser complexos. A documentação deve descrever como a integridade e a exatidão dos AIPs é garantida, começando com o recebimento do produtor e continuando através da criação de AIP e apoiando a preservação a longo prazo.

- o repositório deve fornecer um mecanismo independente para verificar a integridade da coleção/conteúdo do repositório: é necessário para permitir a auditoria da integridade da coleção como um todo. O repositório pode demonstrar que está atendendo a esse requisito utilizando acordos documentados negociados entre o produtor e o repositório; logs de material recebido e associado a datas de ação (recebimento, ação, etc.); e logs de verificações periódicas. É responsabilidade do repositório escolher o mecanismo apropriado para verificar a integridade e exatidão de suas coleções. Em geral, é provável que um repositório que atende a todos os critérios anteriores irá satisfazer este sem a necessidade de demonstrar nada mais. Como um requisito separado, demonstra a importância de ser capaz de auditar a integridade da coleção como um todo e o repositório deve ser capaz de mostrar, para cada item de seu cadastro de acessos, quais AIP(s) possuem conteúdo desse item. Como alternativa, pode ser necessário mostrar que não há AIP para um item, também porque a *ingest* ainda está em

andamento ou porque o item foi rejeitado por algum motivo. Por outro lado, qualquer AIP deve ser capaz de ser relacionado a uma entrada no registro de aquisições.

- o repositório deve ter registros contemporâneos de ações e processos de administração relevantes para a criação de AIP: o repositório pode demonstrar que está atendendo a esse requisito utilizando documentação escrita das decisões e/ou ações realizadas com carimbo de data/hora; e preservação de metadados registrados, armazenados e vinculados a objetos digitais pertinentes. Esses registros devem ser criados no momento ou próximo a essas ações a que se referem e estão relacionadas às ações associadas à criação de AIP. Os registros podem ser automatizados ou podem ser escritos por indivíduos, dependendo da natureza das ações descritas.

c) planejamento de preservação

- o repositório deve ter estratégias de preservação documentadas relevantes: isso é necessário para que fique claro como o repositório planeja garantir que a informação permanecerá disponível e utilizável para as gerações futuras e para fornecer um meio de verificar e validar o trabalho de preservação do repositório. O repositório pode demonstrar que está atendendo a esse requisito através da documentação que identifica cada risco de preservação identificado e a estratégia para lidar com esse risco. Estas estratégias de preservação documentadas irão descrever como o repositório irá agir sobre riscos identificados, como parte do plano estratégico de preservação e normalmente abordará a degradação da mídia de armazenamento, a obsolescência das unidades de mídia e a obsolescência ou inadequação da representação das informações (incluindo formatos) como a base de conhecimento da comunidade designada. Por exemplo, se a migração for a abordagem escolhida para alguns desses problemas, também precisa haver preservação.

- o repositório deve ter mecanismos para monitorar sua preservação: isso é necessário para que o repositório possa reagir às mudanças e, assim, garantir que as informações preservadas permanecem compreensíveis e utilizáveis pela comunidade designada. O repositório pode demonstrar que está atendendo a esse requisito através de pesquisas da comunidade designada. O repositório deve mostrar que tem algum mecanismo ativo para garantir que as informações permanecem compreensíveis e utilizáveis. Para a maioria dos

repositórios, a preocupação será com a representação das informações utilizadas para preservação, que pode incluir informações sobre como lidar com um formato de arquivo ou software que pode ser usado para renderizar ou processar. Às vezes, o formato precisa mudar porque o repositório não pode mais lidar com isso.

> **-** o repositório deve ter mecanismos para monitoramento e notificação quando as informações de representação são inadequadas para a comunidade designada compreender os acervos de dados. Este requisito é necessário para garantir que as informações preservadas permaneçam compreensíveis e utilizáveis Desta forma o repositório pode demonstrar que está atendendo a esse requisito por meio da assinatura de serviço de registro de representação da informação; e pesquisas entre os membros da comunidade designada. Deve-se mostrar que possui algum mecanismo ativo para alertar sobre iminentes obsolescências.

- o repositório deve ter mecanismos para alterar seus planos de preservação como um resultado de suas atividades de monitoramento: isso é necessário para que o repositório esteja preparado para mudanças no sistema, o que pode fazer com que seus planos de preservação atuais sejam uma escolha ruim. Exemplos de maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito será desenvolvendo planos de preservação vinculados a vigilância de tecnologia formal ou informal; planejamento de preservação ou processos que são programados para intervalos mais curtos (por exemplo, não mais de cinco anos); atualizações das políticas de preservação e planos de preservação; e seções das políticas de preservação que abordam como os planos podem ser atualizados e que abordam a frequência com que os planos devem ser revisados, reafirmados ou atualizados. O repositório deve demonstrar ou descrever como ele reage às informações do monitoramento e revisar periodicamente seus planos de preservação e o ambiente de tecnologia e, se necessário, fazer alterações nesses planos para garantir sua eficácia continuada.

> - o repositório deve possuir mecanismos de criação, identificação ou coleta de qualquer representação da informação: este critério é necessário para garantir que as informações preservadas permaneçam compreensíveis e utilizáveis pela comunidade designada, por exemplo, através de assinatura de um serviço de registro de formato; assinatura de um serviço de observação de tecnologia; e planos de preservação. O repositório deve ter mecanismos para monitoramento e notificação quando

informações de representação (incluindo formatos) se aproximam da obsolescência ou não são mais viáveis, devendo ser capaz de mostrar que dispõe de mecanismos para lidar com tais notificações.

- o repositório deve fornecer evidências da eficácia de sua preservação: isso é necessário para garantir à comunidade designada que o repositório será capaz de disponibilizar a informação e utilizá-la a médio e longo prazo. O repositório pode demonstrar que está atendendo a esse requisito através da coleta de metadados de preservação apropriados; prova de usabilidade de objetos digitais selecionados aleatoriamente mantidos dentro do sistema; e pesquisas da comunidade designada. O repositório deve ser capaz de demonstrar a preservação contínua, incluindo compreensibilidade de suas participações. Isso pode ser avaliado em vários graus e depende da especificidade da comunidade designada e se uma comunidade designada for razoavelmente ampla, um auditor pode representar o sujeito no teste na avaliação.

d) preservação de AIP

- o repositório deve ter especificações de como os AIPs são armazenados até o nível de bits: é necessário para garantir que as informações possam ser extraídas do AIP a longo prazo. Maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito será através da documentação do formato dos AIPs. O repositório deve especificar a representação da informação até o nível de bits de cada componente AIP. Frequentemente, os repositórios são tentados a descrever o conteúdo AIP apenas em um nível onde um programa será então usado para converter as informações em uma forma compreensível para suas comunidades designadas. No entanto, se esses programas nunca funcionarem, então o as informações seriam perdidas em todos os AIPs que dependiam desse programa.

- o repositório deve preservar as informações de conteúdo dos AIPs: isso é necessário porque é missão fundamental de um repositório preservar as informações de conteúdo para suas comunidades designadas. O repositório pode demonstrar que está atendendo a esse requisito através da documentação de procedimento de fluxo de trabalho de preservação; documentos de política de preservação que especificam o tratamento de AIPs e em que circunstâncias eles podem ser excluídos; e capacidade

de demonstrar a sequência de conversões para um AIP para qualquer objeto digital específico ou grupo de objetos *ingest*. O repositório deve ser capaz de demonstrar que os AIPs refletem fielmente as informações que foram capturadas durante a *ingest* e que quaisquer transformações planejadas subsequentes ou futuras continuarão a preservar todas as propriedades de informações necessárias das informações de conteúdo. Uma abordagem a este requisito assume que o repositório tem uma política especificando que os AIPs não podem ser excluídos a qualquer momento. Esta implementação particularmente simples e robusta preserva os links entre o que foi originalmente *ingest*, bem como as novas versões que foram transformadas ou alteradas de qualquer forma. Dependendo da implementação, esses objetos mais novos podem ser AIPs completamente novos ou meramente atualizados. De qualquer forma, links persistentes entre o objeto ingerido e o AIP resultante devem ser mantidos.

- o repositório deve monitorar ativamente a integridade dos AIPs: é necessário para proteger a integridade dos objetos de arquivo ao longo do tempo. O repositório pode demonstrar que está atendendo a esse requisito utilizando informações de fixidez (por exemplo, somas de verificação) para cada objeto digital/AIP ingerido; documentação de como as informações de AIPs e Fixity são mantidas separadas; documentação de como os AIPs e os registros de adesão são mantidos separados. Um repositório deve ter logs que mostrem as ações realizadas para verificar a integridade do arquivo de objetos, a fim de garantir financiadores, produtores e usuários e para permitir que eles auditem/validem se o repositório está tomando as medidas necessárias para garantir a longo prazo integridade dos objetos digitais.

- o repositório deve ter registros contemporâneos de ações e processos de administração relevantes para o armazenamento e preservação dos AIPs: Esse requisito é necessário para garantir que a documentação não seja omitida ou errônea ou de autenticidade questionável. O repositório pode demonstrar que está atendendo esse requisito através da documentação escrita das decisões e/ou ações tomadas; e metadados de preservação registrados, armazenados e vinculados a objetos digitais pertinentes. Os registros podem ser automatizados ou podem ser escritos por indivíduos, dependendo da natureza das ações

descritas. Onde os padrões comunitários ou internacionais são usados, o repositório deve demonstrar que todas as ações relevantes são realizadas de forma adequada.

- o repositório deve ter procedimentos para todas as ações realizadas em AIPs: isto é importante para garantir que quaisquer ações realizadas contra um AIP não alterem as informações de uma maneira inaceitável para suas comunidades designadas. Esse requisito será atendido através da documentação escrita, que descreve todas as ações que podem ser executadas em um AIP. Esta documentação é normalmente criada durante o design do repositório e deve detalhar o tratamento normal, todas as ações que podem ser realizadas contra os AIPs, incluindo condições de sucesso, falha e detalhes de como esses processos podem ser monitorados.

- o repositório deve ser capaz de demonstrar que quaisquer ações tomadas em AIPs estavam em conformidade com a especificação dessas ações: isso é necessário para garantir que quaisquer ações realizadas contra um AIP não alterem as informações de uma maneira inaceitável para suas comunidades designadas. Alguns exemplos pelos quais o repositório pode demonstrar que está atendendo a esse requisito será utilizando metadados de preservação registrados, armazenados e vinculados a objetos digitais pertinentes e a documentação dessa ação; e auditorias procedimentais do repositório mostrando que todas as ações estão em conformidade com os processos documentados. A preservação bem-sucedida da informação no arquivo está fortemente ligada aos seguintes procedimentos estabelecidos e documentados para concluir quaisquer ações que afetem o repositório de dados: quanto mais ocorrer o "manuseio especial" dos dados do repositório e mais frequentemente o 'tratamento especial' não é supervisionado de maneira consistente, o mais provável é que os dados mantidos pelo repositório estejam comprometidos. Quando os procedimentos são seguidos regularmente, qualquer desvio de procedimentos susceptíveis de causar uma alteração nos dados será notado ou, se não for notado, pode ser corrigido, e provável mudança poderia ser identificada no futuro.

e) gestão de informações

- o repositório deve especificar os requisitos mínimos de informação para permitir a comunidade designada descobrir e identificar materiais de interesse: isso é necessário

para possibilitar a descoberta dos acervos do repositório. O repositório pode demonstrar que está atendendo esse requisito através das informações descritivas de recuperação, metadados de descoberta, como *Dublin Core* e outras documentações que descrevem o objeto. O repositório deve ser capaz de lidar com os tipos de solicitações que virão de um usuário.

- o repositório deve capturar ou criar informações descritivas mínimas que certifiquem que estão associadas ao AIP: isso é necessário para garantir que as informações descritivas sejam associadas ao AIP. O repositório pode demonstrar que está atendendo a esse requisito com a utilização de metadados descritivos persistentes; identificador ou localizador único que é associado ao AIP; sistema de documentação e arquitetura técnica; acordos de depositantes; política de metadados, incorporando detalhes dos requisitos de metadados e uma declaração que descreve onde recai a responsabilidade por sua aquisição; e documentação do fluxo de trabalho. O repositório deve mostrar que se associa a cada AIP, ou seja, associar as informações descritivas ao objeto é importante, embora não requerem correspondência direta e podem não ser necessariamente armazenados com o AIP. Esquemas hierárquicos de descrição podem permitir que alguns elementos descritivos sejam associados com muitos itens.

- o repositório deve manter a ligação bidirecional entre cada AIP e suas informações descritivas: este item é necessário para garantir que todos os AIPs possam ser localizados e recuperados. Pode-se demonstrar que está atendendo a esse requisito utilizando-se metadados descritivos; identificador ou localizador único e persistente associado ao AIP; relação documentada entre o AIP e seus metadados; documentação do sistema e arquitetura técnica; e documentação do fluxo de trabalho do processo. Os repositórios devem implementar procedimentos para estabelecer e manter relacionamentos para associar informações descritivas para cada AIP.

> - o repositório deve manter as associações entre seus AIPs e suas informações descritivas ao longo do tempo: isso é necessário para garantir que todos os AIPs possam continuar a ser localizados e recuperados. O repositório pode demonstrar que está atendendo a esse requisito utilizando-se Log, detalhando a manutenção contínua ou verificação da integridade dos dados e suas relações com as informações descritivas associadas, especialmente após o reparo ou modificação do AIP;

> informações descritivas legadas; persistência do identificador ou localizador; relação documentada entre AIP e suas informações descritivas; documentação do sistema e arquitetura técnica; e documentação do fluxo de trabalho do processo. Os repositórios devem implementar procedimentos que os avisem quando a relação entre os dados e as informações descritivas associadas são temporariamente quebrados para garantir que possam ser restauradas.

f) gestão de acesso: o termo "acesso" tem vários sentidos diferentes, incluindo o acesso dos usuários ao sistema de repositório, por exemplo, segurança física e autenticação de usuário, e os diferentes estágios de acesso aos registros fazer uma solicitação, verificar os direitos do solicitante e preparar e enviar um *Dissemination Information Package* (DIP). Esta subseção é dividida em dois requisitos principais: um relacionado com a existência e implementação de políticas de acesso, e uma com a capacidade do repositório de fornecer objetos comprovadamente autênticos como DIPs. Assim, o primeiro requisito se refere a solicitações iniciadas por um usuário e como o repositório as trata para garantir que os direitos e acordos sejam respeitados, que a segurança seja monitorada, que as solicitações sejam atendidas, etc. O segundo requisito se relaciona com o que é entregue ao consumidor e a confiança que pode ser colocada nele. Deve ser entendido que as capacidades e sofisticação do sistema de acesso irão variar dependendo da comunidade designada do repositório e dos mandatos de acesso do repositório. Devido à variedade de repositórios e mandatos de acesso, esses critérios podem ser sujeitos a questões sobre aplicabilidade e interpretação a nível local.

- o repositório deve obedecer às Políticas de Acesso: isso é necessário para garantir que o repositório aborde totalmente todos os aspectos de uso, o que pode afetar a confiabilidade do repositório, particularmente com referência ao apoio da comunidade de usuários. Maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito será utilizando declarações de políticas que estão disponíveis para as comunidades de usuários; informação sobre o usuário; logs e trilhas de auditoria de solicitações de acesso; e testes explícitos de alguns tipos de acesso. Dependendo da natureza do repositório, as Políticas de Acesso podem abranger: declarações sobre o que é acessível a cada comunidade e em que condições; requisitos para autenticação e autorização de acessos; cumprimento dos acordos aplicáveis às condições de acesso; e registro de ações de acesso. O acesso pode ser gerenciado em parte por

computadores e em parte por humanos; verificar passaportes, por exemplo, antes de emitir um ID de usuário e senha pode ser uma forma apropriada de gestão de algumas instituições.

> - o repositório deve registrar e revisar todas as falhas de gerenciamento de acesso: isso é necessário para identificar ameaças à segurança e falhas no sistema de gerenciamento de acesso. O repositório pode demonstrar que está atendendo a esse requisito através de registros de acesso, capacidade do sistema de usar ferramentas automatizadas de análise/monitoramento e gerar mensagens de problema/erro; notas de revisões realizadas ou ações tomadas como resultado de avaliações. Um repositório deve ter algum mecanismo automatizado para observar negações anômalas ou incomuns e usá-los para identificar ameaças de segurança ou falhas no sistema de gerenciamento de acesso, como o acesso negado a usuários válidos.

- o repositório deve seguir políticas e procedimentos que possibilitem a disseminação de objetos digitais que são rastreáveis aos originais, com evidências que apoiam sua autenticidade: este requisito é importante para estabelecer uma cadeia auditável de autenticidade do AIP para disseminar objetos digitais. Maneiras pelas quais o repositório pode demonstrar que está atendendo esse requisito será por meio de documentos de design do sistema; instruções de trabalho (se os DIPs envolverem processamento manual); processo passo a passo; produção de cópia de amostra com comprovação de autenticidade; e documentação de requisitos da comunidade para evidências de autenticidade. Esse requisito garante que as ações de *ingest*, preservação e transformação não percam informações que apoiariam uma trilha auditável de autenticidade entre o objeto original depositado e o eventual objeto disseminado. Um repositório deve registrar os processos para construir os DIPs dos AIPs relevantes e isto é uma parte fundamental para estabelecer que os DIPs reflitam o conteúdo dos AIPs e, portanto, do material original, de forma confiável e consistente. Os DIPs podem ser simplesmente uma cópia dos AIPs ou podem ser resultado de uma simples transformação de formato de um AIP, mas em outros casos, eles podem ser derivados de maneiras complexas.

> - o repositório deve registrar e agir sobre relatos de problemas nos dados ou respostas de usuários: é necessário para que os usuários considerem o repositório como uma fonte confiável de informação. Exemplos pelos quais o repositório pode demonstrar que está atendendo a esse requisito será através de documentos de design do sistema;

instruções de trabalho (se os DIPs envolverem processamento manual); processo passo a passo; registros de pedidos e produção DIP; documentação de relatórios de erros e as ações tomadas. O objetivo do gerenciamento de acesso é garantir que um usuário receba uma mensagem utilizável e correta da versão do objeto digital (ou seja, DIP) que ele solicitou. Um repositório deve mostrar que quaisquer problemas que ocorram e sejam trazidos à sua atenção sejam investigados e resolvidos. Essa capacidade de resposta é essencial para que o repositório seja considerado confiável.

GESTÃO DE RISCO DE INFRAESTRUTURA E SEGURANÇA

a) gestão de risco de infraestrutura técnica

- o repositório deve identificar e gerenciar os riscos às suas operações de preservação e objetivos associados à infraestrutura do sistema: isso é necessário para garantir uma infraestrutura segura e confiável. Maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito será utilizando inventário de infraestrutura de componentes do sistema; avaliações periódicas de tecnologia; estimativas da vida útil dos componentes do sistema; exportação de registros autênticos para um sistema independente; uso de software fortemente suportado pela comunidade, por exemplo, Apache, iRODS, Fedora; e recriação de arquivos de backups. O repositório deve realizar ou contratar avaliações dos riscos relacionados ao hardware e infraestrutura de software e procedimentos operacionais. O repositório deve fornecer mecanismos que minimizam o risco de dependências em sistema proprietário ou obsoleto e de erro operacional. Também deve manter um sistema que é evolutivo, ou seja, deve ser projetado de tal forma que os principais componentes possam ser substituídos por tecnologias mais novas sem grandes interrupções do sistema como um todo. O sistema de repositório deve ser extensível, ou seja, deve ser projetado para acomodar formatos futuros (mídia e arquivos). O repositório deve ser capaz de exportar seus acervos para um futuro custodiante e ser capaz de recriar os arquivos após um erro de operação que substitui ou exclui acervos digitais.

- o repositório deve empregar vigilância tecnológica ou outra tecnologia de monitoramento de sistemas de notificação: este requisito é necessário para rastrear quando os componentes de hardware ou software se tornarão obsoletos e a migração será necessária para uma nova infraestrutura. O repositório pode demonstrar que está

atendendo a esse requisito por meio do gerenciamento de relatórios periódicos de avaliação de tecnologia e comparação de tecnologia existente a cada nova avaliação. O objetivo é entender quando algum subsistema apresenta risco de obsolescência, e possibilitar a migração para novas tecnologias antes que os mecanismos de interoperabilidade não sejam mais acessíveis. Isso pode ser impulsionado por dependências de software proprietário (o fornecedor não suporta o componente do subsistema), e pelo surgimento de novos protocolos (o mecanismo para acessar o sistema se tornou obsoleto e não é mais suportado). - o repositório deve possuir tecnologias de hardware adequadas aos serviços fornecidos às suas comunidades designadas: este item é importante para fornecer níveis de serviço esperados, contratados, seguros e persistentes incluindo facilidade de *ingest* e disseminação por meio de depositante apropriado, interfaces de usuário e tecnologias como mecanismos de *upload*; gerenciamento contínuo de objetos digitais; abordagens e soluções de preservação, como migração e segurança do sistema. Através da manutenção de tecnologia, expectativas e uso atualizados da comunidade designada; fornecimento de largura de banda adequada para suportar as demandas de *ingest* e uso; feedback sobre a adequação de hardware e serviço; manutenção de uma corrente inventário de hardware são exemplos de como o repositório pode demonstrar que está de acordo com o requisito. O repositório deve estar ciente dos tipos de armazenamento, gerenciamento de arquivos, preservação e serviços de acesso esperados por sua comunidade designada, incluindo quando aplicável, o tipos de mídia a serem entregues e precisa ter certeza de que seus recursos de hardware podem suportar esses serviços. O objetivo é rastrear quando as mudanças nos requisitos exigem uma mudança correspondente na tecnologia de hardware, quando mudanças nas políticas de *ingest* requerem recursos expandidos, e quando mudanças nas políticas de preservação requerem novos recursos de preservação. Isso pode ser impulsionado por mudanças em requisitos de capacidade (o tempo necessário para ler todas as mídias é maior do que a vida útil da mídia), por mudanças nos mecanismos de entrega (novos clientes para exibir registros autênticos), e mudanças no número e tamanho dos registros arquivados. - o repositório deve ter procedimentos para monitorar e receber notificações quando mudanças de tecnologia de hardware são necessárias: é necessário para garantir níveis de serviço esperados, contratados, seguros e persistentes. Maneiras pelas quais o repositório pode

demonstrar que está atendendo a esse requisito será realizando auditorias de capacidade versus uso real; auditorias de taxas de erro observadas; auditorias de desempenho que limitam a capacidade de atender aos requisitos de acesso da comunidade de usuários; documentação de avaliações sobre observação de tecnologia; e documentação de atualizações de tecnologia de fornecedores. O repositório deve realizar ou contratar varreduras ambientais frequentes em relação ao status do hardware, fontes de falha e interoperabilidade entre componentes de hardware. O repositório também deve entrar em contato com seus fornecedores de hardware sobre atualizações de tecnologia, pontos de provável falha e como os novos componentes podem afetar a integração e o desempenho do sistema. O objetivo é rastrear quando as mudanças nos requisitos de serviço pelas comunidades designadas requerem uma mudança correspondente na tecnologia de hardware, quando mudanças na *ingest* das políticas exigem recursos expandidos, e quando as mudanças nas políticas de preservação exigem novas capacidades de preservação. Isso pode ser impulsionado por mudanças nos requisitos de capacidade (o tempo necessário para ler todas as mídias é maior do que o tempo de vida da mídia), por mudanças na entrega (novos clientes para exibir registros autênticos), e mudanças no número e tamanho dos registros arquivados. - o repositório deve ter procedimentos em vigor para avaliar quando as mudanças são necessárias para o hardware atual (importante para garantir que o repositório tenha a capacidade de tornar informado e oportuno decisões quando as informações indicam a necessidade de novo hardware). Exemplos pelos quais o repositório pode demonstrar que está atendendo a esse requisito será por meio de procedimentos de avaliação implementados; e experiência documentada da equipe em cada subsistema de tecnologia. Recebidas informações de vigilância tecnológica ou outra notificação de monitoramento de tecnologia e sistemas, o repositório deve ter procedimentos e conhecimentos para avaliar esses dados e fazer decisões acertadas em relação à necessidade de novo hardware. O objetivo é rastrear quando fornecedores de tecnologia desenvolveram subsistemas que minimizam o risco, ou que minimizam o custo, ou que melhoram o desempenho. Isso é necessário para rastrear tecnologias emergentes e planejar atualizações antes que os limites de capacidade ocorram. A avaliação deve identificar quando o risco de usar a nova tecnologia supera o benefício esperado, e quando a nova tecnologia é suficientemente amadurecida para minimizar o risco. - o repositório deve ter

procedimentos, compromisso e financiamento para substituir o hardware quando a avaliação indica a necessidade de fazê-lo. isso é necessário para garantir a substituição do hardware em tempo hábil, a fim de evitar a falha no sistema ou inadequação de desempenho. Sem esse compromisso, e mais importante, sem recursos financeiros garantidos ou um fluxo de financiamento seguro, as notificações têm pouco valor. O repositório deve ter mecanismos para avaliar a eficácia dos novos sistemas antes da implementação no sistema de produção. Formas pelas quais o repositório pode demonstrar que está atendendo a esse requisito será com a declaração de compromisso de fornecer os níveis de serviços previstos e contratados; evidência de ativos financeiros contínuos reservados para aquisição de hardware; e demonstração de redução de custos através do custo amortizado do novo sistema. O objetivo é demonstrar que o repositório tem a capacidade de incorporar novas tecnologias, tanto financeiramente por meio de compromissos de financiamento ou redução de custos, e operacionalmente através da verificação das capacidades dos novos sistemas. - o repositório deve ter tecnologias de software adequadas aos serviços que são fornecidos às comunidades designadas: esse item é necessário para fornecer níveis de serviço esperados, contratados, seguros e persistentes incluindo a facilidade de *ingest* e disseminação por meio de depositante apropriado e interfaces de usuário e tecnologias como mecanismos de *upload*; gerenciamento contínuo de objetos digitais; e abordagens e soluções de preservação, como migração e segurança do sistema. O repositório pode demonstrar que está atendendo a esse requisito através da manutenção de tecnologia, expectativas e uso atualizados da comunidade designada; fornecimento de sistemas de software adequados para suportar as demandas de *ingest* e uso; obtenção sistemática de feedback sobre a adequação do software e do serviço; e manutenção de um inventário de software atual. O objetivo é rastrear quando as mudanças nos requisitos de serviço pelas comunidades designadas requerem uma mudança correspondente nos componentes de software, quando mudanças nas políticas de *ingest* requerem suporte para novos formatos de dados e quando mudanças no software requerem novos recursos de migração nos formatos. Isso pode ser impulsionado por mudanças em requisitos de acesso (novos clientes que exigem novos formatos de dados tornam-se preferidos), por mudanças nos mecanismos de entrega (novos mecanismos de transferência de dados) e mudanças no número e tamanho dos registros arquivados que requerem software mais

