

Introdução à Ciência Aberta

Curso 1

A trilha Ciência Aberta abrange apostilas criadas pelo grupo RDPBrasil para treinamentos promovidos pela Rede Nacional de Pesquisa (RNP). O curso 1 intitula-se Introdução à Ciência Aberta.

Pesquisadores

Samile Andréa de Souza Vanz
Caterina Groposo Pavão
Rafael Port da Rocha
Rene Faustino Gabriel Junior
Sônia Elisa Caregnato

Bolsistas

Ana Júlia Lopes
Lucia Helena Cunha Vidal
Wagner Silva Wessfl

Contato

site: www.ufrgs.br/rdpbrasil
Rua Ramiro Barcelos, 2705 - Campus Saúde
Sala: 106 - Anexo 1
Brasil - Porto Alegre - RS - CEP 90.035-007
Telefone: +55 (51) 3308-5942



Esta apostila é licenciada sobre a licença CC BY - Creative Commons Attribution 4.0 International License.

Dados internacionais de catalogação na publicação
Lucia Helena Cunha Vidal CRB 10/2035

V285 Vanz, Samile Andréa de Souza
Introdução à ciência aberta / Pesquisadores Samile Andréa de Souza Vanz ; Caterina Groposo Pavão ; Rafael Port da Rocha ; Rene Faustino Gabriel Junior ; Sônia Elisa Caregnato; Bolsistas Ana Júlia Lopes ; Lucia Helena Cunha Vidal; Wagner Silva Wessfl. – Porto Alegre, RS : UFRGS, 2023.
51 f.

Curso 1 da trilha Ciência Aberta abrange apostilas criadas pelo grupo RDPBrasil para treinamentos promovidos pela Rede Nacional de Pesquisa (RNP).

1. Ciência aberta. 2. Dados de pesquisa. 3. Gestão de dados de pesquisa. 4. Ecossistema de dados de pesquisa. I. Título. II. Pavão, Caterina Groposo. III. Rocha, Rafael Port da. IV. Gabriel Junior, Rene Faustino. V. Caregnato, Sônia Elisa. VI. Lopes, Ana Júlia. VII. Vidal, Lucia Helena Cunha. VIII. Wessfl, Wagner Silva.

Como citar:

VANZ, S. A. de S.; PAVÃO, C. G.; ROCHA, R. P. da; GABRIEL JUNIOR, R. F.; CAREGNATO, S. E.; LOPES, A. J.; VIDAL, L. H. C.; WESSFLL, W. S. Introdução à ciência aberta. Porto Alegre: RDPBrasil, 2023.



ibict

RNP
ORGANIZAÇÃO SOCIAL DO MCTI

Curso 1

Introdução à Ciência Aberta

MÓDULO I

Objetivos

Aprender conceitos básicos sobre Ciência, Comunicação Científica e como os Dados de Pesquisa funcionam dentro dessa engrenagem.