escalonável. - o repositório deve ter procedimentos para monitorar e receber notificações quando mudanças de software são necessárias: este critério é necessário para garantir níveis de serviço esperados, contratados, seguros e persistentes. O repositório pode demonstrar que está atendendo a esse requisito com auditorias de capacidade versus uso real; auditorias de taxas de erro observadas; auditorias de desempenho e gargalos que limitam a capacidade de atender aos requisitos de acesso da comunidade de usuários; documentação de avaliações de observação de tecnologia; e documentação de atualizações de software de fornecedores. O objetivo é rastrear quando as mudanças nos requisitos de serviço pelas comunidades designadas exigem uma mudança correspondente na tecnologia de software, quando mudanças nas políticas de *ingest* exigem recursos expandidos e quando as mudanças nas políticas de preservação requerem novos recursos de preservação. Isso pode ser conduzido por atualizações de segurança (correções fornecidas para vulnerabilidades recentemente identificadas), por mudanças nos mecanismos de entrega (novos clientes de software para exibir registros autênticos) e mudanças no número e tamanho de registros arquivados (requisitos de banco de dados expandidos). O repositório deve conduzir ou contratar varreduras ambientais frequentes em relação à evolução do software, prováveis pontos de falha, e interoperabilidade entre os componentes de software e hardware. O repositório deve também estar em contato com seus fornecedores de software sobre atualizações de tecnologia, pontos de prováveis falhas e como os novos programas podem afetar a integração e o desempenho do sistema. - o repositório deve ter procedimentos em vigor para avaliar quando as mudanças são necessárias para o software atual: é necessário para garantir que o repositório tenha a capacidade de tornar informado e oportuno decisões quando as informações indicam a necessidade de um novo software. Exemplos pelos quais o repositório pode demonstrar que está atendendo a esse requisito será através de procedimentos de avaliação implementados; e experiência documentada da equipe em cada tecnologia de software. Recebidas informações de vigilância tecnológica ou outra notificação de monitoramento de tecnologia de sistemas, o repositório deve ter procedimentos e conhecimentos para avaliar esses dados e fazer decisões acertadas em relação à necessidade de um novo software. O objetivo é rastrear quando fornecedores de tecnologia desenvolveram infraestrutura de software que minimiza o risco, ou que minimiza custos ou melhora o desempenho. Isso é importante para

rastrear emergentes tecnologias e planejar atualizações antes que os limites de capacidade ocorram. A avaliação deve identificar quando o risco de usar uma nova tecnologia supera o benefício esperado, e quando o nova tecnologia é suficientemente madura para minimizar o risco. - o repositório deve ter procedimentos, compromisso e financiamento para substituir software quando a avaliação indica a necessidade de fazê-lo. Este critério é importante para garantir a substituição do software em tempo hábil, a fim de evitar a falha do sistema ou inadequação de desempenho. Sem esse compromisso, ou sem recursos financeiros garantidos, a vigilância tecnológica e as notificações têm pouco valor. O repositório deve ter mecanismos para avaliar a eficácia dos novos sistemas antes da implementação no sistema de produção. Exemplos de maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito será com a declaração de compromisso de fornecer os níveis de serviços previstos e contratados; evidência de ativos financeiros contínuos reservados para aquisição de software; e demonstração de redução de custos através do custo amortizado do novo sistema. O objetivo é demonstrar que o repositório tem a capacidade de incorporar novas tecnologias, tanto financeiramente por meio de compromissos de financiamento ou redução de custos, e operacionalmente através da verificação das capacidades dos novos sistemas.

- o repositório deve ter suporte adequado de hardware e software para backup e funcionalidade suficiente para preservar o conteúdo do repositório e rastrear as funções do repositório: isso é importante para garantir o acesso contínuo e o rastreamento das funções de preservação aplicadas aos objetos digitais sob sua custódia. O repositório pode demonstrar que está atendendo a esse requisito documentando como está sendo feito o backup e com que freqüência; log de auditoria/inventário de backups; validação de backups concluídos; plano, política e documentação de recuperação de desastres; treinamentos; teste de backups; contratos de suporte para hardware e software para backup; e preservação demonstrada de metadados do sistema como controles de acesso, localização de réplicas, trilhas de auditoria, valores de *checksum*. O repositório deve ser capaz de demonstrar a adequação dos processos, hardware e software para seus sistemas de backup e toda a gama de *ingest*, preservação e disseminação exigidas de um repositório encarregado

de preservação a longo prazo. Backup simples devem preservar não apenas o conteúdo principal do repositório, mas também o sistema de metadados gerados pelas funções de preservação. Os repositórios precisam desenvolver planos de backup que garantem a continuidade das operações em todos os modos.

- o repositório deve ter mecanismos eficazes para detectar erros nos dados ou perda de bits: é necessário para garantir que os AIPs e os metadados não estejam corrompidos ou para quaisquer perdas de dados detectadas que se enquadram nas tolerâncias estabelecidas pela política de repositório. O repositório pode demonstrar que está atendendo a esse requisito com a utilização de documentos que especificam os mecanismos de detecção e correção de erros de bits usados; análise de risco; relatórios de erros; análise de ameaças; e análise periódica da integridade dos acervos do repositório. O objetivo é um tratamento abrangente das perdas de fontes e de dados. Quaisquer dados ou metadados que estão temporariamente perdidos devem ser recuperáveis por backups. As falhas sistemáticas de rotina não devem se acumular e causar perda de dados além das tolerâncias estabelecidas pelas políticas do repositório. Mecanismos como somas de verificação (Assinaturas MD5) ou assinaturas digitais devem ser reconhecidas por sua eficácia em detecção de perda de bits e incorporado na abordagem geral do repositório para validação da integridade. - o repositório deve registrar e relatar à sua administração todos os incidentes, erros nos dados ou perda de dados e medidas devem ser tomadas para reparar/substituir dados corrompidos ou perdidos: isso é necessário para garantir que a administração do repositório seja mantida informada sobre incidentes e ações de recuperação, e para permitir a identificação de fontes de erros nos dados ou perdas. O repositório pode demonstrar que está atendendo a esse requisito através de procedimentos relacionados ao relato de incidentes aos administradores; metadados de preservação (por exemplo, PDI) registros; comparação de logs de erros com relatórios para administração; procedimentos de escalação relacionados à perda de dados; rastreamento de fontes de incidentes; ações de remediação tomadas para remover fontes de incidentes. Ter mecanismos eficazes para detectar erros nos dados e perda de bits dentro de um sistema de repositório é crítica, mas é apenas a etapa inicial de um processo maior. Além de registrar, relatar, e reparar o mais rápido possível todas as violações de integridade de dados, esses incidentes e as ações de recuperação e

seus resultados devem ser relatados aos administradores e disponibilizados a todo pessoal relevante. Dada a identificação das fontes de perda de dados, uma avaliação das revisões para sistemas de software e hardware, ou procedimentos operacionais ou políticas de gerenciamento são necessários para minimizar o risco futuro de perda de dados.

- o repositório deve ter um processo para registrar e reagir à disponibilidade de novas atualizações de segurança com base em uma avaliação de risco-benefício: este requisito é necessário para proteger a integridade dos objetos de arquivos não autorizados de alterações ou exclusões. O repositório pode demonstrar que está atendendo a esse requisito através de registro de risco (lista de todos os *patches* disponíveis e análise de documentação de risco); evidência de processos de atualização (por exemplo, *daemon* do gerenciador de atualização do servidor); e documentação relacionada à atualização de instalações. As decisões de aplicar atualizações de segurança provavelmente serão o resultado de uma avaliação de risco-benefício e *patches* de segurança são frequentemente responsáveis por perturbar aspectos alternativos do sistema, funcionalidades ou desempenho. Pode não ser necessário para um repositório implementar todos *patches* de software e a aplicação deve ser cuidadosamente considerada. Cada atualização de segurança implementada pelo repositório deve ser documentada com detalhes sobre como é completada; e atualizações automáticas e manuais são aceitáveis. Atualizações de segurança significativas podem pertencer a software diferente de sistemas operacionais centrais, como aplicativos de banco de dados, servidores da Web, e estes também devem ser documentados. As atualizações de segurança não se limitam a atualizações de segurança de software e as atualizações no hardware real ou no *firmware* do sistema de hardware estão incluídos. Com o tempo, é provável que as atualizações de segurança também sejam necessárias para o repositório. Embora as atualizações de segurança possam ser consideradas como parte do controle de mudança, elas são identificadas separadamente aqui porque muitas vezes há serviços que compilam e divulgam informações sobre problemas e atualizações de segurança. Desta forma, os repositórios devem monitorar esses serviços para garantir que o repositório que mantém os dados não esteja sujeito a comprometimento por ameaças identificadas.

- o repositório deve ter processos definidos para mídia de armazenamento e/ou mudança de hardware (por exemplo, atualização e migração): isso é necessário para garantir que os dados não sejam perdidos quando a mídia falhar ou o hardware de suporte não poder mais ser usado para acessar os dados. Maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito será por meio da documentação de processos de migração; políticas relacionadas ao suporte, manutenção de hardware, e substituição; documentação dos ciclos de vida de suporte esperados do fabricante do hardware; e políticas relacionadas à migração de registros para sistemas de hardware alternativos. O repositório deve ter estimativas da velocidade de acesso e da quantidade de informações para cada tipo de mídia de armazenamento. Em seguida, com estimativas da vida útil confiáveis o repositório pode estimar o tempo necessário para migração de mídia de armazenamento, ou atualização ou cópia entre mídias sem reformatar o fluxo de bits. O repositório pode então definir gatilhos para iniciar a ação em um momento apropriado, portanto, as ações serão concluídas antes que os dados sejam perdidos. Copiar grandes quantidades de dados pode demorar muito tempo e pode afetar outras métricas de desempenho do sistema. Também devem considerar a obsolescência de todo e qualquer componente de hardware dentro do sistema do repositório como possíveis eventos de gatilho para migração.

- o repositório deve ter processos críticos identificados e documentados que podem afetar sua capacidade de cumprir com suas responsabilidades obrigatórias: isso é necessário para garantir que os processos críticos possam ser monitorados para garantir que eles continuem a cumprir as responsabilidades obrigatórias e a garantir que quaisquer alterações a esses os processos sejam examinados e testados. O repositório pode demonstrar que está atendendo a esse requisito utilizando a matriz de rastreabilidade entre processos e requisitos obrigatórios. Exemplos de processos críticos incluem o gerenciamento de dados, acesso, armazenamento de arquivos, *ingest* e processos de segurança. - O repositório deve ter um processo de gerenciamento de mudanças documentado que identifica mudanças em processos críticos que afetam potencialmente a capacidade de cumprir suas responsabilidades obrigatórias: é necessário para garantir que o repositório possa especificar não apenas o atual processo, mas os processos anteriores que foram aplicados aos acervos do

repositório. O repositório pode demonstrar que está atendendo a esse requisito através da documentação do processo de gerenciamento de mudanças; avaliação do risco associado; análise do impacto esperado; e comparação de registros de reais mudanças nos processos versus análises associadas de seu impacto e criticidade. Exemplos disso incluem mudanças nos processos de gerenciamento de dados, acesso, arquivamento armazenamento, *ingest* e segurança. É realmente importante saber quais mudanças foram feitas e quando elas foram feitas. A rastreabilidade torna possível entender o que foi afetado por mudanças específicas nos sistemas. Se as consequências indesejadas ocorrerem mais tarde, então ter este registro pode tornar possível reverter as mudanças ou, pelo menos, documentar as que foram introduzidas. A gestão da mudança é um componente do tópico mais amplo de gerenciamento de configuração descrito pela ISO 10007: 2003, que inclui planejamento de gerenciamento de configuração, identificação de configuração, controle de mudança, contabilidade de status de configuração e auditoria de configuração. Esforços de gerenciamento de configuração deve resultar em uma trilha de auditoria completa de decisões e modificações de design. - o repositório deve ter um processo para testar e avaliar o efeito de mudanças nos processos críticos do repositório: é necessário para proteger a integridade dos processos críticos do repositório, e como que eles continuam em sua capacidade de atender aos requisitos obrigatórios. Pode-se demonstrar que está atendendo a esse requisito com procedimentos de teste documentados; documentação de resultados de testes anteriores e prova de alterações feitas como resultado de testes; e análise do impacto de uma mudança de processo. Mudanças em sistemas críticos devem ser, sempre que possível, pré-testados separadamente. Após as mudanças, os sistemas devem ser monitorados quanto ao comportamento inesperado e inaceitável. Se tal comportamento for descoberto, as mudanças e suas consequências devem ser revertidas. O teste de todo o sistema ou o teste de unidade podem atender a este requisito e não são necessários testes complexos de tipo de segurança.

- o repositório deve gerenciar o número e localização de cópias de todos os objetos digitais: isso é necessário para garantir que o repositório forneça uma cópia autêntica de um objeto digital específico. Exemplos de maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito será realizando testes de recuperação aleatórios; validação da

existência do objeto para cada local registrado; validação de uma localização registrada para cada objeto em sistemas de armazenamento; verificação de proveniência e fixidez em formação; registro de localização de objetos digitais em comparação com o número esperado e localização de cópias de objetos específicos. Um repositório pode ter diferentes políticas de preservação para diferentes classes de objetos, dependendo de fatores como o produtor, o tipo de informação ou seu valor. Repositórios podem exigir um número diferente de cópias para cada classe, ou gerenciar as versões necessárias para atender requisitos de acesso. Pode haver requisitos de identificação adicionais se os mecanismos de integridade dos dados usam cópias alternativas para substituir cópias com falha. A localização de cada objeto digital deve ser descrita de modo que possa ser localizado com precisão, sem ambiguidade e o local pode ser físico absoluto ou um local lógico em uma mídia de armazenamento ou um subsistema de armazenamento. As informações de proveniência sobre como copiar e mover os dados devem ser mantidas e atualizadas, incluindo a identificação dos responsáveis. Isso é necessário a fim de rastrear a cadeia de custódia e afirmar que o repositório está fornecendo uma cópia autêntica de um determinado objeto digital. O repositório deve ser capaz de distinguir entre as versões de objetos ou cópias idênticas, ou seja, é necessário para que um repositório possa afirmar que está fornecendo uma cópia autêntica da versão correta de um objeto.

> - o repositório deve ter mecanismos para garantir que qualquer/várias cópias de objetos digitais sejam sincronizadas: este requisito é importante para garantir que várias cópias de um objeto digital permaneçam idênticas, dentro de um prazo estabelecido como aceitável pelo repositório, e que uma cópia possa ser usada para ser substituída por uma cópia corrompida do objeto. O repositório pode demonstrar que está atendendo a esse requisito através de fluxos de trabalho de sincronização; análise do sistema de quanto tempo leva para as cópias serem sincronizadas; e procedimentos/documentação de processos de sincronização. O plano de recuperação de desastres deve abordar o que fazer em caso de desastre e uma atualização coincidir. Por exemplo, se uma cópia de um objeto for alterada e ocorrer um desastre enquanto a segunda está sendo atualizada, deve haver um mecanismo para garantir que a cópia seja atualizada na primeira oportunidade disponível. Os mecanismos para sincronizar cópias de objetos digitais devem ser capazes de detectar erros de bits e validar as verificações de fixidez antes da sincronização ser tentada.

b) gestão de risco de segurança

- o repositório deve manter uma análise sistemática dos fatores de risco de segurança associada a dados, sistemas, pessoal e planta física: é necessário para garantir um serviço contínuo e ininterrupto à comunidade designada. O repositório pode demonstrar que está atendendo a esse requisito quando emprega os códigos de prática encontrados na série ISO 27000 de sistema de padrões. Deve-se realizar avaliações de risco regulares e manter a segurança adequada a fim de fornecer os níveis de serviço previstos e contratados, seguindo os códigos de prática como a ISO 27000. 'Sistema' aqui se refere a mais do que sistemas de TI, como hardware, software, comunicações equipamentos e instalações e firewalls. Sistemas de proteção contra incêndio e detecção de inundação também são significativos, assim como os meios para avaliar os procedimentos de pessoal, gestão e administração, recursos, bem como operações e prestação de serviços. Perda de receita, orçamento e reputação são ameaças significativas às operações gerais, assim como a perda de mandato. Um curso interno e avaliação externa deve ser realizada para avaliar a qualidade do serviço e a relevância para o usuário. Direito de propriedade intelectual também deve ser revisado regularmente, bem como a responsabilidade do repositório para regulamentar a não conformidade. O repositório deve avaliar as habilidades de sua equipe e garantir a aquisição de novos funcionários ou reciclagem do pessoal existente, conforme necessário. A avaliação de risco regular também deve abordar ameaças e ataques de negação de serviço e perda ou qualidade inaceitável de serviços terceirizados

- o repositório deve ter implementado controles para abordar adequadamente cada um dos riscos de segurança definidos: é importante para garantir que os controles estejam em vigor para atender às necessidades de segurança do repositório. Maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito será empregando os códigos de prática encontrados na série de padrões ISO 27000. Deve-se mostrar como lidou com seus requisitos de segurança. Se alguns tipos de materiais tem maior probabilidade de serem atacados, o repositório precisará fornecer mais proteção, por exemplo: repositórios que experimentaram incidentes podem registrar tais instâncias, incluindo os momentos em que os sistemas ou conteúdo foram afetados e descrever os procedimentos que foram implementados para evitar ocorrências semelhantes no futuro. Os repositórios também podem conduzir uma variedade de exercícios de desastres que podem envolver sua organização ou a comunidade.

Os planos de contingência são especialmente importantes e precisam ser testados, atualizados e revisados em uma base regular.

- a equipe do repositório deve ter funções, responsabilidades e autorizações relacionadas à implementação de mudanças no sistema: este critério é importante para garantir que os indivíduos tenham autoridade para implementar mudanças, que recursos adequados foram atribuídos para o esforço, e que os indivíduos responsáveis serão encarregados pela implementação de tais mudanças. O repositório pode demonstrar que está atendendo a esse requisito empregando os códigos de prática encontrados na série de padrões ISO 27000. As autorizações referem-se a quem pode adicionar usuários, quem tem acesso à mudança de metadados, e quem pode acessar os registros de auditoria. É importante que as autorizações sejam justificadas, para o pessoal entender o que estão autorizados a fazer, qual equipe possui as habilidades necessárias associadas a várias funções e autorizações, e que haja uma visão consistente disso em toda a organização.

- o repositório deve ter plano(s) de preparação para a recuperação de desastres por escrito, incluindo pelo menos um backup externo de todas as informações preservadas, juntamente com uma cópia externa do(s) plano(s) de recuperação: é necessário para garantir que recursos suficientes de backup e recuperação estejam disponíveis para facilitar a preservação contínua e o acesso aos sistemas e seu conteúdo com interrupção limitada dos serviços. Formas pelas quais o repositório pode demonstrar que está atendendo a esse requisito será empregando os códigos de prática encontrados na série de padrões ISO 27000. O nível de detalhe em um plano de desastre e os riscos específicos tratados precisam ser apropriados de acordo com a localização do repositório e as expectativas de serviço. O plano de desastre deve, no entanto, lidar com situações não especificadas que teriam consequências específicas, como falta de acesso a um prédio ou doença generalizada entre o pessoal. No caso de um desastre no repositório, o repositório pode querer entrar em contato com órgãos de desastre local e/ou nacional para recuperação e assistência. Os repositórios também podem conduzir uma variedade de exercícios de desastre que pode envolver sua organização principal ou a comunidade em geral.

Na seção 4.1.3 são apresentados os princípios TRUST.

4.1.3 The Trust Principles For Digital Repositories

Conforme o artigo publicado por Dawei Lin *et al.*(2020), os repositórios devem ganhar a confiança das comunidades demonstrando que são confiáveis e capazes de adequadamente gerenciar os dados que mantêm. Nesse contexto, após uma discussão pública e com base no consenso da comunidade, várias partes interessadas representando vários segmentos da comunidade do repositório digital, desenvolveram de forma colaborativa um conjunto de princípios orientadores para demonstrar a confiabilidade de um repositório digital: os princípios TRUST (*Transparency, Responsibility, User focus, Sustainability and Tecnology*) descritos no quadro 5. Esses princípios fornecem uma estrutura comum para facilitar a discussão e implementação das melhores práticas em preservação digital.

Quadro 5 – Princípios TRUST para repositórios digitais

Princípio	Orientação para repositórios
T (transparência)	Ser transparente sobre os serviços de repositório específicos e acervos de dados.
R (responsabilidade)	Ser responsável por garantir a autenticidade e integridade dos acervos de dados e pela confiabilidade e persistência de seu serviço.
U (foco no usuário)	Para garantir que as normas de gerenciamento de dados e expectativas das comunidades de usuários sejam atendidas.
S (sustentabilidade)	Para sustentar serviços e preservar acervos de dados para o longo prazo.
T (tecnologia)	Para fornecer infraestrutura e recursos para oferecer suporte a serviços seguros, persistentes e confiáveis.

Fonte: Dawei Lin *et al*. (2020).

<u>Princípios TRUST</u>

- *Transparency* (Transparência):

A fim de selecionar o repositório mais apropriado para o uso, usuários potenciais se beneficiam ao encontrar e acessar facilmente informações sobre o escopo, comunidade de usuários, políticas e recursos do repositório de dados. A transparência nessas áreas oferece uma oportunidade de aprender sobre o repositório e considerar sua adequação aos requisitos específicos dos usuários, incluindo depósito, preservação e descoberta de dados. Em

conformidade com este princípio, os repositórios devem garantir que, no mínimo, a declaração de missão e o escopo do repositório sejam claramente indicados. Além disso devem estar declarados de forma transparente os seguintes aspectos:

a) termos de uso, tanto para o repositório quanto para os acervos de dados;
b) prazo mínimo de preservação digital para os acervos de dados;
c) quaisquer recursos ou serviços adicionais pertinentes, por exemplo, a capacidade de administrar com responsabilidade dados confidenciais.

Comunicar claramente as políticas e os termos de uso do repositório, informa aos usuários sobre quaisquer limitações que possam restringir o uso dos dados ou do repositório

- *Responsibility* (Responsabilidade):

Os repositórios confiáveis assumem a responsabilidade pela administração de seus acervos de dados e por servir seus usuários. A responsabilidade é demonstrada por:

a) aderir aos padrões de metadados e curadoria, juntamente com o gerenciamento dos acervos de dados, por exemplo, validação técnica, documentação e controle de qualidade;
b) fornecimento de serviços de dados, por exemplo, download de dados;
c) gerenciar os direitos de propriedade intelectual dos produtores de dados, proteção de recursos de informações confidenciais, e a segurança do sistema e seu conteúdo.

Os usuários do repositório devem ter certeza de que os depositantes de dados são solicitados a fornecer todos os metadados compatíveis com as normas da comunidade, o que aumenta a descoberta e a utilização dos dados. Tanto os depositantes quanto os usuários devem ter confiança de que os dados permanecerão acessíveis ao longo do tempo, portanto poderão ser citados e referenciados em publicações acadêmicas. A responsabilidade pode ser esclarecida através de alguns meios legais (direito de preservar) ou pode assumir a forma de cumprimento de alguma norma (padrões éticos).

***User focus* (Foco no usuário):**

Um repositório confiável precisa se concentrar em servir sua comunidade de usuários-alvo. Cada comunidade de usuários provavelmente tem expectativas diferentes em relação aos seus repositórios. Existe uma visão ampla quanto a comunidade de usuários, ou seja, existem aqueles que acessam acervos de dados e partes interessadas indiretas como, financiadores, editores de periódicos, outros parceiros institucionais e cidadãos. O uso e reutilização de dados de pesquisa são parte integrante do processo científico e, portanto, dos repositórios confiáveis, que deve permitir que sua comunidade encontre, explore e entenda seus acervos de dados em relação ao potencial de reuso. Os repositórios têm um papel vital na aplicação e reforço das normas e padrões, o que inclui os esquemas de metadados, formatos de arquivos de dados, vocabulários controlados, ontologias e outras semânticas. Um repositório confiável deve demonstrar:

a) métricas de dados relevantes e disponibilizá-las aos usuários;
b) fornecer catálogos para facilitar a descoberta de dados;
c) monitorar e identificar as expectativas em relação a comunidade e responder conforme necessário.

***Sustainability* (Sustentabilidade):**

Para garantir a sustentabilidade de um repositório confiável é necessário o acesso ininterrupto do seu valioso acervo de dados à comunidade de usuários atuais e futuros. O acesso contínuo aos dados depende da capacidade do repositório fornecer serviços de longo prazo e responder com serviços novos e aprimorados. Um repositório confiável pode demonstrar sustentabilidade de seus acervos através de:

a) planejamento suficiente para mitigação de riscos, recuperação de desastres e sucessão;
b) garantia de financiamento para permitir o uso contínuo e manter as propriedades desejáveis dos recursos de dados que o repositório foi encarregado de preservar e divulgar;
c) fornecimento de governança para a preservação necessária de dados de longo prazo para que os recursos de dados permaneçam detectáveis, acessíveis e utilizáveis no futuro.

***Technology* (Tecnologia):**

Um repositório depende da interação das pessoas, processos e tecnologias para oferecer suporte seguro, persistente e serviços confiáveis. Suas atividades e funções são suportadas por software, hardware e serviços técnicos, ou seja, juntos eles fornecem as ferramentas para permitir a entrega dos princípios TRUST. Um repositório confiável pode demonstrar a adequação de suas capacidades tecnológicas ao:

a) implementar padrões, ferramentas e tecnologias relevantes e apropriadas para gerenciamento e curadoria de dados;
b) ter planos e mecanismos em vigor para prevenir, detectar e responder à ameaças à segurança física ou cibernética.

Quando os repositórios de dados, financiadores e criadores de dados adotar os Princípios FAIR e implementar os Princípios TRUST, os usuários dos repositórios se beneficiarão diretamente ao fazer o uso eficiente e eficaz dos dados. Conforme Dawei Lin *et al.* (2020), vários estudos descobriram que a transparência está associada a confiança dos repositórios digitais e através do processo de obtenção da certificação é possível contribuir para a transparência do repositório, entre outros benefícios.

Um estudo realizado por Yoon (2014), demonstrou como a consciência dos usuários sobre as funções dos repositórios pode ser um fator para desenvolver a confiança dos usuários, ou seja, os usuários costumam confiar em repositórios com base em suas próprias experiências e práticas e a partir das experiências de outros membros da comunidade. Desta forma, conclui-se que os princípios TRUST são meios que facilitam a comunicação com todas as partes interessadas.

Na seção seguinte são apresentados os princípios FAIR.

4.1.4 Principles FAIR

Em 2016, a *Scientific Data* publicou os “Princípios orientadores FAIR para gestão de dados científicos (*FAIR Guiding Principles for Scientific Data Management and Stewardship*), que destina-se a fornecer diretrizes para melhorar a localização, acessibilidade, interoperabilidade e reuso de ativos digitais. Os princípios tem como objetivo enfatizar a ação da máquina, ou seja, a capacidade dos sistemas computacionais de encontrar, acessar,

interoperar e reutilizar dados com nenhuma ou mínima intervenção humana e referem-se a três tipos de entidades: dados (ou qualquer objeto digital); metadados (informações sobre aquele objeto digital); e infraestrutura (quadro 6).

Quadro 6 – Princípios Fair

Findable **(localizável):** F1. (Meta) dados são atribuídos a um identificador globalmente único e persistente F2. Os dados são descritos com metadados ricos F3. (Meta) dados incluem clara e explicitamente o identificador dos dados que descrevem F4. (Meta) dados são registrados ou indexados em um recurso pesquisável
Accessible **(Acessivel):** A1. (Meta) dados são recuperáveis por seu identificador usando um protocolo de comunicação padronizado - A1.1 O protocolo é aberto, gratuito e universalmente implementável -A1.2 O protocolo permite uma autenticação e autorização quando necessário A2. Os metadados devem estar acessíveis mesmo quando os dados não estão mais disponíveis
Interoperable **(Interoperável):** I1. (Meta) dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para representação de conhecimento I2. (Meta) dados usam vocabulários que seguem os princípios FAIR I3. (Meta) dados incluem referências qualificadas a outros (meta)dados
Reusable **(Reuso):** R1. (Meta) dados são ricamente descritos com pluralidade de atributos precisos e relevantes R1.1 (Meta) dados são liberados com uma licença de uso de dados clara e acessível R1.2 (Meta) dados estão associados à proveniência detalhada R1.3 (Meta) dados atendem aos padrões da comunidade relevantes para o domínio

Fonte: elaborado pela autora.

A seguir são descritos os princípios FAIR:

- ***Findable*** **(localizável):** o primeiro passo para (re)usar dados é encontrá-los. Os metadados e os dados devem ser fáceis de encontrar, tanto para humanos quanto para computadores.

Metadados legíveis por máquina são essenciais para a descoberta automática de conjunto de dados e serviços.

- F1. (Meta) dados são atribuídos a um identificador globalmente único e persistente: o princípio F1 é considerado o mais importante porque será difícil alcançar outros aspectos do FAIR sem identificadores únicos e persistentes, desta forma, a conformidade com F1 já o levará a aplicar dados FAIR. Identificadores únicos e persistentes removem ambiguidade no significado dos dados publicados, atribuindo um identificador único a cada elemento de metadados e a cada conceito/medição no conjunto de dados. Nesse contexto, os identificadores consistem em um link de internet, por exemplo, uma URL que leva a uma página da web que define o conceito. Os identificadores são essenciais para a interoperação homem-máquina que é a chave para a visão de Ciência Aberta. Além disso ajudam outras pessoas a citar corretamente o trabalho e reutilizar os dados. O princípio F1 estipula duas condições para o seu identificador: 1 - deve ser globalmente exclusivo, ou seja, outra pessoa não poderia reutilizar e pode-se obtê-los através de um serviço de registro que usa algoritmos que garantem a exclusividade de identificadores recém criados; 2 – deve ser persistente. Os links tendem a se tornar inválidos com o tempo, desta forma, os serviços de registros garantem a resolução desse link no futuro.
- F2. Os dados são descritos com metadados ricos: na criação de recursos digitais FAIR, os metadados devem ser extensos, incluindo informações descritivas sobre o contexto, qualidade ou características dos dados. Metadados ricos permitem que um computador realize automaticamente a classificação e priorização de tarefas rotineiras que atualmente exigem muita atenção dos pesquisadores. A lógica por trás desse princípio é que alguém deve ser capaz de encontrar dados com base nas informações fornecidas por seus metadados, mesmo sem o identificador de dados. Desta forma, a conformidade com F2 ajuda as pessoas a localizar seus dados e aumentar a reutilização e citações.

-F3. (Meta) dados incluem clara e explicitamente o identificador dos dados que descrevem: os metadados e o conjunto de dados geralmente são arquivos separados. A associação entre um arquivo de metadado e o conjunto de dados deve ser explicitada mencionando um identificador globalmente único e persistente do conjunto de dados nos metadados.

-F4. (Meta) dados são registrados ou indexados em um recurso pesquisável: identificadores e descrições ricas de metadados por si só não garantem a facilidade de localização na internet.

Recursos de dados podem ficar sem uso simplesmente porque ninguém sabe que eles existem, ou seja, se a disponibilidade de um recurso digital como, um conjunto de dados, serviço ou repositório não for conhecida, ninguém e nenhuma máquina poderá descobri-lo. Existem diversas maneiras de tornar os recursos digitais detectáveis, incluindo a indexação. Para os dados de pesquisa acadêmica é necessário ser mais explícito sobre a indexação, desta forma, os princípios F1-F3 fornecerão os elementos principais para a indexação de baixa granularidade por alguns repositórios atuais e serviços futuros (aqueles que precisam de um maior detalhamento).

- ***Accessible*** **(Acessível):** uma vez que o usuário encontra os dados necessários, precisa saber como eles podem ser acessados, possivelmente incluindo autenticação e autorização.

- A1. (Meta) dados são recuperáveis por seu identificador usando um protocolo de comunicação padronizado: a maioria dos usuários da internet recupera dados clicando em um link. Esta é uma interface de alto nível para um protocolo de baixo nível chamado ‘tcp’, que o computador executa para carregar dados no navegador da web do usuário. O princípio A1 afirma que a recuperação de dados FAIR deve ser mediada sem especialização ou ferramentas ou métodos de comunicação proprietários, ou seja, este princípio se concentra em como os dados e metadados podem ser recuperados de seus identificadores.

  Exemplo: - a maioria dos produtores de dados usará http(s) ou ftp;
  - as barreiras de acesso que devem ser evitadas incluem protocolos que têm implementações limitadas, documentação deficiente e componentes que envolvem intervenção humana manual. No entanto, observa-se que pode não ser possível fornecer acesso seguro por meio de um protocolo totalmente mecanizado, por exemplo, dados altamente confidenciais. Nesses casos, é perfeitamente justo fornecer um e-mail, número de telefone ou nome de uma pessoa de contato que possa discutir o acesso aos dados. Este protocolo de contato deve ser claro e explícito nos metadados.

  - A1.1 O protocolo é aberto, gratuito e universalmente implementável: para maximizar a reutilização de dados, o protocolo deve ser gratuito e aberto e implementável globalmente para facilitar a recuperação de dados. Qualquer pessoa com um

computador e uma conexão com a internet pode acessar pelo menos os metadados.

- A1.2 O protocolo permite uma autenticação e autorização quando necessário: dados altamente protegidos e privados podem ser FAIR e o ideal é que a acessibilidade seja especificada de forma que uma máquina possa compreender automaticamente os requisitos, e em seguida, executá-los automaticamente ou alertar o usuário sobre os requisitos. Frequentemente solicita-se aos usuários que criem uma conta de usuário para um repositório. Isso permite autenticar o proprietário de cada conjunto de dados e a definir direitos específicos do usuário.

- A2. Os metadados devem estar acessíveis mesmo quando os dados não estão mais disponíveis: os conjuntos de dados tendem a se degradar ou desaparecer com o tempo porque há um custo para manter uma presença online para recursos de dados. Quando isso acontece, os links se tornam inválidos e os usuários perdem tempo procurando dados que podem não estar mais lá. Armazenar metadados geralmente é muito mais fácil e barato. O princípio A2 afirma que os metadados devem persistir mesmo quando os dados não são mais sustentados.

Interoperable **(Interoperável):** os dados geralmente precisam ser integrados com outros dados. Além disso, os dados precisam interoperar com aplicativos ou fluxos de trabalho para análise, armazenamento e processamento.

I1. (Meta) dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para representação de conhecimento: os humanos devem ser capazes de trocar e interpretar os dados, no entanto, isso também se aplica a computadores, o que significa que os dados devem ser legíveis por máquinas sem a necessidade de algoritmos, tradutores ou mapeamentos especializados ou ad hoc. A interoperabilidade normalmente significa que cada sistema de computador tem conhecimento dos formatos de troca de dados do outro sistema. Para que isso aconteça e para garantir a localização automática e a interoperabilidade dos conjuntos de dados, é fundamental usar vocabulários controlados comumente usados, ontologias, tesauros tendo identificadores únicos e persistentes.

I2. (Meta) dados usam vocabulários que seguem os princípios FAIR: o vocabulário controlado usado para descrever conjuntos de dados precisa ser documentado e resolvido

usando identificadores globais únicos persistentes. Esta documentação deve ser facilmente encontrada e acessível por qualquer pessoa que use o conjunto de dados.

I3. (Meta) dados incluem referências qualificadas a outros (meta)dados: uma referência qualificada é uma referência cruzada que explica sua intenção. Ou seja, o objetivo é criar tantos links significativos quanto possível entre os recursos de (meta)dados para enriquecer o conhecimento contextual sobre os dados, equilibrado com o tempo envolvido na criação de um bom modelo de dados. As ligações científicas entre os conjuntos de dados precisam ser descritas.

Reusable **(Reuso):** o objetivo final do FAIR é otimizar a reutilização de dados. Para isto, os metadados e dados devem ser bem descritos para que possam ser replicados e combinados em diferentes configurações.

R1. (Meta) dados são ricamente descritos com pluralidade de atributos precisos e relevantes: esse princípio se concentra na capacidade de um usuário (máquina ou humano) de decidir se os dados são realmente úteis em determinado contexto. O editor de dados deve fornecer não apenas metadados que permitam a descoberta, mas também metadados que descrevem de forma rica o contexto em que esses dados foram gerados, podendo incluir os protocolos experimentais, o fabricante e a marca da máquina ou sensor que criou os dados, espécies usadas, etc. Deve-se descrever o escopo dos dados e para que finalidade eles foram gerados/coletados; deve-se especificar a data da coleta, as condições do laboratório, quem preparou, nome e versão do software utilizado; se são dados brutos ou processados; e documentar a versão dos dados arquivados.

- R1.1 (Meta) dados são liberados com uma licença de uso de dados clara e acessível: no princípio ‘I” são cobertos elementos de interoperabilidade técnica. R1.1 é sobre interoperabilidade legal. Equivale aos direitos que são atribuídos aos dados e isso deve ser descrito claramente. A ambiguidade pode limitar severamente a reutilização dos dados por organizações que lutam para cumprir as restrições de licenciamento. A clareza do status de licenciamento se tornará mais importante com pesquisas automatizadas envolvendo mais considerações de licenciamento. As condições sob as quais os dados podem ser usados devem ser claras para máquinas e humanos.

- R1.2 (Meta) dados estão associados à proveniência detalhada: para que outros possam reutilizar os dados, deve-se saber de onde os dados vieram, e quem citar. Deve-se incluir quem gerou os dados/coletou; como foi processado em um formato legível por máquina.
- R1.3 (Meta) dados atendem aos padrões da comunidade relevantes para o domínio: se houver padrões da comunidade ou melhores práticas para arquivamento e compartilhamento de dados, eles devem ser seguidos.

4.2 CONFRONTAÇÃO DOS CRITÉRIOS/REQUISITOS

O quadro 7 apresenta a confrontação dos instrumentos selecionados e foi elaborado para nortear o leitor quanto as subdivisões dos instrumentos. A coluna à esquerda apresenta uma correspondência das subdivisões por infraestruturas/áreas de cada um dos requisitos. As colunas à direita demonstram os instrumentos com seus respectivos requisitos (em comum). A seguir, no texto, foram descritas as semelhanças e diferenças.

Quadro 7 – Critérios/Requisitos dos instrumentos selecionados subdivididos por infraestruturas.

	Critério/ Requisito *CoreTrustSeal*	**Critério/ Requisito *ACTDR***	**Critério/ Requisito *The TRUST Principles***	**Critério/ Requisito *Principles FAIR***
Infraestrutura organizacional ***Transparency*** **(Transparência)**	0 Fornecer o contexto do repositório 1. Missão/escopo 2. Licenças 3. Continuidade de acesso 4. Confidencialidade/Ética 5. Infraestrutura organizacional 6. Orientação de especialista	a) governança e viabilidade organizacional b) estrutura organizacional e pessoal c) responsabilidade processual e política de preservação d) sustentabilidade financeira e) contratos, licenças e responsabilidades	a) transparência sobre seus serviços b) fornece informações claras sobre seu escopo, sua missão, sua comunidade alvo e suas políticas c) fornece termos de uso, tanto para o repositório quanto para os acervos de dados d) estabelece um prazo mínimo de preservação digital para os acervos de dados e) demonstra capacidade de administrar com responsabilidade os dados confidenciais	<u>*Reusable* (Reuso):</u> - R1.1 (Meta) dados são liberados com uma licença de uso de dados clara e acessível

	Critério/ Requisito *CoreTrustSeal*	Critério/ Requisito *ACTDR*	Critério/ Requisito *The TRUST Principles*	Critério/ Requisito *Principles FAIR*
Gerenciamento (gestão) de objetos digitais ***Responsibility*** **(Responsabilidade)**	7 Integridade e autenticidade dos dados 8 Avaliação 9 Procedimentos de armazenamento documentados 10 Plano de preservação 11 Qualidade dos dados 12 Fluxos de trabalho 13 Descoberta e identificação de dados 14 Reutilização de dados	a) *Ingest* (ingestão de dados): aquisição de conteúdo b) *Ingest* (ingestão de dados): criação do *Archival Information Package* (AIP) – Pacote de informação para arquivamento c) planejamento de preservação d) preservação de AIP e) gestão de informações f) gestão de acesso	a) o repositório adere os padrões de metadados e curadoria, juntamente com o gerenciamento dos acervos de dados, por exemplo, validação técnica, documentação e controle de qualidade b) fornece serviços de dados, por exemplo, download de dados c) gerencia os direitos de propriedade intelectual dos produtores de dados, proteção de recursos de informações confidenciais, e a segurança do sistema e seu conteúdo	<u>*Findable* (localizável):</u> F1. (Meta) dados são atribuídos a um identificador globalmente único e persistente F2. Os dados são descritos com metadados ricos F3. (Meta) dados incluem clara e explicitamente o identificador dos dados que descrevem F4. (Meta) dados são registrados ou indexados em um recurso pesquisável <u>*Accessible* (Acessível):</u> A2. Os metadados devem estar acessíveis mesmo quando os dados não estão mais disponíveis <u>*Interoperable* (Interoperável):</u> I3. (Meta) dados incluem referências qualificadas a outros (meta)dados <u>*Reusable* (Reuso):</u> R1. (Meta) dados são ricamente descritos com pluralidade de atributos precisos e relevantes ------ R1.2 (Meta) dados estão associados à proveniência detalhada.

	Critério/ Requisito *CoreTrustSeal*	Critério/ Requisito *ACTDR*	Critério/ Requisito *The TRUST Principles*	Critério/ Requisito *Principles FAIR*
***User focus* (Foco no usuário)**			a) O repositório garante que as normas e expectativas de gestão de dados da comunidade-alvo sejam atendidas b) disponibiliza métricas de dados relevantes e disponibiliza aos usuários c) fornece catálogos para facilitar a descoberta de dados d) monitora e identifica as expectativas em relação a comunidade e responde conforme necessário	

	Critério/ Requisito *CoreTrustSeal*	**Critério/ Requisito *ACTDR***	**Critério/ Requisito *The TRUST Principles***	**Critério/ Requisito *Principles FAIR***
Tecnologia **Gestão de risco de infraestrutura e segurança** ***Sustainability* (Sustentabilidade)/*Technology***	15 Infraestrutura técnica 16 Segurança	a) gestão de risco de infraestrutura técnica b) gestão de risco de segurança	a) o repositório possui planejamento suficiente para mitigação de riscos, recuperação de desastres e sucessão b) garante o financiamento para permitir o uso contínuo e mantem as propriedades desejáveis dos recursos de dados que o repositório foi encarregado de preservar e divulgar c) fornece governança para a preservação necessária de dados de longo prazo para que os recursos de dados permaneçam detectáveis, acessíveis e utilizáveis no futuro d) implementa padrões, ferramentas e tecnologias relevantes e apropriadas para gerenciamento e curadoria de dados e) tem planos e mecanismos em vigor para prevenir, detectar e responder à ameaças à segurança física ou cibernética	<u>*Accessible* (Acessível):</u> A1. (Meta) dados são recuperáveis por seu identificador usando um protocolo de comunicação padronizado - A1.1 O protocolo é aberto, gratuito e universalmente implementável -A1.2 O protocolo permite uma autenticação e autorização quando necessário <u>*Interoperable* (Interoperável):</u> I1. (Meta) dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para representação de conhecimento I2. (Meta) dados usam vocabulários que seguem os princípios FAIR <u>*Reusable* (Reuso):</u> R1.3 (Meta) dados atendem aos padrões da comunidade relevantes para o domínio

Fonte: elaborado pela autora.

Mediante ao grau de dificuldade de realizar este trabalho de correspondência, ou seja, de realizar a análise de quatro documentos de natureza distinta, houve a necessidade de discorrer sobre suas diferenças.

Os princípios FAIR, já mencionados, destacam a necessidade de adotar boas práticas definindo características essenciais de objetos de dados para a garantia que sejam reutilizáveis por humanos e máquinas. Para sustentar a implantação dos princípios FAIR, é necessário definir os elementos centrais dos objetos de dados envolvidos e, a partir dessa definição, desenvolver um ecossistema que inclua os serviços necessários para criar, gerir e partilhar os objetos de maneira FAIR (HODSON *et al*., 2018, p.78). Percebe-se, também, que o ecossistema de dados FAIR é altamente distribuído, sendo os repositórios digitais confiáveis responsáveis pelos armazenamento, gerenciamento e preservação dos dados. (HODSON *et al*., 2018, p.39). A proposta de um Ecossistema de dados FAIR compreende, no mínimo, cinco componentes essenciais: políticas, planos de gestão de dados, identificadores, padrões e repositórios (HODSON *et al*., 2018, p.21).

No entanto, para tornar esses dados justos e preservá-los a longo tempo, é necessário ter Repositórios Digitais Confiáveis – *Trustworthy Digital Repositories* (TDRs), com governança sustentável e estruturas organizacionais, infraestrutura confiável e políticas abrangentes de apoio às práticas acordadas pela comunidade. Os TDRs devem demonstrar capacidades essenciais e douradoras para permitir o acesso e reutilização de dados ao longo do tempo para as comunidades que atendem (DAWEI LIN *et al.*, 2020).

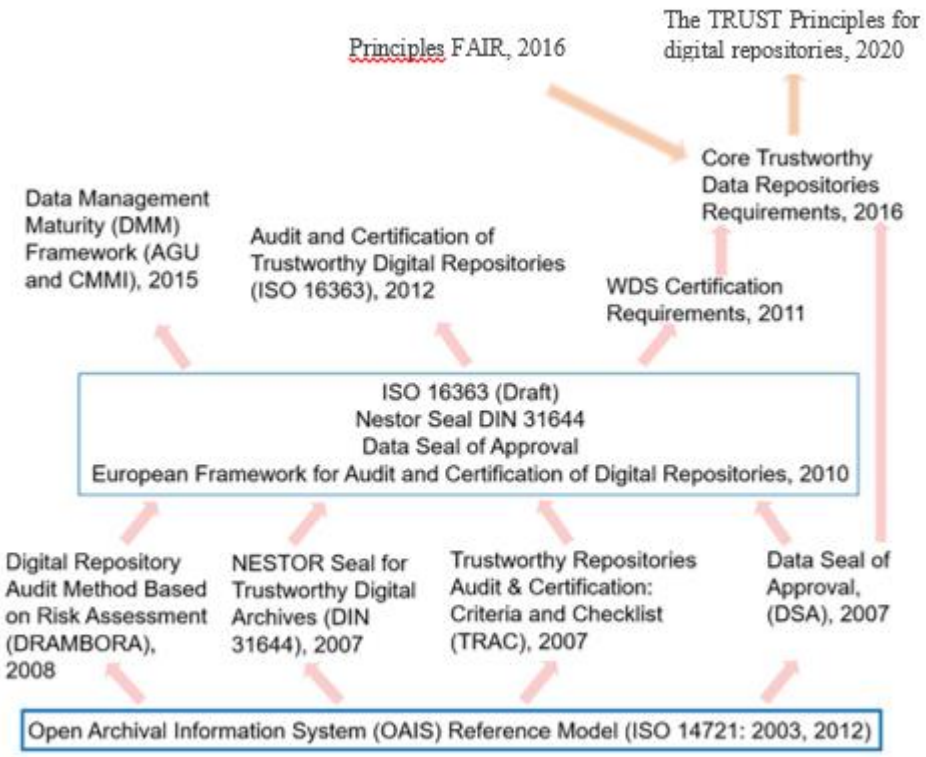
Modelos como OAIS fornece recomendações referente a preservação a longo prazo, princípios e terminologia para a gestão de sistemas de informação, no entanto, para avaliar a confiabilidade, elementos adicionais precisam ser tratados, incluindo governança, recursos e segurança apropriados. Conforme Dawei Lin *et al.* (2020), a confiabilidade é demonstrada por meio de evidências que dependem de transparência, desta forma, os repositórios devem fornecer evidências transparentes, honestas e verificáveis de sua prática. Somente desta forma as partes interessadas poderão estar confiantes que os repositórios garantem a integridade, autenticidade, precisão, confiabilidade e acessibilidade dos dados por longo prazo.

Percebe-se que os requisitos do *CoreTrustSeal* possui uma terminologia baseada no modelo OAIS e os princípios FAIR estão implícitos nos requisitos. O *CoreTrustSeal* também serve para garantir que os repositórios atinjam as propriedades dos Princípios TRUST, os

quais constituem outro conjunto de orientações que objetivam assegurar confiabilidade aos repositórios de dados digitais.

Baseada nessas relações elaborou-se uma figura (2) adaptada da publicação de Downs (2019) onde foram inseridos entre os instrumentos, os princípios FAIR e TRUST.

Figura 2 – Esquema com inserção dos Princípios FAIR e Princípios TRUST.



Fonte: elaborado pela autora e adaptado de Downs (2019).

Nas subseções seguintes foram realizadas as análises para a identificação de semelhanças e diferenças e resolveu-se organizá-las por infraestruturas/áreas.

- Infraestrutura organizacional/*Transparency* (Transparência)

O *CoreTrustSeal*, em sua parte introdutória indica a necessidade que o repositório possui em fornecer seu contexto, ou seja, há a necessidade de relatar o tipo de repositório, item que ajudará os revisores a entender qual a função o repositório executa. Esta explicação

pode fazer referência a coleções de dados relevantes, tipos de dados, formatos e disciplinas com as quais o repositório trabalha: repositório institucional, repositório nacional (governamental), repositório de publicação, biblioteca, museu, arquivo, repositório de projetos de pesquisa, ou outro. Além disso, deve-se fornecer uma breve descrição do repositório, como também uma breve descrição da comunidade designada - a definição clara da comunidade demonstra que o candidato a certificação compreende o escopo e as metodologias - formatos preferidos - da comunidade de usuários que possuem como alvo. Para servir bem a comunidade designada, o repositório deve ter um conhecimento profundo da composição, habilidades e necessidades. O nível de curadoria também deve ser fornecido - este item tem como objetivo verificar se o repositório distribui seu conteúdo para consumidores de dados sem nenhuma alteração mantendo a integridade dos dados (CORETRUSTSEAL, 2020b).

No instrumento ACTDR consta a importância de testar se o repositório atende às necessidades de sua comunidade designada, ou seja, deve-se atender ao requisito ‘o repositório deve ter definido sua comunidade designada e base associada de conhecimento e deve ter essas definições apropriadamente acessíveis’. Uma forma de demonstrar que está atendendo a esse requisito é através de uma definição escrita da comunidade designada. A comunidade designada é definida como "um grupo identificado de consumidores potenciais que deve ser capaz de compreender um determinado conjunto de informações sendo composta de várias comunidades de usuários. Uma comunidade designada é definida pelo arquivo e esta definição pode mudar/evoluir ao longo do tempo” (*CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011).

Os princípios TRUST, no item ‘*User Focus* – foco no usuário’ trata sobre a necessidade de monitorar e identificar as expectativas em relação a comunidade (DAWEI LIN *et al.*, 2020).

O primeiro requisito do *CoreTrustSeal* abrange a missão e o escopo. Percebe-se que o instrumento ACTDR também possui essa subdivisão e, dentro da infraestrutura organizacional, trata sobre a governança e a viabilidade organizacional. Ambos abordam a declaração da missão como requisito e garantia da preservação. Da mesma forma, os princípios TRUST apresentam em seu primeiro requisito ‘*transparency*’ a mesma abordagem.

Estes requisitos semelhantes relatam que o repositório é responsável pela administração dos objetos digitais e por garantir que os materiais sejam mantidos no ambiente apropriado por períodos de tempo. A preservação e o acesso contínuo aos dados são funções explícitas do repositório. O seja, é necessário descrever a missão da organização em preservar e fornecer acesso aos dados. Portanto, deve-se publicar a declaração de missão ou estatuto do repositório ou de sua organização onde deve abordar explicitamente a preservação. A fim de selecionar o repositório mais apropriado para o uso, usuários potenciais se beneficiam ao encontrar e acessar facilmente informações sobre o escopo, comunidade de usuários, políticas e recursos do repositório de dados (CORETRUSTSEAL, 2020b).

A transparência nessas áreas oferece uma oportunidade de aprender sobre o repositório e considerar sua adequação aos requisitos específicos dos usuários, incluindo depósito, preservação e descoberta de dados. Em conformidade com este princípio, os repositórios devem garantir que, no mínimo, a declaração de missão e o escopo do repositório sejam claramente indicados. Além disso devem estar declarados de forma transparente os seguintes aspectos: termos de uso, tanto para o repositório quanto para os acervos de dados; prazo mínimo de preservação digital para os acervos de dados; quaisquer recursos ou serviços adicionais pertinentes, por exemplo, a capacidade de administrar com responsabilidade dados confidenciais (DAWEI LIN *et al.*, 2020).

O segundo requisito do CoreTrustSeal refere-se as ‘licenças’, equivalente ao requisito ‘contratos, licenças e responsabilidades’ do instrumento ACTDR. Ambos requisitos assemelham-se ao requisito ‘*Responsibility*’ dos princípios TRUST. O Princípio FAIR R1.1 também possui relação com estes requisitos.

Conforme os instrumentos citados (CORETRUSTSEAL, 2020b; *CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011; DAWEI LIN *et al.*, 2020; WILKINSON *et al*., 2016), este requisito refere-se aos regulamentos de acesso e licenças aplicáveis estabelecidas pelo próprio repositório de dados, bem como quaisquer códigos de conduta que são geralmente aceitos no setor relevante para o intercâmbio e uso adequado de conhecimento e informação. Deve-se descrever os contratos de licenças, condições de uso (direitos de propriedade intelectual, uso pretendido, proteção de dados confidenciais, etc). Deve-se considerar as consequências se o não cumprimento for detectado, no caso de divulgação de dados pessoais confidenciais, pode haver penalidades legais severas que afetam tanto o

usuário quanto o repositório, ou seja, os repositórios devem ter uma política pública em vigor para o descumprimento. Ou seja, este requisito é importante para garantir que o repositório tenha os direitos e autorizações necessários para permitir que ele colete e preserve conteúdo digital ao longo do tempo, faça com que essas informações estejam disponíveis para sua comunidade designada, e para defender esses direitos quando questionados.

Deve-se ter contratos de depósito e licenças devidamente assinados e executados de acordo com as normas locais, leis e regulamentos nacionais e internacionais; políticas sobre acordos de depósito de terceiros; definições de níveis de serviço e usos permitidos; políticas de repositório sobre o tratamento de 'Obras órfãs' e resolução de disputas de direitos autorais; relatórios de avaliações de risco; procedimentos para revisar e manter regularmente acordos, contratos e licenças. Os repositórios podem precisar mostrar evidências de que seus contratos estão sendo cumpridos. Contratos e acordos de depósito formais devem ser legítimos, ou seja, eles precisam ser assinados e atuais. Os repositórios confiáveis assumem a responsabilidade pela administração de seus acervos de dados e por servir seus usuários.

Os usuários do repositório devem ter certeza de que os depositantes de dados são solicitados a fornecer todos os metadados compatíveis com as normas da comunidade, o que aumenta a descoberta e a utilização dos dados. Tanto os depositantes quanto os usuários devem ter confiança de que os dados permanecerão acessíveis ao longo do tempo, portanto poderão ser citados e referenciados em publicações acadêmicas. A responsabilidade pode ser esclarecida através de alguns meios legais (direito de preservar) ou pode assumir a forma de cumprimento de alguma norma (padrões éticos).

O Princípio FAIR R1.1 menciona que os (Meta) dados devem ser disponibilizados com licenças de uso claras e acessíveis, ou seja, é fundamental que o responsável pelos dados e metadados defina explicitamente quem pode ter acesso a eles, para que e sob quais condições.

O terceiro requisito do *CoreTrustSeal*, 'continuidade de acesso', se assemelha ao requisito do ACTDR 'governança e a viabilidade organizacional' destacando que o repositório deve ter um Plano Estratégico de Preservação que define a abordagem que o repositório terá no suporte de longo prazo. Os princípios TRUST tratam sobre esta questão através do princípio '*Sustainability'.*

O requisito do *CoreTrustSeal* (CORETRUSTSEAL, 2020b) cobre a governança relacionada à operação contínua do repositório ao longo do tempo e durante desastres, bem como evidências em relação ao planejamento de sucessão, designadamente, as medidas em vigor para garantir o acesso e a disponibilidade dos acervos de dados, atualmente e no futuro. Para este requisito deve-se descrever: o nível de responsabilidade assumido por acervos de dados, incluindo quaisquer períodos de preservação garantidos; planos de médio prazo (três a cinco anos) e longo prazo (cinco anos) em vigor para garantir a disponibilidade e acessibilidade contínua dos dados.

Neste requisito é necessário descrever o nível de responsabilidade assumido pelos dados e o nível de risco para a organização atual, ou seja, é importante descrever se o depositante compartilha a responsabilidade pelo futuro dos dados, se o repositório fornece acesso, preservação e/ou armazenamento de dados a algum nível mínimo de qualidade por um período mínimo de tempo. Essas informações possibilitam o julgamento em relação a sustentabilidade do repositório em termos de finanças e processos, que pode não estar sob responsabilidade do repositório em si, e sim por um host superior ou uma organização principal. Sendo assim, se faz necessário um acordo formal para garantir quem assumirá a responsabilidade em caso de descontinuidade do serviço.

O ACTDR menciona que é através de um Plano Estratégico de Preservação; atas de reuniões; e documentação de decisões administrativas que pode-se atingir esse requisito. O plano estratégico deve ser baseado na missão estabelecida da organização, e em seu valores, visão e objetivos definidos. O repositório deve ter um plano de sucessão adequado, planos de contingência, e/ou acordos de custódia em vigor no caso de o repositório deixar de operar ou a instituição governamental ou de financiamento alterar substancialmente seu escopo.

Deve-se utilizar a Sucessão escrita e plano(s) de contingência; declaração explícita e específica com a intenção de garantir a continuidade do repositório; custódia de código crítico, software e metadados suficientes para permitir a reconstituição do repositório e seu conteúdo em caso de falha do repositório; fundos de *escrow* e/ou reserva para contingências; acordos explícitos com o sucessor documentando as medidas a serem tomadas para garantir a completa e formal transferência de responsabilidade pelo conteúdo digital do repositório e ativos relacionados, e concessão dos direitos necessários para garantir a continuidade dos serviços de conteúdo e repositório.

De acordo com o ACTDR, o repositório deve monitorar seu ambiente organizacional para determinar quando executar seu plano de sucessão, planos de contingência e/ou acordos de custódia através de políticas administrativas, procedimentos, protocolos, requisitos; orçamentos e análise financeira; calendários fiscais; Planos de negócio; qualquer evidência de monitoramento ativo e preparação. A gestão de um repositório deve ter procedimentos formais para periodicamente verificar a viabilidade do repositório. Esta verificação periódica deve ser usada para determinar se, ou quando, executar o plano de sucessão formal do repositório, planos de contingência e/ou arranjos de custódia (*CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011).

Conforme os princípios TRUST, um repositório confiável pode demonstrar sustentabilidade de seus acervos através de planejamento suficiente para mitigação de riscos, recuperação de desastres e sucessão; garantia de financiamento para permitir o uso contínuo e manter as propriedades desejáveis dos recursos de dados que o repositório foi encarregado de preservar e divulgar; fornecimento de governança para a preservação necessária de dados de longo prazo para que os recursos de dados permaneçam detectáveis, acessíveis e utilizáveis no futuro (DAWEI LIN *et al.*, 2020).

O quarto requisito do *CoreTrustSeal*, 'Confidencialidade/Ética', menciona que o repositório deve garantir, na medida do possível, que os dados sejam criados, com curadoria, acessados e usados em conformidade com as normas disciplinares e éticas. Da mesma forma, o requisito 'sustentabilidade financeira" do ACTDR, trata em parte o assunto confidencialidade. Os princípios TRUST, no item '*Responsibility'* aborda as informações confidenciais.

O instrumento *CoreTrustSeal* relata que este requisito se refere às disposições éticas e de privacidade que afetam a criação, curadoria e uso dos dados. Desta forma, o repositório deve demonstrar que tem boas práticas para dados que possuem riscos de divulgação e deve incluir orientações para os depositantes e usuários. Esse procedimento é necessário para manter a confiança daqueles que concordam em ter dados pessoais/confidenciais armazenados no repositório.

Para este requisito deve-se responder como o repositório cumpre as normas disciplinares aplicáveis, como o repositório solicita confirmação de que a coleta ou criação de dados foi realizada de acordo com os critérios legais e éticos, como os dados com risco de

divulgação são gerenciados de maneira adequada para limitar acesso, se os funcionários são treinados em relação ao gerenciamento de dados com risco de divulgação, se existem medidas em vigor caso as condições não forem cumpridas. Ou seja, espera-se que as organizações responsáveis pelos dados possuam o dever ético de gerenciá-los no nível esperado pela prática científica da comunidade designada.

Para repositórios que contêm dados sobre indivíduos, organizações ou áreas e espécies protegidas, existem expectativas legais e éticas adicionais de que os direitos dos titulares dos dados serão protegidos. Ou seja, a divulgação desses dados pode representar um risco de dano pessoal, uma violação da confidencialidade comercial ou a divulgação de informações críticas. Se houver algum risco de que esses dados sejam depositados, por exemplo, por acidente, o repositório deve tomar as medidas adequadas para lidar com esses dados e que eles sejam tratados de acordo com os regulamentos legais. O repositório deve apresentar evidências de que possui os procedimentos documentados em vigor para garantir a conformidade (CORETRUSTSEAL, 2020b).

Contudo, o ACTDR menciona que é necessário se proteger contra má conduta ou outra atividade desagradável que possa ameaçar a viabilidade econômica do repositório. O repositório não pode apenas reivindicar transparência, mas deve mostrar que ajusta seus negócios para mantê-los transparentes, compatíveis e auditáveis. Requisitos de confidencialidade podem proibir a divulgação de informações sobre as finanças do repositório, mas o repositório deve ser capaz de demonstrar que está satisfazendo as necessidades de sua comunidade designada (*CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011).

Os princípios TRUST mencionam que a responsabilidade é demonstrada com o gerenciamento dos direitos de propriedade intelectual dos produtores de dados, proteção de recursos de informações confidenciais, e a segurança do sistema e seu conteúdo. Os usuários do repositório devem ter certeza de que os depositantes de dados são solicitados a fornecer todos os metadados compatíveis com as normas da comunidade, o que aumenta a descoberta e a utilização dos dados. Tanto os depositantes quanto os usuários devem ter confiança de que os dados permanecerão acessíveis ao longo do tempo, portanto poderão ser citados e referenciados em publicações acadêmicas. A responsabilidade pode ser esclarecida através de

alguns meios legais/direito de preservar ou pode assumir a forma de cumprimento de alguma norma/padrões éticos (DAWEI LIN *et al.*).

O quinto requisito do *CoreTrustSeal*, 'Infraestrutura organizacional', menciona que o repositório deve ter financiamento adequado e número suficiente de funcionários qualificados gerenciados por meio de um sistema claro de governança para realizar a missão com eficácia. O ACTDR possui um requisito semelhante, 'estrutura organizacional e pessoal', e relata que o repositório deve ter identificado e estabelecido as funções de que necessita para executar e deve ter nomeado pessoal com habilidades e experiência adequadas para cumprir essas funções. Os princípios TRUST relatam que um repositório depende da interação das pessoas, processos e tecnologias para oferecer suporte seguro, persistente e serviços confiáveis (CORETRUSTSEAL, 2020b; *CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011; DAWEI LIN *et al.*, 2020).

O sexto requisito do *CoreTrustSeal*, 'Orientação de especialista' assemelhasse a um dos requisitos do ACTDR, 'contratos, licenças e responsabilidades'. Segundo o instrumento do *CoreTrustSeal*, um repositório eficaz se esforça para realizar evoluções e adotar novas tecnologias mais eficazes a fim de permanecer valioso para sua comunidade designada.

Devido ao ritmo rápido de mudanças é aconselhável que o repositório possua um aconselhamento e feedback de especialistas para garantir sua relevância e melhorias contínuas. Para este requisito deve-se responder se o repositório possui consultores internos ou um comitê consultivo externo que podem incluir especialistas técnicos, de curadoria e ciência dos dados; deve-se responder como o repositório se comunica com os especialistas para aconselhamento; e como o repositório se comunica com a comunidade designada para feedback.

O repositório deve dar evidências de que está ligado a uma rede mais ampla de experiência, a fim de demonstrar acesso a aconselhamento e orientação para suas atividades e monitoramento de novos desafios potenciais. Da mesma forma, o ACTDR orienta que as Políticas de Preservação e Planos de Implementação de Preservação do repositório e mecanismos devem ser examinados por autoridades institucionais apropriadas e/ou especialistas jurídicos para garantir que as respostas aos desafios cumpram as leis e requisitos relevantes (CORETRUSTSEAL, 2020b; *CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011). Os princípios TRUST não mencionam este requisito.

- Gerenciamento (gestão) de objetos digitais/*Responsibility* (Responsabilidade)

O sétimo requisito do *CoreTrustSeal* corresponde a ‘Integridade e autenticidade dos dados’. O instrumento ACTDR possui alguns requisitos que assemelham-se ao do *CoreTrustSeal*: responsabilidade processual e política de preservação (o repositório deve definir, coletar, rastrear e fornecer adequadamente suas medições de integridade da informação; o repositório deve ter um processo de *ingest* que verifica cada SIP para integridade e correção; o repositório deve obter controle suficiente sobre os objetos digitais para preservá-los; o repositório deve verificar cada AIP quanto à integridade e exatidão; e o repositório deve fornecer um mecanismo independente para verificar a integridade da coleção /conteúdo do repositório).

O requisito preservação de AIP também trata sobre integridade: (o repositório deve ter especificações de como os AIPs são armazenados até o nível de bits). No requisito ‘gestão de informações’ também há especificações referentes a integridade. Na subdivisão: ‘Gestão de risco de infraestrutura e segurança’, requisito ‘gestão de risco de infraestrutura técnica’ do ACTDR, informa que o repositório deve identificar e gerenciar os riscos às suas operações de preservação e objetivos associados à infraestrutura do sistema como também, gerenciar o número e localização de cópias de todos os objetos digitais. O Princípio FAIR R1.2 possui relação com estes requisitos. Nos princípios TRUST não foram encontrados requisitos relacionados a integridade e autenticidade.

De acordo com o requisito do *CoreTrustSeal*, o repositório deve fornecer evidências para mostrar que opera um sistema de gerenciamento de dados e metadados adequado para garantir integridade e autenticidade durante processos de armazenamento e acesso a dados. Este requisito cobre todo o ciclo de vida dos dados dentro do repositório. Para proteger a integridade dos dados e metadados, quaisquer alterações intencionais devem ser documentadas, incluindo a justificativa e o originador da mudança. Convém que medidas sejam implementadas para garantir que mudanças não intencionais ou não autorizadas possam ser detectadas e versões corretas de dados e metadados recuperadas.

A autenticidade cobre o grau de confiabilidade dos dados originais depositados e sua procedência, incluindo a relação entre os dados originais e os divulgados e se as relações

existentes entre os conjuntos de dados e/ou metadados são mantidas ou não. Para este requisito deve-se descrever a documentação da integridade dos dados e metadados; descrever em detalhes como todas as alterações nos dados e metadados são registradas; a descrição estratégica de controle de versão; deve-se utilizar o uso de padrões em convenções internacionais apropriados. Para incluir evidências sobre gerenciamento de autenticidade deve-se responder se o repositório tem uma estratégia para mudanças de dados; se mantém links para metadados e outros conjuntos de dados; e se é verificado as identidades dos depositantes (CORETRUSTSEAL, 2020b).

Os Princípios FAIR trata sobre o reuso dos dados, ou seja, pode-se avaliar se os dados são precisos e descritos com atributos relevantes; se os dados possuem uma licença de uso clara e acessível, se está claro como, por que e por quem os dados foram criados e processados; e se os dados e os metadados atendem a padrões e domínio relevantes (WILKINSON *et al*., 2016).

O ACTDR menciona que as maneiras pelas quais o repositório pode demonstrar que está atendendo a esse requisito será a partir da definição ou especificação escrita das medidas de integridade do repositório; documentação dos procedimentos e mecanismos para monitorar medições de integridade e responder aos resultados das medições de integridade que indicam que o conteúdo digital está em risco; um processo de auditoria para coletar, rastrear e apresentar medições de integridade; política de preservação e documentação do fluxo de trabalho.

É responsabilidade do repositório escolher o mecanismo apropriado para verificar a integridade e exatidão de suas coleções. Em geral, é provável que um repositório que atende a todos os critérios anteriores irá satisfazer este sem a necessidade de demonstrar nada mais.

Como um requisito separado, demonstra a importância de ser capaz de auditar a integridade da coleção como um todo. Um repositório deve ser capaz de mostrar, para cada item de seu cadastro de acessos, quais AIP(s) possuem conteúdo desse item. Como alternativa, pode ser necessário mostrar que não há AIP para um item, também porque a *ingest* ainda está em andamento ou porque o item foi rejeitado por algum motivo. Por outro lado, qualquer AIP deve ser capaz de ser relacionado a uma entrada no registro de aquisições. Desta forma, o repositório deve monitorar ativamente a integridade dos AIPs.

Um dos requisitos do ACTDR relacionado a integridade diz respeito ao processo para testar e avaliar o efeito de mudanças nos processos críticos do repositório (é necessário para proteger a integridade dos processos críticos do repositório, como que eles continuam em sua capacidade de atender aos requisitos obrigatórios do repositório). O repositório pode demonstrar que está atendendo a esse requisito através de procedimentos de teste documentados; documentação de resultados de testes anteriores e prova de alterações feitas como resultado de testes; e análise do impacto de uma mudança de processo. Mudanças em sistemas críticos devem ser, sempre que possível, pré-testados separadamente. Após as mudanças, os sistemas devem ser monitorados quanto ao comportamento inesperado e se tal comportamento for descoberto, as mudanças e suas consequências devem ser revertidas.

Para garantir que o repositório forneça uma cópia autêntica de um objeto digital específico devem ser realizados testes de recuperação aleatórios; validação da existência do objeto para cada local registrado; validação de uma localização registrada para cada objeto em sistemas de armazenamento; verificação de proveniência e fixidez em formação; registro de localização de objetos digitais em comparação com o número esperado e localização de cópias de objetos específicos. Um repositório pode ter diferentes políticas de preservação para diferentes classes de objetos, dependendo de fatores como o produtor, o tipo de informação ou seu valor.

Repositórios podem exigir um número diferente de cópias para cada classe, ou gerenciar as versões necessárias para atender requisitos de acesso. Pode haver requisitos de identificação adicionais se os mecanismos de integridade dos dados usam cópias alternativas para substituir cópias com falha. A localização de cada objeto digital deve ser descrito de modo que possa ser localizado com precisão, sem ambiguidade. O local pode ser físico absoluto ou um local lógico em uma mídia de armazenamento ou um subsistema de armazenamento.

As informações de proveniência sobre como copiar e mover os dados devem ser mantidos/atualizados, incluindo a identificação dos responsáveis. Isso é necessário a fim de rastrear a cadeia de custódia e afirmar que o repositório está fornecendo uma cópia autêntica de um determinado objeto digital. O repositório deve ser capaz de distinguir entre as versões de objetos ou cópias idênticas. Isso é necessário para que um repositório possa afirmar que

está fornecendo uma cópia autêntica da versão correta de um objeto (*CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011).

O oitavo requisito do *CoreTrustSeal*, 'Avaliação', orienta o repositório a aceitar dados e metadados com base em critérios definidos para garantir relevância e compreensibilidade para os usuários de dados. O ACTDR possui alguns requisitos relacionados à avaliação: 'O repositório deve se comprometer com um cronograma regular de autoavaliação e certificação externa'; 'O repositório deve fornecer evidências da eficácia de sua preservação'; 'O repositório deve manter uma análise sistemática dos fatores de risco de segurança associados a dados, sistemas, pessoal e planta física'; 'O repositório deve ter uma Política de Coleta ou outro documento que especifique o tipo de informação que irá preservar, reter, gerenciar e fornecer acesso. Após análise dos princípios TRUST percebe-se que não há abordagem deste requisito.

De acordo com o *CoreTrustSeal* (CORETRUSTSEAL, 2020b), a função da avaliação é fundamental para avaliar se os dados atendem a todos os critérios de seleção e para garantir o manejo adequado para a sua preservação. A avaliação e a reavaliação ao longo do tempo garantem que os dados permaneçam relevantes e compreensíveis para a comunidade designada. Para este requisito deve-se incluir evidências como, por exemplo, se o repositório possui uma política de desenvolvimento de coleção para orientar a seleção de dados; evidências sobre qual abordagem é usada para dados que não se enquadram no perfil de missão/coleta; se o repositório possui procedimentos em vigor para determinar se são fornecidos os metadados necessários para interpretar e usar os dados; se existe alguma avaliação automatizada da aderência dos metadados aos esquemas relevantes; qual é a abordagem do repositório e se os metadados fornecidos são insuficientes para a preservação a longo prazo; se o repositório publica uma lista de formatos preferidos; se existem verificações para garantir que os produtores de dados sigam formatos preferidos; e qual é o processo de remoção de itens de sua coleção tendo em mente o impacto nos identificadores persistentes. Os Princípios FAIR tratam sobre a encontrabilidade dos dados, ou seja, deve-se verificar se são encontráveis. Para realizar a avaliação é necessário averiguar se os dados estão associados a um identificador persistente e se existem metadados ricos descrevendo os dados.

Para que a coleção permaneça relevante e utilizável pela comunidade designada, especialmente à luz das mudanças na tecnologia, cultura e legislação (como por exemplo,

proteção de dados ou direitos de propriedade intelectual), os critérios de seleção podem ter que ser revisados ao longo do tempo e os ativos digitais reavaliados adequadamente.

Contudo, o ACTDR (*CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011) menciona que o repositório deve ser capaz de demonstrar a preservação contínua, incluindo compreensibilidade de suas participações. Isso pode ser avaliado em vários graus e depende da especificidade da comunidade designada. Se uma comunidade designada for razoavelmente ampla, um auditor pode representar o sujeito na avaliação e garantir que o repositório continue a ser confiável e que não haja ameaça ao seu conteúdo. Através de listas de verificação preenchidas e datadas de autoavaliações e/ou auditorias de terceiros; certificados concedidos em conformidade com os padrões ISO; cronogramas e evidências de alocações de orçamento para certificação futura é possível atender a esse requisito. Uma verificação única de confiabilidade não é adequada porque muitas coisas vão mudar com o tempo.

Um compromisso de longo prazo deve ser demonstrado, ou seja, o repositório deve ter procedimentos em vigor para avaliar quando as mudanças são necessárias para o hardware atual (necessário para garantir que o repositório tenha a capacidade de tornar informado e oportuno decisões quando as informações indicam a necessidade de novo hardware). Portanto, por meio de procedimentos de avaliação implementados e experiência documentada da equipe em cada subsistema de tecnologia é possível verificar se o repositório está atendendo ao requisito.

O repositório deve ter procedimentos, compromisso e financiamento para substituir o hardware quando a avaliação indica a necessidade de fazê-lo (isso é necessário para garantir a substituição do hardware em tempo hábil, a fim de evitar a falha no sistema ou inadequação de desempenho). Sem esse compromisso, e mais importante, sem recursos financeiros garantidos ou um fluxo de financiamento seguro, as notificações têm pouco valor. O repositório deve ter mecanismos para avaliar a eficácia dos novos sistemas antes da implementação no sistema de produção. O objetivo é demonstrar que o repositório tem a capacidade de incorporar novas tecnologias, tanto financeiramente por meio de compromissos de financiamento ou redução de custos, e operacionalmente através da verificação das capacidades dos novos sistemas.

O repositório deve realizar avaliações de risco regulares e manter a segurança adequada a fim de fornecer os níveis de serviço previstos e contratados. 'Sistema' aqui se refere a mais do que sistemas de TI, como hardware, software, comunicações equipamentos e instalações e firewalls. Sistemas de proteção contra incêndio e detecção de inundação também são significativos, assim como os meios para avaliar os procedimentos de pessoal, gestão e administração, recursos, bem como operações e prestação de serviços. Perda de receita, orçamento e reputação são ameaças significativas às operações gerais, assim como a perda de mandato. Um curso interno e avaliação externa deve ser realizada para avaliar a qualidade do serviço e a relevância para o usuário. O direito de propriedade intelectual também deve ser revisado regularmente, bem como a responsabilidade do repositório para regulamentar a não conformidade.

O repositório deve avaliar as habilidades de sua equipe e garantir a aquisição de novos funcionários ou reciclagem do pessoal existente, conforme necessário. A avaliação de risco regular também deve abordar ameaças e ataques de negação de serviço e perda ou qualidade inaceitável de serviços terceirizados. O requisito que trata sobre a política de coleta/depósito menciona a importância de um documento que especifique o tipo de informação que irá preservar, reter, gerenciar e fornecer acesso, ou seja, através da política de cobrança e documentos comprobatórios; política de preservação, missão, objetivos e visão do repositório. A política de coleta pode ser usada para entender o que o repositório mantém, o que ele não mantém, e por quê. A política de coleta apoia a missão mais ampla do repositório, sem tal política, o repositório provavelmente coletará de maneira aleatória ou armazenará grandes quantidades de conteúdo digital de baixo valor. A política de coleta ajuda a organização a identificar que conteúdo digital irá aceitar ou não, para a gestão. Em uma organização com uma ampla missão, a política de coleção ajuda a definir o papel do repositório dentro do contexto organizacional mais amplo.

O nono requisito do *CoreTrustSeal* trata sobre ‘Procedimentos de armazenamento documentados’, ou seja, o repositório aplica processos e procedimentos documentados durante o gerenciamento do armazenamento de arquivos de dados. O instrumento ACTDR possui requisito semelhante: ‘o repositório deve ter processos documentados para aquisição de PDI’. Os princípios TRUST não abordam este requisito.

De acordo com o *CoreTrustSeal* (CORETRUSTSEAL, 2020b), os repositórios precisam armazenar dados e metadados desde o ponto de depósito até o ponto de acesso. Para este requisito, o repositório deve oferecer evidências relacionadas as seguintes questões: como os processos e procedimentos são documentados e gerenciados; se o repositório tem uma estratégia para várias cópias; quais verificações existem para garantir a consistência entre as cópias de arquivo; e como a deterioração da mídia de armazenamento é tratada e monitorada. Os procedimentos são documentados e padronizados de forma que diferentes gerenciadores de dados, embora realizem as mesmas tarefas separadamente, cheguem ao mesmo resultado.

O ACTDR (*CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011) orienta que o repositório deve executar seus processos documentados para aquisição de PDI como também deve garantir que o PDI esteja persistentemente associado às informações de conteúdo relevantes. Isso é necessário para garantir que uma trilha auditável que apoia as reivindicações de autenticidade esteja disponível e para que alterações não autorizadas nos acervos digitais possam ser detectados, e que os objetos digitais possam ser identificados e colocados em seu contexto apropriado. O PDI é necessário não apenas para o repositório, mas para ajudar a garantir que as informações de conteúdo não sejam corrompidas (Fixity). O PDI deve estar permanentemente associado às informações de conteúdo.

O décimo requisito do *CoreTrustSeal*, ‘Plano de preservação’, diz que o repositório assume a responsabilidade pela preservação a longo prazo e gerencia essa função de forma planejada e documentada. O ACTDR possui entre seus requisitos, o seguinte critério: O repositório deve ter estratégias de preservação documentadas relevantes. O Princípio FAIR A2, trata sobre estratégias de preservação e possui relação com estes requisitos.

De acordo com o *CoreTrustSeal*, o repositório, os depositantes de dados e a comunidade designada precisam entender o nível de responsabilidade assumido para cada item depositado no repositório. O repositório deve ter direitos para assumir essas responsabilidades, ou seja, os procedimentos devem ser documentados e sua conclusão assegurada.

Para este requisito, as respostas de uma autoavaliação devem fornecer informações referentes as seguintes questões: se o repositório possui uma abordagem documentada para preservação; se o nível de responsabilidade pela preservação de cada item é compreendido e como isso é definido; se existem planos relacionados a migrações futuras ou medidas

semelhantes para lidar com a ameaça de obsolescência; se o contrato entre o depositante e o repositório prevê todas as ações necessárias para cumprir as responsabilidades; se o repositório tem direitos para copiar, transformar e armazenar os itens, bem como fornecer acesso a eles; as ações relevantes para a preservação estão especificadas na documentação, incluindo transferência de custódia, padrões de informações de envio e padrões de informações de arquivo.

Desta forma, o repositório deve fornecer em suas respostas, se possui uma documentação clara para garantir uma abordagem organizada para preservação de longo prazo, acesso contínuo para tipos de dados apesar das mudanças de formatos e se há documentação suficiente para apoiar a usabilidade pela a comunidade designada. Deve-se abordar se o repositório tem níveis de preservação definidos e, em caso afirmativo, como eles são aplicados. O plano de preservação deve ser gerenciado para garantir que as mudanças na tecnologia de dados e nos requisitos do usuário sejam tratadas de maneira estável e oportuna (CORETRUSTSEAL, 2020b).

Dito isto, o ACTDR menciona que isso é necessário para que fique claro como o repositório planeja garantir que a informação permanecerá disponível e utilizável para as gerações futuras e para fornecer um meio de verificar e validar o trabalho de preservação do repositório. Estas estratégias de preservação documentadas irão descrever como o repositório irá agir sobre riscos identificados, como parte do plano estratégico de preservação. Essas estratégias de preservação e o plano estratégico de preservação normalmente abordará a degradação da mídia de armazenamento, a obsolescência das unidades de mídia e a obsolescência ou inadequação da representação das informações (incluindo formatos) como a base de conhecimento da comunidade designada (*CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011).

O Princípio FAIR A2 menciona que quando necessário, dependendo das restrições ao acesso aos dados e/ou metadados, um mecanismo para autenticação e autorização para o acesso deve ser permitido pelo protocolo de comunicação (WILKINSON *et al*., 2016).

O requisito 11 do *CoreTrustSeal*, ‘Qualidade dos dados’, refere-se a experiência apropriada para lidar com dados técnicos e qualidade de metadados e a garantia que informações suficientes estejam disponíveis para os usuários finais fazerem avaliações relacionadas à qualidade. O requisito do ACTDR, ‘gestão de risco de segurança’ orienta que

deve-se manter uma análise sistemática dos fatores de risco de segurança associado a dados, sistemas, pessoal e planta física descrevendo alguns aspectos relacionados a qualidade. Os princípios TRUST aborda este requisito qualidade no item '*Responsibility*', informando que a responsabilidade é demonstrada aderindo aos padrões de metadados e curadoria, juntamente com o gerenciamento dos acervos de dados, por exemplo, validação técnica, documentação e controle de qualidade. Os Princípios FAIR I3 e R1 estão implícitos neste requisito (CORETRUSTSEAL, 2020b; *CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011; DAWEI LIN *et al*., 2020; WILKINSON *et al*., 2016).

De acordo com o *CoreTrustSeal*, os repositórios devem garantir que haja informações suficientes sobre os dados para que a comunidade designada avalie a qualidade dos dados. A avaliação da qualidade torna-se cada vez mais relevante quando a comunidade designada é multidisciplinar, onde os usuários podem não ter experiência pessoal para fazer uma avaliação da qualidade apenas a partir dos dados. Os repositórios devem ser capazes de avaliar a integridade e qualidade dos dados e metadados. Para este requisito é necessário descrever a abordagem da qualidade dos dados e metadados feita pelo repositório; deve-se descrever se o repositório possui verificações de controle de qualidade para garantir a integridade e a compreensibilidade dos dados depositados (em caso afirmativo, deve-se fornecer as referências aos padrões de controle de qualidade e mecanismos de relatório aceitos pela comunidade de prática relevante e incluir detalhes de como quaisquer problemas são resolvidos). Também é importante descrever a capacidade da comunidade de comentar e/ou classificar dados e metadados e se as citações de trabalhos relacionados ou links para índices de citações são fornecidos.

O ACTDR orienta que um curso interno e avaliação externa deve ser realizada para avaliar a qualidade do serviço e a relevância para o usuário. Direito de propriedade intelectual também deve ser revisado regularmente, bem como a responsabilidade do repositório para regulamentar a não conformidade. O repositório deve avaliar as habilidades de sua equipe e garantir a aquisição de novos funcionários ou reciclagem do pessoal existente, conforme necessário. A avaliação de risco regular também deve abordar ameaças e ataques de negação de serviço e perda ou qualidade inaceitável de serviços terceirizados O repositório pode realizar avaliações de risco geral.

Os Princípios FAIR I3 menciona que (Meta) dados devem incluir referências qualificadas para outros (Meta) dados, ou seja, deve-se referenciar os conjuntos de dados devidamente, possibilitando que conjuntos de dados gerados, a partir de outros conjuntos de dados, sejam ligados. O R1 estabelece que os (Meta) dados devem ser descritos com uma pluralidade de atributos precisos e relevantes.

O requisito 12 do *CoreTrustSeal* corresponde a 'Fluxos de trabalho', ou seja, o arquivamento deve ocorrer de acordo com fluxos de trabalho definidos, desde o armazenamento até a disseminação. No instrumento ACTDR, questões relacionadas a fluxos de trabalho, estão presentes em 3 requisitos: o repositório deve ter 'Políticas de Preservação em vigor para garantir que o seu Plano Estratégico de Preservação seja cumprido'; 'o repositório deve ter um processo de *ingest* que verifica cada SIP para integridade e correção'; e 'o repositório deve gerenciar o número e localização de cópias de todos os objetos digitais'. Não foram identificadas questões relacionadas a fluxo de trabalho nos princípios TRUST e FAIR.

De acordo com o instrumento *CoreTrustSeal*, para garantir a consistência das práticas entre conjunto de dados e serviços, os fluxos de trabalho devem ser definidos de acordo com as atividades do repositório e claramente documentados. O modelo de referência OAIS pode ajudar a especificar as funções de fluxo de trabalho de um repositório. Para este requisito deve-se descrever: os fluxos de trabalho (descrições de processos de negócios); a comunicação clara para depositantes e usuários sobre o manuseio de dados; os níveis de segurança e impacto nos fluxos de trabalho (proteção da privacidade dos sujeitos); a verificação qualitativa dos resultados; os tipos de dados gerenciados e qualquer impacto no fluxo de trabalho; e o gerenciamento de mudanças de fluxos de trabalho. Este requisito confirma que todos os fluxos de trabalho estão documentados, ou seja, o repositório deve adotar uma abordagem consistente, rigorosa e documentada para gerenciar todas as atividades em seus processos e que as alterações nesses processos são implementadas, avaliadas, registradas e administradas de maneira adequada. O requisito não exige descrições detalhadas dos fluxos de trabalho, mas busca evidências de como e onde esses fluxos de trabalho são documentados (CORETRUSTSEAL, 2020b).

O ACTDR indica utilizar fluxos de trabalho nos mecanismos para revisão, nas atualizações e desenvolvimento de suas políticas de preservação à medida que o repositório

cresce e à medida que a tecnologia e a prática da comunidade evolui. É necessário, para que o repositório tenha políticas completas e atualizadas, procedimentos em vigor que refletem os requisitos e práticas de preservação para sua comunidade. O instrumento também indica a utilização de fluxos de trabalho para detectar e corrigir erros no SIP quando criado e potencial erros de transmissão entre o depositante e o repositório. Além disso, utiliza-se fluxos de trabalho em mecanismos para garantir que qualquer/várias cópias de objetos digitais sejam sincronizadas, ou seja, é necessário para garantir que várias cópias de um objeto digital permaneçam idênticas, dentro de um prazo estabelecido como aceitável pelo repositório, e que uma cópia possa ser usada para ser substituída por uma cópia corrompida do objeto (*CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011).

O Requisito 13 do CoreTrustSeal, 'Descoberta e identificação de dados', orienta que o repositório deve permitir que os usuários descubram os dados e os consultem de forma persistente por meio de citações adequadas. O ACTDR possui um requisito semelhante ao *CoreTrustSeal:* 'Gestão de informações', onde o repositório deve especificar os requisitos mínimos de informação para permitir a comunidade designada a descobrir e identificar materiais de interesse. Os princípios FAIR (F1, F2, F3, F4) estão de certa forma, implícitos neste requisito.

Segundo o *CoreTrustSeal*, a descoberta de dados eficaz é a chave para o compartilhamento de dados. Uma vez descobertos, os conjuntos de dados devem ser referenciados por meio de citações completas, incluindo identificadores persistentes para ajudar a garantir que os dados possam ser acessados no futuro. Para este requisito, o repositório deve incluir evidências referentes aos seguintes aspectos: se o repositório oferece recursos de pesquisa; se o repositório mantém um catálogo de metadados pesquisável para padrões apropriados; quais sistemas de identificadores persistentes o repositório usa; se o repositório facilita a coleta automática dos metadados; e se o repositório oferece recomendação de citação de dados. Ou seja, o repositório deve dar evidências de que toda a curadoria de dados e metadados apoia a descoberta de objetos digitais claramente definidos e identificados e permite sua vinculação com objetos digitais relacionados de acordo com os padrões de domínio. Deve ficar claro para a comunidade como os dados são citados, de modo que o crédito e a atribuição apropriados sejam dados aos indivíduos/organizações que contribuíram para sua criação.

Da mesma forma, o ACTDR, menciona que este requisito é necessário para permitir a descoberta dos acervos do repositório. Pode-se atingir esse requisito através das informações descritivas de recuperação, metadados de descoberta, como Dublin Core e outras documentações que descrevem o objeto. O repositório deve ser capaz de lidar com os tipos de solicitações que virão de um típico usuário da comunidade designada.

Os Princípios FAIR F1, F2, F2 e F4 mencionam que os (meta)dados devem ter identificadores globais, persistentes e identificáveis como, por exemplo, DOI, ARK entre outros e ser descritos com metadados ricos o suficiente para que, uma vez indexados, possam ajudar o usuário a encontrar os dados mesmo que não haja um identificador. Os metadados devem incluir claramente e explicitamente os identificadores dos dados que descrevem, ou seja, como não podemos prever que os dados e seus metadados estejam sempre juntos, por exemplo, quando os metadados são indexados por um mecanismo de busca e, portanto, está em uma plataforma diferente dos dados, a associação entre eles deve ser feita pela inclusão do identificador dos dados pelos metadados. Para que os dados sejam encontrados, seus metadados devem ser indexados por mecanismos de busca que, por sua vez, permitem aos usuários encontrá-los por meio de elementos desses metadados (CORETRUSTSEAL, 2020b; *CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011; WILKINSON *et al*., 2016).

O requisito 14 do *CoreTrustSeal*, ‘Reutilização de dados’, menciona que o repositório deve permitir a reutilização dos dados ao longo do tempo, garantindo que os metadados apropriados estejam disponíveis para apoiar a compreensão e o uso dos dados. Da mesma forma, o requisito ‘O repositório deve rastrear e gerenciar os direitos de propriedade intelectual e restrições ao uso do conteúdo do repositório, conforme exigido pelo acordo de depósito, contrato ou licença’, do ACTDR também trata sobre a questão da reutilização de dados. Observou-se que os princípios TRUST, da mesma forma que o *CoreTrustSeal* e o ACTDR, também aborda a mesma questão através do critério ‘*User focus* (Foco no usuário)’.

O *CoreTrustSeal* orienta que os repositórios devem garantir que os dados continuem a ser compreendidos e usados de forma eficaz no futuro, apesar das mudanças na tecnologia e na base de conhecimento da comunidade designada. Este requisito avalia as medidas tomadas para garantir que os dados sejam reutilizáveis. Para este requisito deve-se dar evidências em relação a quais metadados são fornecidos pelo repositório quando os dados são acessados;

como o repositório garante a compreensão contínua dos dados; se os dados são fornecidos em formatos usados pela comunidade designada e quais são esses formatos; e se são tomadas medidas para levar em consideração a possível evolução dos formatos.

Para atender a este requisito o repositório deve demonstrar um conhecimento profundo dos cenários de reutilização e das necessidades da comunidade em termos de suas práticas, ambiente técnico e adesão aos padrões aplicáveis. Mudanças na tecnologia e nas metodologias e normas empregadas podem levar a necessidade de reconsiderar o formato em que os dados são divulgados. Da mesma forma, metadados apropriados de alta qualidade em conformidade com um esquema generalizado e/ou disciplinar específico desempenham um papel essencial e devem ser mencionados nas evidências fornecidas (CORETRUSTSEAL, 2020b).

O ACTDR relata que isso é necessário para permitir que o repositório rastreie, atue e verifique os direitos e restrições relacionadas ao uso dos objetos digitais dentro do repositório. Pode-se atender a esse requisito através de uma declaração de Política de Preservação que define e especifica os requisitos do repositório e processo de gestão de direitos de propriedade intelectual; acordos de depositantes; amostras de acordos e outros documentos que especificam e tratam dos direitos de propriedade intelectual; documentação de monitoramento ao longo do tempo de mudanças no status de propriedade intelectual em conteúdo digital detido pelo repositório; resultados do monitoramento e metadados que capturam informações de direitos. O repositório deve ter um mecanismo de rastreamento de licenças e contratos aos quais é obrigado. Qualquer que seja o formato do sistema de rastreamento, ele deve ser suficiente para a instituição rastrear, agir e verificar os direitos e restrições relacionados ao uso dos objetos digitais dentro do repositório (*CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011).

Os Princípios TRUST mencionam que os repositórios têm um papel vital na aplicação e reforço das normas e padrões, o que inclui os esquemas de metadados, formatos de arquivos de dados, vocabulários controlados, ontologias e outras semânticas (DAWEI LIN *et al.*, 2020).

- *User focus* (Foco no usuário)

Os princípios TRUST (DAWEI LIN *et al.*, 2020) mencionam que um repositório confiável precisa se concentrar em servir sua comunidade de usuários-alvo. Cada comunidade de usuários provavelmente tem expectativas diferentes em relação aos seus repositórios. Existe uma visão ampla quanto a comunidade de usuários, ou seja, existem aqueles que acessam acervos de dados e partes interessadas indiretas como, financiadores, editores de periódicos, outros parceiros institucionais e cidadãos.

O uso e reutilização de dados de pesquisa são parte integrante do processo científico e, portanto, dos repositórios confiáveis, que deve permitir que sua comunidade encontre, explore e entenda seus acervos de dados em relação ao potencial de reuso. Os repositórios têm um papel vital na aplicação e reforço das normas e padrões, o que inclui os esquemas de metadados, formatos de arquivos de dados, vocabulários controlados, ontologias e outras semânticas. Um repositório confiável deve demonstrar métricas de dados relevantes e disponibilizá-las aos usuários; fornecer catálogos para facilitar a descoberta de dados e monitorar e identificar as expectativas em relação a comunidade e responder conforme necessário. Este requisito não foi encontrado explicitamente nos demais instrumentos analisados.

- Tecnologia/ Gestão de risco de infraestrutura e segurança/ *Sustainability* (Sustentabilidade)/*Technology*

O requisito 15 do *CoreTrustSeal* corresponde a ‘Infraestrutura técnica’ e orienta que o repositório funcione em sistemas operacionais bem suportados e em outro software de infraestrutura central, e use tecnologias de hardware e software apropriadas para os serviços que fornece à sua comunidade designada. O ACTDR possui uma subdivisão ‘Gestão de risco de infraestrutura e segurança’ e o requisito ‘gestão de risco de infraestrutura técnica’, indicando que o repositório deve identificar e gerenciar os riscos às suas operações de preservação e objetivos associados à infraestrutura do sistema. Os Princípios FAIR A1, A1.1, I1, I2 e R1.3 possuem relação com este requisito e tratam sobre os dados serem acessíveis. Não foram encontrados estes requisitos entre os princípios TRUST.

O *CoreTrustSeal* orienta que os repositórios precisam operar em infraestruturas centrais confiáveis e estáveis que maximizam a disponibilidade do serviço. Ademais, o

hardware e o software usados devem ser relevantes e apropriados para a comunidade e para as funções que o repositório desempenha. Para este requisito, deve-se responder as seguintes questões: quais padrões o repositório usa para referência (se são padrões internacionais ou comunitários) e com que frequência são revisados; de que forma esses padrões são implementados; se o repositório possui um plano de desenvolvimento de infraestrutura; se existe um inventário de software e se a documentação do sistema está disponível; se a disponibilidade e conectividade são suficientes para atender as necessidades da comunidade; e se existem procedimentos e arranjos para fornecer recuperação rápida ou backup de serviços essenciais no caso de uma interrupção. Assim sendo, os fluxos de trabalho e os atores humanos que fornecem serviços de repositório devem ser suportados por uma infraestrutura tecnológica adequada que atenda às necessidades da comunidade e permita que o repositório se recupere de desastres de curto prazo. O repositório deve demonstrar que compreende o ecossistema mais amplo de padrões, ferramentas e tecnologias disponíveis para gerenciamento e curadoria de dados de pesquisa e que selecionou opções que se alinham com requisitos locais. Se possível deve-se comprovar utilizando-se um modelo de referência como, por exemplo, *Spatial Data Infrastructure* (SDI), *Open Geospatial Consortium* (OGC), W3C ou padrões ISO (CORETRUSTSEAL, 2020b).

De acordo com ACTDR este requisito é necessário para garantir uma infraestrutura segura e confiável. Deve-se utilizar um inventário de infraestrutura de componentes do sistema; avaliações periódicas de tecnologia; estimativas da vida útil dos componentes do sistema; exportação de registros autênticos para um sistema independente; uso de software fortemente suportado pela comunidade, por exemplo, Apache, iRODS, Fedora e recriação de arquivos de backups. O repositório deve realizar ou contratar avaliações dos riscos relacionados ao hardware e infraestrutura de software e procedimentos operacionais.

O repositório deve fornecer mecanismos que minimizam o risco de dependências em sistema proprietário ou obsoleto e de erro operacional. O grau de suporte necessário está relacionado a criticidade do(s) subsistema(s) envolvido(s) na preservação de longo prazo. O repositório deve manter um sistema que seja escalável sem uma grande interrupção do sistema. O repositório deve manter um sistema que é evolutivo, ou seja, o sistema deve ser projetado de tal forma que os principais componentes do sistema possam ser substituídos por tecnologias mais novas sem grandes interrupções do sistema como um todo. O sistema de

repositório deve ser extensível, ou seja, o sistema deve ser projetado para acomodar formatos futuros (mídia e arquivos) sem grande perturbação do sistema como um todo. O repositório deve ser capaz de exportar seus acervos para um futuro custodiante e deve ser capaz de recriar os arquivos após um erro de operação que substitui ou exclui acervos digitais.

Os subitens deste requisito mencionam que o repositório deve empregar vigilância tecnológica ou outra tecnologia de monitoramento de sistemas de notificação (deve possuir tecnologias de hardware adequadas aos serviços e fornecer às suas comunidades designadas; deve ter procedimentos para monitorar e receber notificações quando mudanças de tecnologia de hardware são necessárias; deve ter procedimentos, compromisso e financiamento para substituir o hardware quando a avaliação indica a necessidade de fazê-lo; deve ter procedimentos para monitorar e receber notificações quando mudanças de software são necessárias; ter procedimentos em vigor para avaliar quando as mudanças são necessárias para o software atual; e ter procedimentos, compromisso e financiamento para substituir software quando a avaliação indica a necessidade de fazê-lo).

O repositório deve ter suporte adequado de hardware e software para backup e funcionalidade suficiente para preservar o conteúdo do repositório e rastrear as funções do repositório. Esse item é necessário para garantir o acesso contínuo e o rastreamento das funções de preservação aplicadas aos objetos digitais sob sua custódia.

Além disso, o repositório deve ter mecanismos eficazes para detectar erros nos dados ou perda de bits, ou seja, é necessário para garantir que os AIPs e os metadados não estejam corrompidos ou para quaisquer perdas de dados detectadas que se enquadram nas tolerâncias estabelecidas pela política de repositório (o repositório deve registrar e relatar à sua administração todos os incidentes, erros nos dados ou perda de dados e medidas devem ser tomadas para reparar/substituir dados corrompidos ou perdidos).

O repositório deve ter um processo para registrar e reagir à disponibilidade de novas atualizações de segurança com base em uma avaliação de risco-benefício. Entende-se que este requisito é necessário para proteger a integridade dos objetos de arquivos não autorizados de alterações ou exclusões. Também deve ter processos definidos para mídia de armazenamento e/ou mudança de hardware (por exemplo, atualização, migração), pois é necessário para garantir que os dados não sejam perdidos quando a mídia falhar ou o hardware de suporte não poder mais ser usado para acessar os dados.

O repositório deve ter processos críticos identificados e documentados que podem afetar sua capacidade de cumprir com suas responsabilidades obrigatórias tornando-se necessário para garantir que os processos críticos possam ser monitorados para garantir que eles continuam a cumprir as responsabilidades obrigatórias e a garantir que quaisquer alterações a esses os processos sejam examinados e testados, ou seja, o repositório deve ter um processo de gerenciamento de mudanças documentado que identifica mudanças em processos críticos que afetam potencialmente a capacidade do repositório de cumprir suas responsabilidades obrigatórias; como também possuir um processo para testar e avaliar o efeito de mudanças nos processos críticos do repositório (*CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011).

Os Princípios FAIR A1, A1.1 mencionam que os (Meta) dados devem ser recuperáveis pelos seus identificadores usando um protocolo de comunicação padronizado, aberto, gratuito e universalmente implementável. Já os Princípios FAIR I1 e I2 mencionam que os (Meta)dados devem ser representados por meio de uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento. Como os dados e metadados devem possuir referências aos vocabulários que contenham os conceitos utilizados, devemos garantir que sejam utilizados vocabulários que também sigam os princípios FAIR. Já o Princípio R1.3, diz que os (Meta) dados devem estar alinhados com padrões relevantes ao seu domínio, ou seja, deve atender os padrões específicos da comunidade da área e às boas práticas de arquivamento e ao compartilhamento do campo de pesquisa específico (WILKINSON *et al*., 2016).

O requisito 16 do *CoreTrustSeal* refere-se a ‘Segurança’, ou seja, a infraestrutura técnica do repositório deve oferecer proteção para as instalações e seus dados, produtos, serviços e usuários. O instrumento ACTDR possui alguns requisitos relacionados à segurança: Gestão de acesso (o repositório deve obedecer às Políticas de Acesso). Na subdivisão ‘Gestão de risco de infraestrutura e segurança’, no requisito ‘gestão de risco de infraestrutura técnica’ (o repositório deve identificar e gerenciar os riscos às suas operações de preservação e objetivos associados à infraestrutura do sistema). E no requisito ‘gestão de risco de segurança’ (o repositório deve manter uma análise sistemática dos fatores de risco de segurança associado a dados, sistemas, pessoal e planta física). Entre os princípios TRUST, nos itens

Responsibility e Technology foram encontradas menções sobre segurança. O Princípio FAIR A 1.2 está implícito neste requisito.

Conforme consta no *CoreTrustSeal* (CORETRUSTSEAL, 2020b), o repositório deve analisar ameaças potenciais, avaliar riscos e criar um sistema de segurança consistente. Deve descrever cenários de danos com base em ações maliciosas, erro humano ou falha técnica que representam uma ameaça ao repositório e seus dados, produtos, serviços e usuários. Deve-se medir a probabilidade e o impacto de tais cenários, decidir quais níveis de risco são aceitáveis e determinar quais medidas devem ser tomadas para combater as ameaças ao repositório e sua comunidade designada. Para este requisito deve ser descrito: o sistema de segurança de TI, funcionários com funções relacionadas à segurança e quaisquer ferramentas de análise de risco utilizados; os níveis de segurança que são exigidos e como eles são suportados; deve-se descrever quaisquer procedimentos de autenticação e autorização empregados para gerenciar com segurança o acesso aos sistemas em uso.

É necessário demonstrar que compreende todos os riscos técnicos aplicáveis ao serviço prestado bem como ao ambiente físico. Além disso, deve-se demonstrar que possui mecanismos para prevenir, detectar e responder a um incidente de segurança, ou seja, deve demonstrar de que forma a segurança da infraestrutura técnica é controlada pelo repositório e por sua instituição hospedeira/terceirizada e quem está no comando. Também deve demonstrar se os procedimentos de autenticação e autorização são suficientes para garantir a segurança dos acervos de dados em cada estágio do fluxo de trabalho e mostrar quais políticas de segurança da empresa estão em vigor para controlar segurança de todos os sistemas, incluindo segurança de rede, verificações de intrusos, segurança de instalações físicas e política de senhas.

Conforme os requisitos citados e encontrados no ACTDR (*CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS*, 2011), o termo "acesso" tem vários sentidos diferentes, incluindo o acesso dos usuários ao sistema de repositório, por exemplo, segurança física e autenticação de usuário, e os diferentes estágios de acesso aos registros (fazer uma solicitação, verificar os direitos do solicitante e preparar e enviar um Pacote de Informação de Disseminação DIP). Esta subseção é dividida em dois requisitos principais, um relacionado com o existência e implementação de políticas de acesso, e uma com capacidade do repositório para fornecer objetos comprovadamente autênticos como DIPs. Assim, o primeiro

requisito se refere a solicitações iniciadas por um usuário e como o repositório as trata para garantir que os direitos e acordos sejam respeitados, que a segurança seja monitorada, que as solicitações sejam atendidas, etc. o segundo requisito se relaciona com o que é entregue ao consumidor e a confiança que pode ser colocada nele. Deve ser entendido que as capacidades e sofisticação do sistema de acesso irão variar dependendo da comunidade designada do repositório e dos mandatos de acesso do repositório. Devido à variedade de repositórios e mandatos de acesso, esses critérios podem ser sujeitos a questões sobre aplicabilidade e interpretação a nível local.

Ainda tratando sobre segurança, o repositório deve registrar e revisar todas as falhas de gerenciamento de acesso e anomalias, desta maneira é necessário identificar ameaças à segurança e falhas no sistema de gerenciamento de acesso. Um repositório deve ter algum mecanismo automatizado para observar negações anômalas ou incomuns e usá-los para identificar ameaças de segurança ou falhas no sistema de gerenciamento de acesso, como o acesso negado a usuários válidos.

O repositório deve ter um processo para registrar e reagir à disponibilidade de novas atualizações de segurança com base em uma avaliação de risco-benefício, assim este requisito é necessário para proteger a integridade dos objetos de arquivos não autorizados de alterações ou exclusões. As decisões de aplicar atualizações de segurança provavelmente serão o resultado de uma avaliação de risco-benefício; *patches* de segurança são frequentemente responsáveis por perturbar aspectos alternativos do sistema, funcionalidades ou desempenho. Pode não ser necessário para um repositório implementar todos *patches* de software e a aplicação de qualquer um devem ser cuidadosamente considerados.

Cada atualização de segurança implementada pelo repositório deve ser documentada com detalhes sobre como é completado e atualizações automáticas e manuais são aceitáveis. Atualizações de segurança significativas podem pertencer a software diferente de sistemas operacionais centrais, como aplicativos de banco de dados e servidores da Web, e estes também devem ser documentados. As atualizações de segurança não se limitam a atualizações de segurança de software. Atualizações no hardware real ou no *firmware* do sistema de hardware estão incluídos. Com o tempo, é provável que as atualizações de segurança também sejam necessárias para o repositório. Embora as atualizações de segurança possam ser consideradas como parte do controle de mudança, elas são identificadas separadamente aqui

porque muitas vezes há serviços que compilam e divulgam informações sobre problemas e atualizações de segurança. Em um mínimo, os repositórios devem monitorar esses serviços para garantir que o repositório que mantém os dados não esteja sujeito a comprometimento por ameaças identificadas.

O repositório deve mostrar como lidou com seus requisitos de segurança. Se alguns tipos de materiais tem maior probabilidade de ser atacado, o repositório precisará fornecer mais proteção, por exemplo, os repositórios que experimentaram incidentes podem registrar tais instâncias, incluindo os momentos em que os sistemas ou conteúdo foram afetados e descrever os procedimentos que foi implementado para evitar ocorrências semelhantes no futuro. Também podem conduzir uma variedade de exercícios de desastre que podem envolver sua organização mãe ou a comunidade. Os planos de contingência são especialmente importantes e precisam ser testados, atualizados e revisados em uma base regular.

Entre os princípios TRUST existem especificações para gerenciar os direitos de propriedade intelectual dos produtores de dados, proteção de recursos de informações confidenciais, e a segurança do sistema e seu conteúdo; como também ter planos e mecanismos em vigor para prevenir, detectar e responder à ameaças à segurança física ou cibernética (DAWEI LIN *et al.*, 2020).

De acordo com o Princípio FAIR A 1.2, quando necessário, dependendo das restrições ao acesso aos dados e/ou metadados, um mecanismo para autenticação e autorização para o acesso deve ser permitido pelo protocolo de comunicação (WILKINSON *et al*., 2016).

Foi observado que alguns requisitos que constavam no ACTDR não estavam contemplados de forma clara nos instrumentos *CoreTrustSeal,* TRUST e FAIR. Comparando os instrumentos relacionados nesta pesquisa nota-se que o ACTDR apresenta seus requisitos de forma mais extensa e detalhada, o que pode dificultar a compreensão e aplicação.

4.3 REUNIÃO DOS CRITÉRIOS/REQUISITOS PARA GARANTIR A CONFIABILIDADE DE REPOSITÓRIOS DE DADOS DE PESQUISA

A partir da revisão bibliográfica e análise documental minuciosa sobre os critérios/requisitos de avaliação de repositórios de dados de pesquisa, elaborou-se um conjunto de critérios/requisitos, conforme descritos no quadro 8.

Quadro 8 – Conjunto de critérios/requisitos elaborado a partir da análise documental para ser utilizado por repositórios de dados de pesquisa em um contexto brasileiro.

INFRAESTRUTURA ORGANIZACIONAL/TRANSPARÊNCIA
Requisitos/Princípios: - Missão/escopo (*CoreTrustSeal*) - Governança e viabilidade organizacional (ACTDR) - *Transparency* (TRUST)
Estratégias a serem utilizadas pelos repositórios
1 Descrever na missão da organização a missão em preservar e fornecer acesso aos dados.
2 Declarar de forma transparente o prazo mínimo de preservação digital para os acervos de dados.
Requisitos/Princípios: - Licenças (*CoreTrustSeal*) - Contratos, licenças e responsabilidades(ACTDR) - *Responsibility* (TRUST) - *Reusable* (FAIR)
Estratégias a serem utilizadas pelos repositórios
1 Possuir regulamentos de acesso e licenças aplicáveis cobrindo o uso de dados.
2 Possuir descrito as condições de uso e a proteção de dados confidenciais.
3 Possuir uma política pública em vigor para o descumprimento no caso de divulgação de dados pessoais.

4 Possuir contratos de depósitos devidamente assinados e executados de acordo com as normas locais, leis e regulamentos nacionais e internacionais.
5 Possuir políticas sobre acordos de depósitos de terceiros e usos permitidos.
Requisitos/Princípios: - Continuidade de acesso (*CoreTrustSeal*) - Governança e viabilidade organizacional (ACTDR) - *Sustainability* (TRUST)
Estratégias a serem utilizadas pelos repositórios
1 Possuir um plano de continuidade para garantir o acesso contínuo e a preservação de seus acervos.
2 Possuir descrito a governança relacionada à operação contínua do repositório ao longo do tempo e durante desastres, bem como evidências em relação ao planejamento de sucessão, designadamente, as medidas em vigor para garantir o acesso e a disponibilidade dos acervos de dados, atualmente e no futuro.
3 Ter acordo formal para garantir quem assumirá a responsabilidade em caso de descontinuidade do serviço.
4 Possuir um plano estratégico de preservação com definições administrativas que defina a abordagem que o repositório terá no suporte de longo prazo.
5 Possuir um plano de sucessão adequado, planos de contingência, e/ou acordos de custódia em vigor no caso do repositório deixar de operar ou a instituição governamental ou de financiamento alterar substancialmente seu escopo.
Requisitos/Princípios: - Confidencialidade/Ética (*CoreTrustSeal*) - *Responsibility* (TRUST)
Estratégias a serem utilizadas pelos repositórios
1 Garantir que os dados sejam tratados com curadoria, usados e acessados em conformidade com as normas.
2 Garantir que a coleta e criação de dados seja realizada com os critérios legais e éticos.
3 Gerenciar de maneira adequada os dados com risco de divulgação para limitar acesso.
4 Treinar os funcionários do repositório em relação ao gerenciamento de dados com risco de divulgação.

5 Possuir medidas em vigor caso as condições não sejam cumpridas.
6 Possuir os procedimentos em vigor documentados para garantir a conformidade.
7 Gerenciar os direitos de propriedade intelectual dos produtores de dados, proteger os recursos de informações confidenciais, a segurança do sistema e o seu conteúdo.
Requisitos/Princípios: - Infraestrutura organizacional (*CoreTrustSeal*) - Estrutura organizacional e pessoal (ACTDR) - *Technology* (TRUST)
Estratégias a serem utilizadas pelos repositórios
1 Possuir financiamento adequado e número suficiente de funcionários.
2 Ser hospedado por uma instituição reconhecida para a garantia da estabilidade e sustentabilidade de longo prazo.
3 Ter financiamento suficiente incluindo recursos de equipe e TI.
4 Possuir um plano de pessoal, definições de competência, e desenvolvimento profissional de equipe.
Requisitos/Princípios: - Orientação de especialista (*CoreTrustSeal*) - Contratos, licenças e responsabilidades (ACTDR)
Estratégias a serem utilizadas pelos repositórios
1 Possuir um aconselhamento e feedback de especialistas para garantir sua relevância e melhorias contínuas.
2 Possuir consultores internos ou um comitê consultivo (especialistas técnicos).
3 Examinar por meio de especialistas as políticas de preservação e planos de implementação do repositório.

GERENCIAMENTO DE OBJETOS DIGITAIS
Requisitos/Princípios: - Integridade e autenticidade dos dados (*CoreTrustSeal*) - Preservação de AIP (ACTDR) - Gestão de informações (ACTDR) - *Reusable* (FAIR)
Estratégias a serem utilizadas pelos repositórios
1 Fornecer evidências mostrando que opera um sistema de gerenciamento de dados e metadados adequado para garantir integridade e autenticidade durante o processo de armazenamento e acesso a dados.
2 Descrever a documentação da integridade em detalhes, assim como todas as alterações e descrições de controle de versões.
3 Possuir especificações de como os AIPs (*Archival Information Package*) – conjunto de informações contendo qualidades referente a preservação de determinado objeto de informação - são armazenados.
4 Disponibilizar a representação da informação para cada AIP.
5 Utilizar metadados descritivos e identificadores únicos persistentes associado ao AIP.
Requisitos/Princípios: - Avaliação (*CoreTrustSeal*) - Responsabilidade processual e política de preservação (ACTDR)
Estratégias a serem utilizadas pelos repositórios
1 Revisar os critérios de seleção ao longo do tempo e os ativos digitais devem ser reavaliados adequadamente.
2 Possuir uma política de desenvolvimento de coleção para orientar a seleção de dados.
3 Realizar uma avaliação automatizada da aderência dos metadados aos esquemas relevantes.
4 Possuir algum cronograma regular de autoavaliação e certificação externa.

5 Possuir listas de verificação preenchidas e datadas de autoavaliações e/ou auditorias.
Requisitos/Princípios: - Procedimentos de armazenamento documentados (*CoreTrustSeal*) - *Ingest*: criação do *Archival Information Package* (AIP). (ACTDR)
Estratégias a serem utilizadas pelos repositórios
1 Documentar os procedimentos e padronizar de forma que diferentes gerenciadores de dados, embora realizem as mesmas tarefas separadamente, cheguem ao mesmo resultado.
Requisitos/Princípios: - Plano de preservação (*CoreTrustSeal*) - Planejamento de preservação (ACTDR) - *Accessible* (FAIR)
Estratégias a serem utilizadas pelos repositórios
1 Especificar na documentação as ações relevantes para a preservação, incluindo a transferência de custódia, padrões de informações de envio e padrões de informações de arquivo.
2 Desenvolver alguma documentação que identifique os riscos de preservação e as estratégias para lidar com esses riscos.
3As estratégias de preservação e o plano estratégico de preservação deve abordar a degradação da mídia de armazenamento, a obsolescência das unidades de mídia e a obsolescência ou inadequação da representação das informações (incluindo formatos) como a base no conhecimento da comunidade designada.
Requisitos/Princípios: - Qualidade dos dados (*CoreTrustSeal)* - *Responsibility* (TRUST) *Interoperable/Reusable* (FAIR)
Estratégias a serem utilizadas pelos repositórios
1 Possuir verificações de controle de qualidade para garantir a integridade e a compreensibilidade dos dados depositados.

2 Possibilitar que a comunidade designada comente ou classifique dados e metadados.
3 Fornecer citações de trabalhos relacionados aos dados ou links para índices de citações.
4 Possuir validação técnica e controle de qualidade.
Requisitos/Princípios: - Fluxos de trabalho (*CoreTrustSeal*) -ACTDR
Estratégias a serem utilizadas pelos repositórios
1 Definir os fluxos de trabalho (descrições de processos de negócios) e a comunicação clara para depositantes e usuários sobre o manuseio de dados.
2 Definir os níveis de segurança e impacto nos fluxos de trabalho (proteção da privacidade dos sujeitos).
3 Adotar uma abordagem consistente, rigorosa e documentada para gerenciar todas as atividades em seus processos.
Requisitos/Princípios: - Descoberta e identificação de dados (*CoreTrustSeal*) - Gestão de informações (ACTDR) *Findable* (FAIR)
Estratégias a serem utilizadas pelos repositórios
1 Oferecer recursos de pesquisa e manter um catálogo de metadados pesquisável para padrões apropriados.
2 Facilitar a coleta automática dos metadados e oferecer recomendação de citação de dados.
3 Fornecer evidências de que toda a curadoria de dados e metadados apoia a descoberta de objetos digitais claramente definidos e identificados e permitir sua vinculação com objetos digitais relacionados de acordo com os padrões de domínio.
4 Esclarecer para a comunidade como os dados são citados, de modo que o crédito e a atribuição apropriado seja dado aos indivíduos/organizações que contribuíram para sua criação.
5 Especificar os requisitos mínimos de informação para permitir a comunidade designada a descobrir e

identificar materiais de interesse.
Requisitos/Princípios: - Reutilização de dados (*CoreTrustSeal*) - *User focus* (TRUST)
Estratégias a serem utilizadas pelos repositórios
1 Garantir que os dados continuem a ser compreendidos e usados de forma eficaz no futuro, apesar das mudanças na tecnologia e na base de conhecimento da comunidade designada.
2 Fornecer catálogos para facilitar a descoberta de dados.
TECNOLOGIA/GESTÃO DE RISCO DE INFRAESTRUTURA E SEGURANÇA/*SUSTAINABILITY*
Requisitos/Princípios: - Infraestrutura técnica (*CoreTrustSeal*) - Gestão de risco de infraestrutura técnica (ACTDR) - *Reusable* (FAIR)
Estratégias a serem utilizadas pelos repositórios
1 Operar em infraestruturas centrais confiáveis e estáveis que maximizam a disponibilidade do serviço.
2 Descrever quais padrões o repositório usa para referência (se são padrões internacionais ou comunitários) e com que frequência são revisados.
3 Possuir um plano de desenvolvimento de infraestrutura.
4 Desenvolver um inventário de software e a documentação do sistema deve estar disponível.
5 Elaborar procedimentos e arranjos para fornecer recuperação rápida ou backup de serviços essenciais no caso de uma interrupção.
6 Realizar avaliações periódicas de tecnologia e fazer estimativas da vida útil dos componentes do sistema.
7 O repositório deve realizar ou contratar avaliações dos riscos relacionados ao hardware e infraestrutura de

software e procedimentos operacionais.
Requisitos/Princípios: - Segurança (*CoreTrustSeal*) - Gestão de risco de segurança (ACTDR) - *Technology* (TRUST) - *Accessible* (FAIR)
Estratégias a serem utilizadas pelos repositórios
1 Analisar ameaças potenciais, avaliar riscos e criar um sistema de segurança consistente.
2 Possuir mecanismos para prevenir, detectar e responder a um incidente de segurança, ou seja, demonstrar de que forma a segurança da infraestrutura técnica é controlada pelo repositório e por sua instituição hospedeira/terceirizada e quem está no comando.
3 Realizar avaliações de risco regulares e manter a segurança adequada a fim de fornecer os níveis de serviço previstos e contratados.
4 Manter uma análise sistemática dos fatores de risco de segurança associado a dados, sistemas, pessoal e planta física.
5 Ter planos e mecanismos em vigor para prevenir, detectar e responder à ameaças à segurança física ou cibernética.

Fonte: elaborado pela autora com base nos instrumentos analisados.

O quadro 8 ilustra o conjunto de critérios/requisitos elaborado para ser utilizado por repositórios de dados de pesquisa brasileiros em autoavaliações e foi construído da seguinte maneira: Nas linhas/cabeçalhos grafadas em tom rosa e azul constam respectivamente a infraestrutura/área e os requisitos/princípios dos instrumentos analisados neste estudo. Nas linhas abaixo foram elencadas as estratégias a serem utilizadas pelo repositório. Essas estratégias são critérios/requisitos interpretados a partir da discussão e confrontação dos instrumentos selecionados.

A partir de uma autoavaliação utilizando-se o conjunto de critérios/requisitos elaborado nesta pesquisa é possível examinar um repositório numa abordagem direcionada.

As estratégias a serem utilizadas pelos repositórios devem ser registradas em documentos específicos como políticas, manuais e regulamentos, servindo como evidências em futuras avaliações e certificações. Com esta estrutura de avaliação poderá ser possível melhorar os serviços disponibilizados atualmente pelos repositórios de dados de pesquisa no contexto brasileiro e a entender como os gestores dos repositórios estão trabalhando no desenvolvimento de repositórios de qualidade e confiáveis.

5 CONSIDERAÇÕES FINAIS

Instituições brasileiras estão em busca do desenvolvimento e implementação de repositórios de dados de pesquisa e observa-se a grande necessidade de ampliar os estudos referentes à avaliação, pois trata-se de uma etapa de extrema importância na implementação e acompanhamento de sistemas de informação. Internacionalmente, existem instituições que avaliam e oferecem certificações para repositórios de dados, no entanto, no Brasil, os estudos estão em fases iniciais, com falta de padronização e clareza na definição de critérios avaliativos.

Sobre estudos de repositórios confiáveis, entende-se que é uma das linhas de pesquisa em preservação digital mais trabalhadas em nível internacional e que se busca definir metodologias e ferramentas de avaliação de acordo com padrões e pode-se perceber a partir da literatura abordada nesta pesquisa que as publicações sobre o assunto, repositórios de dados de pesquisa confiáveis, ainda são incipientes no contexto brasileiro.

Repositório confiável é aquele cuja missão é fornecer acesso de longo prazo a recursos digitais gerenciados e para manter o status de confiança, um repositório precisará realizar autoavaliações e um ciclo regular de auditoria e/ou certificação permitindo que sejam auditadas e medidas suas políticas, práticas e desempenho. Avaliações e autoavaliações podem ser uma contribuição importante para garantir a confiabilidade e a durabilidade dos repositórios de dados de pesquisa.

Assim sendo, esta pesquisa foi motivada inicialmente pelo interesse da autora em assuntos relacionados a dados de pesquisa. Ainda é fundamental destacar que, dentre a literatura existente, houve a percepção da falta de um estudo específico que abordasse esses novos padrões em um contexto brasileiro.

Nesse sentido, o objetivo geral estabelecido como norteador da pesquisa foi verificar quais critérios/requisitos são pertinentes em uma avaliação de confiabilidade de repositórios de dados de pesquisa em um contexto brasileiro. No intuito de atender a esse objetivo principal, os objetivos específicos foram elaborados a partir de um levantamento, em âmbito internacional, dos instrumentos de avaliação existentes para repositórios de dados de pesquisa confiáveis.

Buscou-se compreender, através de um estudo bibliográfico, sobre o processo de avaliações e autoavaliações e constatou-se diversos benefícios, como por exemplo, a contribuição para o desenvolvimento de repositórios de dados de pesquisa a partir da construção de uma base organizacional sólida para submeter-se futuramente às certificações e auditorias e consequentemente obter selos de repositórios confiáveis e de qualidade. O estudo se fundamenta em uma metodologia de pesquisa documental realizado com base em quatro instrumentos selecionados: *CoreTrustSeal Trustworthy Data Repositories Requirements*; ACTDR - *Audit and Certification of Trustworthy Digital Repositories; Principles* FAIR; e *The TRUST Principles for digital repositories*. Os instrumentos foram escolhidos por serem considerados conceituados internacionalmente.

A tese apresentada postula que um conjunto de critérios/requisitos de avaliação da confiabilidade para ser utilizado por repositórios de dados de pesquisa brasileiros pode auxiliar no desenvolvimento de repositórios confiáveis e de qualidade. A fim de demonstrar isso empiricamente, e com base na literatura sobre o tema, foi criado um conjunto de critérios/requisitos para auxiliar nas autoavaliações da confiabilidade de repositórios de dados de pesquisa brasileiros. O conjunto desenvolvido poderá auxiliar os repositórios brasileiros na realização de autoavaliações e se prepararem para futuras avaliações e certificações. Ademais, o estabelecimento de diretrizes nesta área pode auxiliar no treinamento de profissionais envolvidos em avaliação de repositórios de dados de pesquisa e direcioná-los no desenvolvimento e implementação de novos repositórios.

O conjunto de critérios/requisitos elaborados, a partir da análise dos instrumentos internacionais, possibilita uma autoavaliação das áreas/infraestruturas de um repositório de dados de pesquisa que divide-se em três: infraestrutura organizacional/transparência; gerenciamento de objetos digitais; e tecnologia/gestão de risco de infraestrutura e segurança/*sustainability*. As estratégias elaboradas possuem relação com os principais critérios/princípios inseridos nos instrumentos selecionados. O conjunto foi proposto a partir da confrontação dos requisitos/princípios, ou seja, as estratégias a serem seguidas baseou-se na existência de similaridade.

Esta pesquisa não tem como propósito a substituição das leituras dos documentos publicados pelas organizações, simplesmente atuará de modo complementar para facilitar a compreensão e despertar a reflexão auxiliando nos processos de planejamento e

implementação tendo em vista o preparo para futura certificação. Mesmo sem a pretensão de uma certificação formal, os requisitos contemplados no estudo podem ser usados como referência e para identificar as lacunas e áreas que precisam de maior atenção.

Com a análise dos 4 documentos selecionados nesta pesquisa foi possível elaborar um conjunto de critérios/requisitos para orientação à implantação da confiabilidade em repositórios de dados de pesquisa brasileiros.

Conforme citado, os dados mostraram similaridades entre alguns requisitos: o primeiro requisito do *CoreTrustSeal* abrange a missão e o escopo. Nota-se que o instrumento ACTDR também possui essa subdivisão e, dentro da infraestrutura organizacional, trata sobre a governança e a viabilidade organizacional. Ambos abordam a declaração da missão como requisito e garantia da preservação. Da mesma forma, os princípios TRUST apresentam em seu primeiro requisito '*transparency*' a mesma abordagem. Estes requisitos semelhantes relatam que o repositório é responsável pela administração dos objetos digitais e por garantir que os materiais sejam mantidos no ambiente apropriado por períodos de tempo. Em conformidade com este princípio, os repositórios devem garantir que, no mínimo, a declaração de missão e o escopo do repositório sejam claramente indicados.

O segundo requisito do *CoreTrustSeal* refere-se as 'licenças', equivalente ao requisito 'contratos, licenças e responsabilidades' do instrumento ACTDR. Ambos requisitos assemelham-se ao requisito '*Responsibility*' dos princípios TRUST. O Princípio FAIR R1.1 também possui relação com estes requisitos. Este requisito refere-se aos regulamentos de acesso e licenças aplicáveis estabelecidas pelo próprio repositório de dados, bem como quaisquer códigos de conduta que são geralmente aceitos no setor relevante para o intercâmbio e uso adequado de conhecimento e informação. Este requisito é importante para garantir que o repositório tenha os direitos e autorizações necessários para permitir que ele colete e preserve conteúdo digital ao longo do tempo, faça com que essas informações estejam disponíveis para sua comunidade designada, e para defender esses direitos quando questionados.

O terceiro requisito do *CoreTrustSeal*, 'continuidade de acesso', se assemelha ao requisito do ACTDR 'governança e a viabilidade organizacional' destacando que o repositório deve ter um Plano Estratégico de Preservação que define a abordagem que o repositório terá no suporte de longo prazo. Os princípios TRUST tratam sobre esta questão

através do princípio '*Sustainability*'. O requisito do *CoreTrustSeal* cobre a governança relacionada à operação contínua do repositório ao longo do tempo e durante desastres, bem como evidências em relação ao planejamento de sucessão, designadamente, as medidas em vigor para garantir o acesso e a disponibilidade dos acervos de dados, atualmente e no futuro.

O quarto requisito do *CoreTrustSeal*, 'Confidencialidade/Ética', menciona que o repositório deve garantir, na medida do possível, que os dados sejam criados, com curadoria, acessados e usados em conformidade com as normas disciplinares e éticas. Da mesma forma, o requisito 'sustentabilidade financeira" do ACTDR, trata em parte o assunto confidencialidade. Os princípios TRUST, no item '*Responsibility*' aborda as informações confidenciais. O instrumento *CoreTrustSeal* relata que este requisito se refere às disposições éticas e de privacidade que afetam a criação, curadoria e uso dos dados. Desta forma, o repositório deve demonstrar que tem boas práticas para dados que possuem riscos de divulgação e deve incluir orientações para os depositantes e usuários. Esse procedimento é necessário para manter a confiança daqueles que concordam em ter dados pessoais/confidenciais armazenados no repositório. Contudo, o ACTDR menciona que é necessário se proteger contra má conduta ou outra atividade desagradável que possa ameaçar a viabilidade econômica do repositório. O repositório não pode apenas reivindicar transparência, mas deve mostrar que ajusta seus negócios para mantê-los transparentes, compatíveis e auditáveis. Os princípios TRUST mencionam que a responsabilidade é demonstrada com o gerenciamento dos direitos de propriedade intelectual dos produtores de dados, proteção de recursos de informações confidenciais, e a segurança do sistema e seu conteúdo. Os usuários do repositório devem ter certeza de que os depositantes de dados são solicitados a fornecer todos os metadados compatíveis com as normas da comunidade, o que aumenta a descoberta e a utilização dos dados.

O quinto requisito do *CoreTrustSeal,* 'Infraestrutura organizacional', menciona que o repositório deve ter financiamento adequado e número suficiente de funcionários qualificados gerenciados por meio de um sistema claro de governança para realizar a missão com eficácia. O ACTDR possui um requisito semelhante, 'estrutura organizacional e pessoal', e relata que o repositório deve ter identificado e estabelecido as funções de que necessita para executar e deve ter nomeado pessoal com habilidades e experiência adequadas para cumprir essas

funções. Os princípios TRUST relatam que um repositório depende da interação das pessoas, processos e tecnologias para oferecer suporte seguro, persistente e serviços confiáveis.

O sexto requisito do *CoreTrustSeal*, 'Orientação de especialista' assemelha-se a um dos requisitos do ACTDR, 'contratos, licenças e responsabilidades'. Segundo o instrumento do *CoreTrustSeal,* um repositório eficaz se esforça para realizar evoluções e adotar novas tecnologias mais eficazes a fim de permanecer valioso para sua comunidade designada. Devido ao ritmo rápido de mudanças é aconselhável que o repositório possua um aconselhamento e feedback de especialistas para garantir sua relevância e melhorias contínuas. Da mesma forma, o ACTDR orienta que as Políticas de Preservação e Planos de Implementação de Preservação do repositório e mecanismos devem ser examinados por autoridades institucionais apropriadas e/ou especialistas jurídicos para garantir que as respostas aos desafios cumpram as leis e requisitos relevantes. Os princípios TRUST não mencionam este requisito.

O sétimo requisito do *CoreTrustSeal* corresponde a 'Integridade e autenticidade dos dados'. O instrumento ACTDR possui alguns requisitos que assemelham-se ao do *CoreTrustSeal*: responsabilidade processual e política de preservação (o repositório deve definir, coletar, rastrear e fornecer adequadamente suas medições de integridade da informação; o repositório deve ter um processo de *ingest* que verifica cada SIP para integridade e correção; o repositório deve obter controle suficiente sobre os objetos digitais para preservá-los; o repositório deve verificar cada AIP quanto à integridade e exatidão; e o repositório deve fornecer um mecanismo independente para verificar a integridade da coleção/conteúdo do repositório). O requisito preservação de AIP também trata sobre integridade: (o repositório deve ter especificações de como os AIPs são armazenados até o nível de bits). No requisito 'gestão de informações' também há especificações referentes a integridade. Na subdivisão: 'Gestão de risco de infraestrutura e segurança', requisito 'gestão de risco de infraestrutura técnica' do ACTDR, informa que o repositório deve identificar e gerenciar os riscos às suas operações de preservação e objetivos associados à infraestrutura do sistema como também, gerenciar o número e localização de cópias de todos os objetos digitais. O Princípio FAIR R1.2 possui relação com estes requisitos. Nos princípios TRUST não foram encontrados requisitos relacionados a integridade e autenticidade.

O oitavo requisito do *CoreTrustSeal*, ‘Avaliação’, orienta o repositório a aceitar dados e metadados com base em critérios definidos para garantir relevância e compreensibilidade para os usuários de dados. O ACTDR possui alguns requisitos relacionados à avaliação: ‘O repositório deve se comprometer com um cronograma regular de autoavaliação e certificação externa’; ‘O repositório deve fornecer evidências da eficácia de sua preservação’; ‘O repositório deve manter uma análise sistemática dos fatores de risco de segurança associados a dados, sistemas, pessoal e planta física’; ‘O repositório deve ter uma Política de Coleta ou outro documento que especifique o tipo de informação que irá preservar, reter, gerenciar e fornecer acesso. Após análise dos princípios TRUST percebe-se que não há abordagem deste requisito. De acordo com o *CoreTrustSeal,* a função da avaliação é fundamental para avaliar se os dados atendem a todos os critérios de seleção e para garantir o manejo adequado para a sua preservação. A avaliação e a reavaliação ao longo do tempo garantem que os dados permaneçam relevantes e compreensíveis para a comunidade designada. Contudo, o ACTDR menciona que o repositório deve ser capaz de demonstrar a preservação contínua, incluindo compreensibilidade de suas participações. Isso pode ser avaliado em vários graus e depende da especificidade da comunidade designada.

O nono requisito do *CoreTrustSeal* trata sobre ‘Procedimentos de armazenamento documentados’, ou seja, o repositório aplica processos e procedimentos documentados durante o gerenciamento do armazenamento de arquivos de dados. O instrumento ACTDR possui requisito semelhante: ‘o repositório deve ter processos documentados para aquisição de PDI’. Os princípios TRUST não abordam este requisito. De acordo com o *CoreTrustSeal*, os repositórios precisam armazenar dados e metadados desde o ponto de depósito até o ponto de acesso. O ACTDR orienta que o repositório deve executar seus processos documentados para aquisição de PDI como também deve garantir que o PDI esteja persistentemente associado às informações de conteúdo relevantes.

O décimo requisito do *CoreTrustSeal,* ‘Plano de preservação’, diz que o repositório assume a responsabilidade pela preservação a longo prazo e gerencia essa função de forma planejada e documentada. O ACTDR possui entre seus requisitos, o seguinte critério: O repositório deve ter estratégias de preservação documentadas relevantes. O Princípio FAIR A2, que trata sobre estratégias de preservação e possui relação com estes requisitos. De acordo com o *CoreTrustSeal,* o repositório, os depositantes de dados e a comunidade

designada precisam entender o nível de responsabilidade assumido para cada item depositado no repositório. O repositório deve ter direitos para assumir essas responsabilidades, ou seja, os procedimentos devem ser documentados e sua conclusão assegurada. Dito isto, o ACTDR menciona que isso é necessário para que fique claro como o repositório planeja garantir que a informação permanecerá disponível e utilizável para as gerações futuras e para fornecer um meio de verificar e validar o trabalho de preservação do repositório. Estas estratégias de preservação documentadas irão descrever como o repositório irá agir sobre riscos identificados, como parte do plano estratégico de preservação. Essas estratégias de preservação e o plano estratégico de preservação normalmente abordará a degradação da mídia de armazenamento, a obsolescência das unidades de mídia e a obsolescência ou inadequação da representação das informações (incluindo formatos) como a base de conhecimento da comunidade designada. O Princípio FAIR A2 menciona que quando necessário, dependendo das restrições ao acesso aos dados e/ou metadados, um mecanismo para autenticação e autorização para o acesso deve ser permitido pelo protocolo de comunicação.

O requisito 11 do *CoreTrustSeal*, 'Qualidade dos dados', refere-se a experiência apropriada para lidar com dados técnicos e qualidade de metadados e a garantia que informações suficientes estejam disponíveis para os usuários finais fazerem avaliações relacionadas à qualidade. O requisito do ACTDR, 'gestão de risco de segurança' orienta que deve-se manter uma análise sistemática dos fatores de risco de segurança associado a dados, sistemas, pessoal e planta física descrevendo alguns aspectos relacionados a qualidade. Os princípios TRUST aborda este requisito qualidade no item '*Responsibility*', informando que a responsabilidade é demonstrada aderindo aos padrões de metadados e curadoria, juntamente com o gerenciamento dos acervos de dados, por exemplo, validação técnica, documentação e controle de qualidade. Os Princípios FAIR I3 e R1 estão implícitos neste requisito. De acordo com o *CoreTrustSeal*, os repositórios devem garantir que haja informações suficientes sobre os dados para que a comunidade designada avalie a qualidade dos dados. A avaliação da qualidade torna-se cada vez mais relevante quando a comunidade designada é multidisciplinar, onde os usuários podem não ter experiência pessoal para fazer uma avaliação da qualidade apenas a partir dos dados. Os repositórios devem ser capazes de avaliar a integridade e qualidade dos dados e metadados. O ACTDR orienta que um curso interno e

avaliação externa deve ser realizada para avaliar a qualidade do serviço e a relevância para o usuário. Direito de propriedade intelectual também deve ser revisado regularmente, bem como a responsabilidade do repositório para regulamentar a não conformidade. O repositório deve avaliar as habilidades de sua equipe e garantir a aquisição de novos funcionários ou reciclagem do pessoal existente, conforme necessário. A avaliação de risco regular também deve abordar ameaças e ataques de negação de serviço e perda ou qualidade inaceitável de serviços terceirizados O repositório pode realizar avaliações de risco geral. Os Princípios FAIR I3 menciona que (Meta) dados devem incluir referências qualificadas para outros (Meta) dados, ou seja, deve-se referenciar os conjuntos de dados devidamente, possibilitando que conjuntos de dados gerados, a partir de outros conjuntos de dados, sejam ligados. O R1 estabelece que os (Meta) dados devem ser descritos com uma pluralidade de atributos precisos e relevantes.

O requisito 12 do *CoreTrustSeal* corresponde a ‘Fluxos de trabalho’, ou seja, o arquivamento deve ocorrer de acordo com fluxos de trabalho definidos, desde o armazenamento até a disseminação. No instrumento ACTDR, questões relacionadas a fluxos de trabalho, estão presentes em 3 requisitos: o repositório deve ter ‘Políticas de Preservação em vigor para garantir que o seu Plano Estratégico de Preservação seja cumprido’; ‘o repositório deve ter um processo de *ingest* que verifica cada SIP para integridade e correção’; e ‘o repositório deve gerenciar o número e localização de cópias de todos os objetos digitais’. Não foram identificadas questões relacionadas a fluxo de trabalho nos princípios TRUST e FAIR. De acordo com o instrumento *CoreTrustSeal*, para garantir a consistência das práticas entre conjunto de dados e serviços, os fluxos de trabalho devem ser definidos de acordo com as atividades do repositório e claramente documentados. O ACTDR indica utilizar fluxos de trabalho nos mecanismos para revisão, nas atualizações e desenvolvimento de suas políticas de preservação à medida que o repositório cresce e à medida que a tecnologia e a prática da comunidade evolui. É necessário, para que o repositório tenha políticas completas e atualizadas, procedimentos em vigor que refletem os requisitos e práticas de preservação para sua comunidade.

O Requisito 13 do *CoreTrustSeal*, ‘Descoberta e identificação de dados’, orienta que o repositório deve permitir que os usuários descubram os dados e os consultem de forma persistente por meio de citações adequadas. O ACTDR possui um requisito semelhante ao

CoreTrustSeal: ‘Gestão de informações’, onde o repositório deve especificar os requisitos mínimos de informação para permitir a comunidade designada a descobrir e identificar materiais de interesse. Os princípios FAIR (F1, F2, F3, F4) estão de certa forma, implícitos neste requisito. Segundo o *CoreTrustSeal*, a descoberta de dados eficaz é a chave para o compartilhamento de dados. Uma vez descobertos, os conjuntos de dados devem ser referenciados por meio de citações completas, incluindo identificadores persistentes para ajudar a garantir que os dados possam ser acessados no futuro. Da mesma forma, o ACTDR, menciona que este requisito é necessário para permitir a descoberta dos acervos do repositório. Pode-se atingir esse requisito através das informações descritivas de recuperação, metadados de descoberta, como Dublin Core e outras documentações que descrevem o objeto. O repositório deve ser capaz de lidar com os tipos de solicitações que virão de um típico usuário da comunidade designada. Os Princípios FAIR F1, F2, F2 e F4 mencionam que os (meta)dados devem ter identificadores globais, persistentes e identificáveis como, por exemplo, DOI, ARK entre outros e ser descritos com metadados ricos o suficiente para que, uma vez indexados, possam ajudar o usuário a encontrar os dados mesmo que não haja um identificador. Os metadados devem incluir claramente e explicitamente os identificadores dos dados que descrevem, ou seja, como não podemos prever que os dados e seus metadados estejam sempre juntos, por exemplo, quando os metadados são indexados por um mecanismo de busca e, portanto, está em uma plataforma diferente dos dados, a associação entre eles deve ser feita pela inclusão do identificador dos dados pelos metadados. Para que os dados sejam encontrados, seus metadados devem ser indexados por mecanismos de busca que, por sua vez, permitem aos usuários encontrá-los por meio de elementos desses metadados.

O requisito 14 do *CoreTrustSeal*, ‘Reutilização de dados’, menciona que o repositório deve permitir a reutilização dos dados ao longo do tempo, garantindo que os metadados apropriados estejam disponíveis para apoiar a compreensão e o uso dos dados. Da mesma forma, o requisito ‘O repositório deve rastrear e gerenciar os direitos de propriedade intelectual e restrições ao uso do conteúdo do repositório, conforme exigido pelo acordo de depósito, contrato ou licença’, do ACTDR também trata sobre a questão da reutilização de 149 dados. Observou-se que os princípios TRUST, da mesma forma que o *CoreTrustSeal* e o ACTDR, também aborda a mesma questão através do critério ‘*User focus* (Foco no usuário)’. O *CoreTrustSeal* orienta que os repositórios devem garantir que os dados continuem a ser

compreendidos e usados de forma eficaz no futuro, apesar das mudanças na tecnologia e na base de conhecimento da comunidade designada. Este requisito avalia as medidas tomadas para garantir que os dados sejam reutilizáveis. O ACTDR relata que isso é necessário para permitir que o repositório rastreie, atue e verifique os direitos e restrições relacionadas ao uso dos objetos digitais dentro do repositório. Os Princípios TRUST mencionam que os repositórios têm um papel vital na aplicação e reforço das normas e padrões, o que inclui os esquemas de metadados, formatos de arquivos de dados, vocabulários controlados, ontologias e outras semânticas. Também mencionam que um repositório confiável precisa se concentrar em servir sua comunidade de usuários-alvo. Cada comunidade de usuários provavelmente tem expectativas diferentes em relação aos seus repositórios.

O requisito 15 do *CoreTrustSeal* corresponde a ‘Infraestrutura técnica’ e orienta que o repositório funcione em sistemas operacionais bem suportados e em outro software de infraestrutura central, e use tecnologias de hardware e software apropriadas para os serviços que fornece à sua comunidade designada. O ACTDR possui uma subdivisão ‘Gestão de risco de infraestrutura e segurança’ e o requisito ‘gestão de risco de infraestrutura técnica’, indicando que o repositório deve identificar e gerenciar os riscos às suas operações de preservação e objetivos associados à infraestrutura do sistema. Os Princípios FAIR A1, A1.1, I1, I2 e R1.3 possuem relação com este requisito e tratam sobre os dados serem acessíveis. Não foram encontrados estes requisitos entre os princípios TRUST. O *CoreTrustSeal* orienta que os repositórios precisam operar em infraestruturas centrais confiáveis e estáveis que maximizam a disponibilidade do serviço. Ademais, o hardware e o software usados devem ser relevantes e apropriados para a comunidade e para as funções que o repositório desempenha. De acordo com ACTDR este requisito é necessário para garantir uma infraestrutura segura e confiável. Os Princípios FAIR A1, A1.1 mencionam que os (Meta) dados devem ser recuperáveis pelos seus identificadores usando um protocolo de comunicação padronizado deve ser aberto, gratuito e universalmente implementável. Já os Princípios FAIR I1 e I2 mencionam que os (Meta) dados devem ser representados por meio de uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento. Como os dados e metadados devem possuir referências aos vocabulários que contenham os conceitos utilizados, devemos garantir que sejam utilizados vocabulários que também sigam os princípios FAIR. Já o Princípio R1.3, diz que os (Meta) dados devem estar alinhados com

padrões relevantes ao seu domínio, ou seja, deve atender os padrões específicos da comunidade da área e às boas práticas de arquivamento e ao compartilhamento do campo de pesquisa específico.

O requisito 16 do *CoreTrustSeal* refere-se a 'Segurança', ou seja, a infraestrutura técnica do repositório deve oferecer proteção para as instalações e seus dados, produtos, serviços e usuários. O instrumento ACTDR possui alguns requisitos relacionados à segurança: Gestão de acesso (o repositório deve obedecer às Políticas de Acesso). Na subdivisão 'Gestão de risco de infraestrutura e segurança', no requisito 'gestão de risco de infraestrutura técnica' (o repositório deve identificar e gerenciar os riscos às suas operações de preservação e objetivos associados à infraestrutura do sistema). E no requisito 'gestão de risco de segurança' (o repositório deve manter uma análise sistemática dos fatores de risco de segurança associado a dados, sistemas, pessoal e planta física). Entre os princípios TRUST, nos itens *Responsibility* e *Technology* foram encontradas menções sobre segurança. O Princípio FAIR A 1.2 está implícito neste requisito. Conforme consta no *CoreTrustSeal*, o repositório deve analisar ameaças potenciais, avaliar riscos e criar um sistema de segurança consistente. Conforme os requisitos citados e encontrados no ACTDR, o termo "acesso" tem vários sentidos diferentes, incluindo o acesso dos usuários ao sistema de repositório, por exemplo, segurança física e autenticação de usuário, e os diferentes estágios de acesso aos registros (fazer uma solicitação, verificar os direitos do solicitante e preparar e enviar um Pacote de Informação de Disseminação DIP). Esta subseção é dividida em dois requisitos principais, um relacionado com o existência e implementação de políticas de acesso, e uma com capacidade de repositório para fornecer objetos comprovadamente autênticos como DIPs. Assim, o primeiro requisito se refere a solicitações iniciadas por um usuário e como o repositório as trata para garantir que os direitos e acordos sejam respeitados, que a segurança seja monitorada, que as solicitações sejam atendidas. O segundo requisito se relaciona com o que é entregue ao consumidor e a confiança que pode ser colocada nele. Entre os princípios TRUST existem especificações para gerenciar os direitos de propriedade intelectual dos produtores de dados, proteção de recursos de informações confidenciais, e a segurança do sistema e seu conteúdo; como também ter planos e mecanismos em vigor para prevenir, detectar e responder à ameaças à segurança física ou cibernética. De acordo com o Princípio FAIR A 1.2, quando necessário, dependendo das restrições ao acesso aos dados e/ou metadados, um

mecanismo para autenticação e autorização para o acesso deve ser permitido pelo protocolo de comunicação.

Foi observado que alguns requisitos que constavam no ACTDR não estavam contemplados de forma clara nos instrumentos *CoreTrustSeal*, TRUST e FAIR. Comparando os instrumentos relacionados nesta pesquisa nota-se que o ACTDR apresenta seus requisitos de forma mais extensa e detalhada, o que pode dificultar a compreensão e aplicação. Assim sendo, esta pesquisa forneceu *insights* sobre questões metodológicas e evidenciou a necessidade de ampliação das investigações sobre o tema para promover e consolidar iniciativas no Brasil.

Como proposições para novas pesquisas, sugere-se o desenvolvimento de um instrumento de autoavaliação da confiabilidade adaptado ao contexto brasileiro para ser utilizado por repositórios de dados de pesquisa. Destaca-se também a necessidade e que enriqueceria a pesquisa, utilizar-se do conhecimento e da experiência de especialistas e de suas opiniões que auxiliariam no desenvolvimento e na verificação da importância da aplicação do instrumento.

REFERÊNCIAS

ALISSON, Elton. Repositórios científicos buscam certificação. Agência FAPESP, 2021. Disponível em: **https://agencia.fapesp.br/repositorios-cientificos-buscam-certificacao/35313/**. Acesso em: 5 abr. 2021.

ARELLANO, Miguel Ángel Márdero. **Auditoria e certificação de repositórios digitais.** 2017. Rede Cariniana – IBICT. Disponível em: http://www.cnen.gov.br/images/CIN/PDFs/Auditoria_e_certificacao_repositorios.pdf. Acesso em: 18 set 2020.

ASSIS, Tainá Batista de. Rede brasileira de repositórios e o impacto dos trabalhos das subredes. I Encontro Rede Sudeste de Repositórios Institucionais. **Anais.** 2019. Disponível em:https://www.arca.fiocruz.br/bitstream/icict/33642/2/Anais_I_Encontro_Sudeste_RIAA_2019.pdf. Acesso em: 13 set. 2020.

AUSTIN, Claire C. *et al.* Research data repositories: review of current features, gap analysis, and recommendations for minimum requirements. **IASSIST**, 2015. Disponível em: https://www.researchgate.net/publication/280303920_Research_Data_Repositories_Review_of_Current_Features_Gap_Analysis_and_Recommendations_for_Minimum_Requirements. Acesso em: 18 out. 2020.

AVENTURIER, Pascal. **Princípios FAIR:** critérios de qualidade para dados de pesquisa. A publicação científica, 2017a. Disponível em: https://publicient.hypotheses.org/1456. Acesso em: 25 jul. 2020.

AVENTURIER, Pascal. Reporte do CWTS e do Elsevier sobre dados abertos e perspectivas do pesquisador. 2017b. **A publicação científica**. Disponível em: https://publicient.hypotheses.org/1798. Acesso em: 27 maio. 2020.

AVENTURIER, Pascal. ALENCAR, Maria de Cléofas Faggion. Os desafios dos dados de pesquisa abertos. 2016.**RECIIS – Revista Eletronica de Comunicação, Informação e Inovação em Saúde.** Disponível em:https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1069/pdf1069. Acesso em: 23 dez. 2019.

BARZELAY, Michael. Instituições centrais de auditoria e auditoria de desempenho: uma análise comparativa das estratégias organizacionais na OCDE. **Revista Do Serviço Público**, 2014. Disponível em: https://doi.org/10.21874/rsp.v53i2.283. Acesso em 2 ago. 2020.

BERNERS-LEE, Tim. **Linked Data.** W3C, 2006. Disponível em: https://www.w3.org/DesignIssues/LinkedData.html. Acesso em: 2 mar. 2020.

BERNERS-LEE, Tim. I Invented the World Wide Web. Here's How We Can Fix It. **The New Work Times**, 2019. Disponível em:

https://www.nytimes.com/2019/11/24/opinion/world-wide-web.html. Acesso em 10 jun. 2020.

BOURDIEU, Pierre. **A miséria do mundo**. Tradução de Mateus S. Soares. Petrópolis: Vozes, 1999.

CAMPÊLO, Leonard Richard Rodrigues Rufino; BARRETO NETO, Vanderlino Coelho. Comparando softwares gratuitos para criação de repositórios de dados abertos. **Ciência da Informação,** v.48, n.3, p. 341-346, 2019. Disponível em: http://revista.ibict.br/ciinf/article/view/5004. Acesso em: 6 maio. 2020.

CARVALHO, José *et al*. Auditoria ISO 16363 a repositórios institucionais. **Cadernos BAD**, 2014. Disponível em: http://repositorium.sdum.uminho.pt/bitstream/1822/30499/4/ConfOA2014_Auditoria%20ISO%2016363%20-%20RCAAP%20-%20Cadernos%20BAD-FINAL.pdf. Acesso em: 23 jul. 2020.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO – CNPq. **CNPq e IBICT firmam acordo para implementação de repositório de dados científicos.** 2019. Disponível em: http://portal.abipti.org.br/cnpq-e-ibict-firmam-acordo-para-implementacao-de-repositorio-de-dados-cientificos/. Acesso em: 22 abr. 2020.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO – CNPq. **Plano de Dados Abertos CNPq**, data Disponível em: http://ftp.cnpq.br/pub/CKAN/Plano_de_Dados_Abertos_CNPq_2019.pdf. Acesso em: 5 jul. 2020.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS- CCSDS. **Audit and Certification of Trustworthy Digital Repositories - ACTDR**. Magenta Book, 2011.

CONTANDRIOPOULOS, André-Pierre *et al*. A avaliação na área da saúde: conceitos e métodos; HARTZ, Zulmira Maria de Araújo, org. **Avaliação em Saúde:** dos modelos conceituais à prática na análise da implantação de programas [online]. Rio de Janeiro: Editora FIOCRUZ, 1997. Disponível em:https://static.scielo.org/scielobooks/3zcft/pdf/hartz-9788575414033.pdf#page=26. Acesso em: 25 jul. 2020.

CORETRUSTSEAL. **The World Data System of the International Science Council (WDS) and the Data Seal of Approval (DSA) are pleased to announce the launch of a new certification organization:** CoreTrustSeal. 2020a. Disponível em: https://www.coretrustseal.org/about/. Acesso em: 26 abr. 2020.

CORETRUSTSEAL. **Coretrustseal Trustworthy Data Repositories Requirements 2020-2022**. 2020b. Disponível em: https://www.coretrustseal.org/wp-content/uploads/2017/01/Core_Trustworthy_Data_Repositories_Requirements_01_00.pdf. Acesso em: 17 jun. 2020.

CORETRUSTSEAL. **Update: Draft Requirements 2023-2025.** Disponível em: https://www.coretrustseal.org/why-certification/meeting-community-needs/trustworthy-data-repository-requirements-review-2023-2025/. Acesso em: 16 jul. 2022.

COSTA, Michele Pereira da. **Fatores que Influenciam a Comunicação de Dados de Pesquisa sobre o Vírus da Zika na Perspectiva de Pesquisadores**. 2017. Tese (doutorado) - Universidade de Brasília, Brasília, 2017. Disponível em: https://repositorio.unb.br/handle/10482/32428. Acesso em 12 dez. 2019.

CURTY, Renata. Para onde os dados devem ir afinal? **Dados de pesquisa abertos.** 2018. Disponível em: https://dadosdepesquisa.rnp.br/para-onde-os-dados-devem-ir-afinal/. Acesso em: 9 jan. 2020.

DAWEI LIN, Jonathan Crabtree *et al*. **The TRUST Principles for digital repositories**. Scientific Data, 2020. Disponível em: https://www.nature.com/articles/s41597-020-0486-7. Acesso em: 26 jul. 2020.

DIGITAL SCIENCE REPORT. **The state of Open Data 2018**. 2018. Disponível em: https://www.digital-science.com/resources/portfolio-reports/state-open-data-2018/. Acesso em: 23 jan. 2021.

DILLO, Ingrid. **CoreTrustSeal for Trustworthy Data Repositories.** Word Data System. 2018. Disponível em: https://www.worlddatasystem.org/community/wds-members-forum/data-repositories-day-2018/2018-presentations/11_CTS_certification. Acesso em: 13 abr. 2021.

DONALDSON, Devan Ray. **Certification information on trustworthy digital repository websites:** A content analysis. PLOS ONE, 2020. Disponível em: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0242525. Acesso em 5 fev. 2021.

DONALDSON, Devan Ray *et al*. **The Perceived Value of Acquiring Data Seals of Approval.** International Journal of Digital Curation, 2017. Disponível em: https://core.ac.uk/download/pdf/192905532.pdf. Acesso em: 27 abr. 2021.

DOWNS, Robert R. **Improving Opportunities for New Value of Open Data:** Assessing and Certifying Research Data Repositories. Data Science Journal, 2021. Disponível em: https://datascience.codata.org/article/10.5334/dsj-2021-001/. Acesso em: 4 fev. 2021.

DOWNS, Robert R. International Standards for Trustworthy Data Repositories. National Institutes of Health (NIH) **Trustworthy Data Repositories Workshop**. National Institute for Allergy and Infectious Diseases (NIAID). Rockville, 2019. Disponível em: https://academiccommons.columbia.edu/doi/10.7916/d8-h7xf-ha70. Acesso em: 4 fev. 2021.

DUDZIAK, Elisabeth. Dados de Pesquisa agora devem ser armazenados e citados. **SIBIUSP**, 2016. Disponível em: http://www.sibi.usp.br/noticias/dados-materiais-metodos-revistas-exigem-dados-pesquisa-estejam-disponiveis/. Acesso em: 5 jan. 2020.

EDMUNDS, Rorie. **Second Latin America and Caribbean Scientific Data Management Workshop**, 2021. Disponível em: https://fapesp.br/eventos/wds. Acesso em: 3 maio. 2021.
EUROPEAN COMMISSION**. Guidelines on open access to scientific publications and research data in Horizon 2020**, 2014. Disponível em: https://www.openaire.eu/guidelines-on-open-access-to-scientific-publications-and-research-data-in-horizon-2020. Acesso em: 16 jan. 2020.

EUROPEAN COMMISSION. **H2020 Programme:** Guidelines on FAIR Data Management in Horizon 2020. Disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. Acesso em: 16 mai. 2020.

FAPESP. Open data under the COVID-19 pandemic. **Research Webinars**, 2020. Disponível em: https://fapesp.br/eventos/covid4. Acesso em: 9 mar. 2021.

FORCE11. ***The FAIR data principles***. 2016. Disponível em: https://www.force11.org/group/fairgroup/fairprinciples. Acesso em: 18 jan. 2020.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO - FAPESP. **Política para Propriedade Intelectual da FAPESP**. 2011. Disponível em: http://www.fapesp.br/6594#2. Acesso em: 18 mar. 2020.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO - FAPESP. Planos de gestão de dados se incorporam a projetos de pesquisa no Brasil. **Revista FAPESP.** 2017. Disponível em: https://revistapesquisa.fapesp.br/planos-de-gestao-de-dados-se-incorporam-a-projetos-de-pesquisa-no-brasil/. Acesso em: 3 jul. 2020.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO - FAPESP. **Programa de pesquisa em eScience**. 2018. Disponível em: www.fapesp.br. Acesso em: 21 jan. 2020.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO FAPESP. **Plano de gestão de dados** – FAPESP. 2019. Disponível em:http://www.fapesp.br/gestaodedados/. Acesso em: 22 jan. 2020.

GABRIEL JUNIOR Rene Faustino *et al*. Acesso aberto a dados de pesquisa no Brasil: mapeamento de repositórios, práticas e percepções dos pesquisadores e tecnologias. **Ciência da Informação**, 2019. Disponível em: https://lume.ufrgs.br/bitstream/handle/10183/212266/001115316.pdf?sequence=1&isAllowed=y. Acesso em: 28 jan. 2021.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social.** 5.ed. São Paulo: Atlas, 1999.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa.** 5. ed. São Paulo: Atlas, 2010.

GO FAIR BRASIL SAÚDE. **I Congresso Internacional da Escola de Enfermagem Alfredo Pinto**. 2020. Disponível em: https://portal.fiocruz.br/noticia/seminario-virtual-marca-o-lancamento-da-rede-go-fair-brasil-saude-enfermagem. Acesso em: 18 set. 2020.

GUEDES, Renan Mastrange *et al*. **Maturidade de gestão de projetos de sistemas de informação:** um estudo exploratório quantitativo no Brasil. Production, 2014. Disponível em: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-65132014000200010&lng=pt&tlng=pt. Acesso em: 24 set. 2020.

HENNING, Patricia Corrêa *et al*. GOFAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. **Em Questão,** 2019. Disponível em: https://seer.ufrgs.br/EmQuestao/article/view/84753. Acesso em: 19 jun. 2020.

HENNING, Patricia Corrêa; MOREIRA, J. Ciência aberta, dados abertos e princípios FAIR: uma contribuição dos Países Baixos. *In*: SHINTAKU, Milton; SALES, Luana Farias; COSTA, Michelli. (org).**Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. Disponível em: https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf. Acesso em: 13 abr. 2020.

HENNING, Patricia et al. **Desmistificando os princípios fair: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados fair.** Tendências da Pesquisa Brasileira e Ciência da Informação, ANCIB, v. 11, n. 1. 2018. Disponível em: https://revistas.ancib.org. Acesso em: 18 dez. 2021.

HODSON, S. et al. **Turning FAIR into reality**. Interim report of the European Commission Expert Group on FAIR data. 2018. Disponível em: https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_1.pdf. .Acesso em: 31, jul. 2018.

INGRAM, Caroline. **How and why you should manage your research data: a guide for researchers.** JISC, 2016. Disponível em: https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data. Acesso em: 20 jan. 2020.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA – IBICT. **Plano de dados abertos, 2020-2021**, 2020. Disponível em: http://www.ibict.br/images/conteudo/PDA_2020___2021.pdf. Acesso em: 23 jan. 2020.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA – IBICT. **Instituições comemoram o cumprimento do 3º compromisso do 4º Plano de Ação Nacional para Governo Aberto,** 2020. Disponível em: https://ibict.br/sala-de-

imprensa/noticias/item/2322-instituicoes-comemoram-o-cumprimento-do-3-compromisso-do-4-plano-de-acao-nacional-para-governo-aberto. Acesso em: 3 fev. 2021.

JANTZ, Ronald; GIARLO, Mike. **Digital Preservation:** Architecture and Technology for Trusted Digital Repositories. 2006. Disponível em: http://www.dlib.org/dlib/june05/jantz/06jantz.html. Acesso em: 4 maio. 2021.

KIM, Suntae. Functional requirements for research data repositories. **International Journal of Knowledge Content Development & Technology**, 2018. Disponível em: http://koreascience.or.kr/article/JAKO201823954941675.page. Acesso em: 10 fev. 2021.
KÖCHE, José Carlos. **Fundamentos de metodologia científica:** teoria da ciência e iniciação à pesquisa. Petrópolis: Vozes, 2003.

LAGUARDIA, Josué; PORTELA, Margareth Crisóstomo; VASCONCELLOS, Miguel Murat. Avaliação em ambientes virtuais de aprendizagem. **Educação e Pesquisa**, São Paulo, v.33, n.3, p. 513-530, set./dez. 2007. Disponível em: https://www.scielo.br/pdf/ep/v33n3/a09v33n3.pdf. Acesso em:10 ago. 2020.

LAMEIRA, Ana Kelly Alves. Avaliação de repositórios institucionais brasileiros: uma proposta de método de avaliação. **Cadernos BAD**, 2016. Disponível em: https://bad.pt/publicacoes/index.php/cadernos/article/view/1594. Acesso em: 17 abr. 2020.

L'HOURS, Hervé; KLEEMOLA, Mari; LEEUW, Lisa de. CoreTrustSeal: From academic collaboration to sustainable services, **IASSIST Quarterly**, 2019. Disponível em: https://www.iassistquarterly.com/index.php/iassist/article/view/936/932. Acesso em: 7 maio. 2021.

MÁRDERO ARELLANO, Miguel Ángel. Critérios para a preservação digital da informação científica. 354f. **Tese** (Doutorado em Ciência da Informação) - Universidade Federal de Brasília, Departamento de Ciência da Informação, 2008. Disponível em: https://core.ac.uk/download/pdf/11884842.pdf. Acesso em: 27 maio. 2021.

MARQUES, Fabrício. Ciência transparente. **Pesquisa FAPESP**, 2014. Disponível em: https://revistapesquisa.fapesp.br/2014/04/24/ciencia-transparente/. Acesso em: 1 jun. 2020.

MEDEIROS, Jackson da Silva. **Uma investigação sobre a autoria de dados científicos: teias de uma rede em construção.** 2015. Tese (doutorado) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2015. Disponível em: https://lume.ufrgs.br/handle/10183/116504. Acesso em: 3 abr. 2020.

MEDINA-SMITH, Andrea. **A Self-Audit of the NIST Public Data Repository Using the CoreTrustSeal Trustworthy Data Repositories Requirements.** NIST, 2021. Disponível em: https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8341.pdf. Acesso em: 17 dez. 2020.

MINAYO, Maria Cecília de Souza *et al*. **Pesquisa social:** teoria, método e criatividade. Petrópolis: Vozes, 1994.

MISGAR, Safat Mushtaq; BHAT, Ajra; WANI, Zahid Ashraf. **A study of Open Access research data repositories developed by BRICS countries.** 2020. Disponível em: https://www-emerald.ez45.periodicos.capes.gov.br/insight/content/doi/10.1108/DLP-02-2020-0012/full/pdf?title=a-study-of-open-access-research-data-repositories-developed-by-brics-countries. Acesso em: 4 fev. 2021.

OPENAIRE. **OpenAIRE Guidelines for literature repository managers.** 2018. Disponível em: https://openaire-guidelines-for-literature-repository managers.readthedocs.io/en/v4.0.0/introduction.html. Acesso em: 11 set 2020.

OPENAIRE. **OpenAIRE Guidelines for Data Archives.** 2018. Disponível em: https://guidelines.openaire.eu/en/latest/data/use_of_datacite.html. Acesso em: 18 abr. 2021.

OPEN KNOWLEDGE FOUNDATION *et al.* **Open data handbook**. 2009. Disponível em: http://opendatahandbook.org/guide/pt_BR/. Acesso em: 30 jan. 2020.

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT – OECD. **Declaration on Access to Research Data from Public Funding** 2004. Disponível em: https://www.oecd.org/newsroom/38528123.pdf. Acesso em: 13 maio. 2020.

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT – OECD. **Principles and Guidelines for Access to Research Data from Public Funding.** 2007. Disponível em: https://www.oecd.org/sti/inno/38500813.pdf. Acesso em: 15 jan. 2020.

PAVÃO, Caterina Groposo; SILVA, Fabiano Couto Corrêa da; SILVEIRA, Lúcia da. Gestão de dados em periódicos científicos. *In*: SHINTAKU, Milton; SALES, Luana Farias; COSTA, Michelli. (org). **Tópicos sobre dados abertos para editores científicos.** Botucatu, SP: ABEC, 2020. Disponível em: https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf. Acesso em: 13 abr. 2020.

PAVÃO, Caterina Groposo *et al*. **Acesso aberto a dados de pesquisa no Brasil: repositórios brasileiros de dados de pesquisa:** relatório 2018. Disponível em: http://hdl.handle.net/10183/185138. Acesso em: 2 mar. 2022.

PINHEIRO, Lena Vania Ribeiro. Do acesso livre à ciência aberta: conceitos e implicações na comunicação científica. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde - RECIIS**, v. 8, n. 2, 2014. Disponível em: https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/629/1269. Acesso em: 29 jan. 2020.

PIWOWAR, Heather A; DAY, Roger S; FRIDSMA, Douglas B. Sharing Detailed Research Data Is Associated with Increased Citation Rate. **PLoS ONE**, 2007. Disponível em: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308. Acesso em: 16 jan. 2021.

PRESSER, Nadi Helena; SILVA, Eli Lopes da. O gerenciamento dos dados de pesquisa como wicked problem. **Navus**, Florianópolis, v. 8, n. 4, 2018. Disponível em: http://navus.sc.senac.br/index.php/navus/article/view/855/pdf. Acesso em: 19 fev. 2020.

RAUEN, Cristiane Vianna. A relevância de uma política nacional de acesso aberto a dados de pesquisa. **Revista Construção**, 2018. Disponível em: http://revistaconstrucao.org/ciencia-e-tecnologia/relevancia-de-uma-politica-nacional-de-acesso-aberto-dados-de-pesquisa/. Acesso em 14 maio. 2020.

RESEARCH DATA CANADA – RDC. **Trust principles Mini Symposium:** the future of Digital Repositories, 2020. Disponível em: https://www.rdc-drc.ca/trust-principles-mini-symposium-the-future-of-digital-repositories/. Acesso em: 28 abr. 2021.

RIBEIRO, Cláudio José Silva. Modelo de maturidade para repositórios digitais: um caminho para sua adoção na gestão de dados de pesquisa. **Liinc em Revista**, Rio de Janeiro, v.15, n.2, p. 224-243, novembro 2019. Disponível em: http://revista.ibict.br/liinc/article/view/4816. Acesso em: 6 abr. 2020.

RIBEIRO, Odília Barbosa; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Otimização do acesso à informação científica: discussão sobre a aplicação de elementos da arquitetura da informação em repositórios digitais. **Biblos**, 2009. Disponível em: https://periodicos.furg.br/biblos/article/view/1309/593. Acesso em: 4 jun. 2020.

RLG/OCLC WORKING GROUP ON DIGITAL ARCHIVE ATTRIBUTES. **Trusted digital repositories:** attributes and responsibilities. May 2002. Disponível em: https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf. Acesso em: 24 mar. 2021.

ROCHA, Lucas. As potencialidades do data paper na ciência atual. *In*: SHINTAKU, Milton; SALES, Luana Farias; COSTA, Michelli. (org). **Tópicos sobre dados abertos para editores científicos.** Botucatu, SP: ABEC, 2020. Disponível em: https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf. Acesso em: 13 abr. 2020.

ROCHA, Rafael Port da; CAREGNATO, Sonia; GABRIEL JUNIOR Rene Faustino. Aspectos de inovação na implantação de um centro de digitalização e gestão de dados da pesquisa. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, v. 23, n. esp., 2018. Disponível em: https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23nespp1. Acesso em: 21 fev. 2010.

ROYAL SOCIETY. **Science as an open enterprise**, 2012. Disponível em: https://royalsociety.org/-/media/policy/projects/sape/2012-06-20-saoe.pdf. Acesso em: 23 maio. 2020.

SALES, Luana Farias. **Integração semântica de publicações científicas e dados de pesquisa:** proposta de modelo de publicação ampliada para a área de Ciências Nucleares. 2014. Tese (Doutorado em Ciência da Informação) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014. Disponível em: https: //ridi.ibict.br/bitstream/123456789/874/1/LUANA%20SALES%20D.pdf. Acesso em: 28 jul. 2020.

SALES, Luana Farias; SAYÃO, Luís Fernando. Ampliando as fronteiras da editoração científica: o papel dos repositórios de dados. *In*: SHINTAKU, Milton; SALES, Luana Farias; COSTA, Michelli. (org). **Tópicos sobre dados abertos para editores científicos.** Botucatu, SP: ABEC, 2020.Disponível em: https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf. Acesso em: 13 abr. 2020.

SALES, Luana Farias. **Princípios FAIR aplicados à repositórios**, 2021. Disponível em: https://www.arca.fiocruz.br/bitstream/icict/50571/2/RIAA_Curso_Princ%c3%adpios_FAIR_20_10_2021.pdf. Acesso em: 3 jan. 2022.

SANCHEZ, Fernanda Alves; VECHIATO, Fernando Luiz. VIDOTTI, Silvana Aparecida Borsetti Gregorio. Encontrabilidade da Informação em Repositórios de Dados: uma análise do DataONE. **Informação & Informação.** Londrina, v. 23, n. 1, p. 51 – 79, jan./abr. 2019. Disponível em: http://www.uel.br/revistas/uel/index.php/informacao/article/view/30725. Acesso em: 9 abr. 2020.

SANTOS, Gildenir Carolino. Como os indexadores estão vendo a questão de dados abertos de pesquisa para o processo de indexação?. *In* SHINTAKU, Milton; SALES, Luana Farias; COSTA, Michelli. (org).**Tópicos sobre dados abertos para editores científicos.** Botucatu, SP: ABEC, 2020. Disponível em: https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf. Acesso em: 13 abr. 2020.

SANTOS, Paula Xavier dos; ALMEIDA, Bethânia de Araújo; HENNING, Patrícia (org.). **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. Disponível em: https: //www.arca.fiocruz.br/bitstream/icict/24117/2/Livro-Verde-07-06-2018.pdf. Acesso em: 5 abr. 2020.

SANTOS, Henrique Machado dos; FLORES, Daniel. Políticas de preservação digital para documentos arquivísticos. **Perspectivas em Ciência da Informação**, 2015. Disponível em: http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2542. Acesso em 3 jun. 2021.

SANTOS, Henrique Machado dos; FLORES, Daniel. Introdução aos conceitos básicos do modelo Open Archival Information System no contexto da arquivística. **Researchgate.net**, 2019. Disponível em: https://www.researchgate.net/publication/333297791_Introducao_aos_conceitos_basicos_do_

modelo_Open_Archival_Information_System_no_contexto_da_arquivistica. Acesso em: 2 out. 2021.

SANTOS, Henrique Machado dos. Manual para auditoria de repositórios arquivisticos digitais confiáveis, 2018. Disponivel em https://repositorio.ufsm.br/bitstream/handle/1/15909/DIS_PPGPC_2018_SANTOS_HENRIQUE.pdf?sequence=1&isAllowed=y. Acesso em 15 ago. 2021.

SAYÃO, Luis Fernando; SALES, Luana. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **RECIIS | 2014**, jun., 8(2) – p.76-92, 2014. Disponível em: www.reciis.icict.fiocruz.br/index.php/reciis/article/download/611/1252. Acesso em: 20 jan. 2020

SAYÃO, Luis Fernando; SALES, Luana Farias. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação.** Londrina, maio/ago., 2016. Disponível em: http:www.uel.br/revistas/informacao/. Acesso em: 4 ago. 2020.

SAYÃO, Luis Fernando; SALES, Luana Farias. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores.** CNEN: Rio de Janeiro, 2015.

SAYÃO, Luis Fernando; SALES, Luana Farias. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, v. 21, n. 2, p. 90- 115, 2016. Disponível em: http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939. Acesso em: 26 mar. 2020.

SAYÃO, Luis Fernando; SALES, Luana Farias. Subsídios para a construção de um modelo de avaliação de sistemas de gestão de dados de pesquisa. **Ponto de Acesso**, Salvador, v.12, n.3, p.80-108, dez. 2018. Disponível em: http://www.brapci.inf.br/index.php/res/download/119953. Acesso em: 3 abr. 2020.

SAYÃO, Luis Fernando; SALES, Luana Farias**.** A ciência invisível: revelando os dados da cauda longa da pesquisa. **XIX ENANCIB** - Encontro Nacional de Pesquisa em Ciência da Informação. Londrina/Paraná, 2019. Disponível em: https://www.researchgate.net/publication/330449075_A_CIENCIA_INVISIVEL_REVELANDO_OS_DADOS_DA_CAUDA_LONGA_DA_PESQUISA. Acesso em: 5 abr. 2020.

SAYÃO, Luis Fernando; SALES, Luana Farias**.** A grande e a pequena ciência: análise das diferenças na gestão de dados de pesquisa**. Research Gate**, 2019. Disponível em: https://www.researchgate.net/publication/331385359. Acesso em: 25 mar. 2020.

SEMELER, Alexandre Ribas. **Ciência da Informação em contextos de E-Science:** bibliotecários de dados em tempos de data Science. 2017 Tese (doutorado) – Universidade Federal de Santa Catarina, Florianópolis, 2017. Disponível em: https://repositorio.ufsc.br/bitstream/handle/123456789/185593/PCIN0168-T.pdf?sequence=-1&isAllowed=y. Acesso em: 11 jun. 2020.

SHINTAKU, Milton. **Federação de repositórios científicos:** identificação, análise e proposta de modelo baseado nas tendências tecnológicas e da Ciência. 2014. Tese (doutorado). Universidade de Brasília, Brasília, 2014. Disponível em: https://repositorio.unb.br/bitstream/10482/18125/1/2014_MiltonShintaku.pdf. Acesso em: 27 ago. 2020.

SHINTAKU, Milton; LANNE, Suzana. Caetano da Silva. Dados abertos e a Fundação de Amparo à Pesquisa do Estado de São Paulo. *In*: SHINTAKU, Milton; SALES, Luana Farias; COSTA, Michelli. (org).**Tópicos sobre dados abertos para editores científicos.** Botucatu, SP: ABEC, 2020. Disponível em: https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf. Acesso em: 13 abr. 2020.

SILVA, Fabiano Couto Corrêa da. **Gestão de dados científicos.** Rio de Janeiro: Interciência. 2019.

SILVA, Rosane Mendes da. **Avaliação de qualidade de repositórios institucionais:** o caso do repositório da ENSP. 2013. Dissertação (mestrado), Fiocruz, Rio de Janeiro, 2013. Disponível em: https://www.arca.fiocruz.br/bitstream/icict/24737/1/410.pdf. Acesso em: 18 jul. 2020.

SILVA, Lucas Henrique Alves da, *et al*. Os Princípios TRUST como ferramenta de avaliação de repositórios de dados. **Brazilian Journal of Information Studies:** Research trends, 2021. Disponível em: https://revistas.marilia.unesp.br/index.php/bjis/article/view/11283/7184. Acesso em: 29 mar. 2021.

SOUZA, Luciana Gonçalves Silva; AGANETTE, Elisângela Cristina. Repositórios digitais confiáveis. **Informação & Sociedade,** 2020. Disponível em: https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/45426/29688. Acesso em: 14 abr. 2020.

TARTAROTTI, Roberta Cristina Dal´Evedove; DAL´EVEDOVE, P. R.; FUJITA, Mariângela Spotti Lopes. Biblioteconomia de dados em repositórios de pesquisa: perspectivas para a atuação bibliotecária. **Informação & Informação**, v. 24, n. 3, p. 207-226, 2019. Disponível em: https://www.brapci.inf.br/index.php/res/v/134221. Acesso em: 27 mar. 2020.

TEIXEIRA, Enise Barth. A análise de dados na pesquisa científica: importância e desafios em estudos organizacionais. **Desenvolvimento em questão**, 2003. Disponível em: https://www.revistas.unijui.edu.br/index.php/desenvolvimentoemquestao/article/view/84. Acesso em: 19 out. 2020.

TÉRMENS, Miquel; LEIJA, David. Auditoría de preservación digital con NDSA Levels: methodology of digital preservation audits with NDSA Levels. **El profesional de la información,** 2017. Disponível em: https://fbd.ub.edu/pub/termens/docs/EPI-v26n3.pdf. Acesso em: 06 maio. 2021.

THOMAZ, Katia de Padua. Repositórios Digitais confiáveis e certificação. **Arquivística.net**, Rio de Janeiro, v. 3, n.1, p. 80-89, jan./jun. 2007. Disponível em: www.brapci.inf.br/index.php/article/download/10726. Acesso em: 2 jul. 2020.

TORINO, Emanuelle; ROA-MARTÍNEZ, Sandra Milena; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Dados de pesquisa: disponibilização ou publicação? *In*: SHINTAKU, Milton; SALES, Luana Farias; COSTA, Michelli. (org).**Tópicos sobre dados abertos para editores científicos.** Botucatu, SP: ABEC, 2020. Disponível em: https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf. Acesso em: 13 abr. 2020.

VECHIATO, Fernando Luiz; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Encontrabilidade da informação: atributos e recomendações para ambientes informacionais digitais. **Informação & Tecnologia (ITEC)**: Marília/João Pessoa, v. 1, n. 2, p. 42-58, jul./dec., 2014. Disponível em: https://repositorio.unesp.br/bitstream/handle/11449/114982/ISSN23583908-2014-01-02-42-58.pdf?sequence=2&isAllowed=y. Acesso em: 4 ago. 2020.

WILKINSON, Mark D. *et al*. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data,** 2016**.** Disponível em: https://www.nature.com/articles/sdata201618. Acesso em: 23 jan. 2020.

WORLD DATA SYSTEM. **Trusted Data Services for Global Science**. 2020. Disponível em: https://www.icsu-wds.org/organization/intro-to-wds. Acesso em: 1 jul. 2020.

YOON, Ayoung. End users' trust in data repositories: definition and influences on trust development. **Arch. Sci.**, 2014. Disponível em: https://scholarworks.iupui.edu/bitstream/handle/1805/13592/Yoon_2014.pdf?sequence=1&isAllowed=y. Acesso em: 20 maio. 2021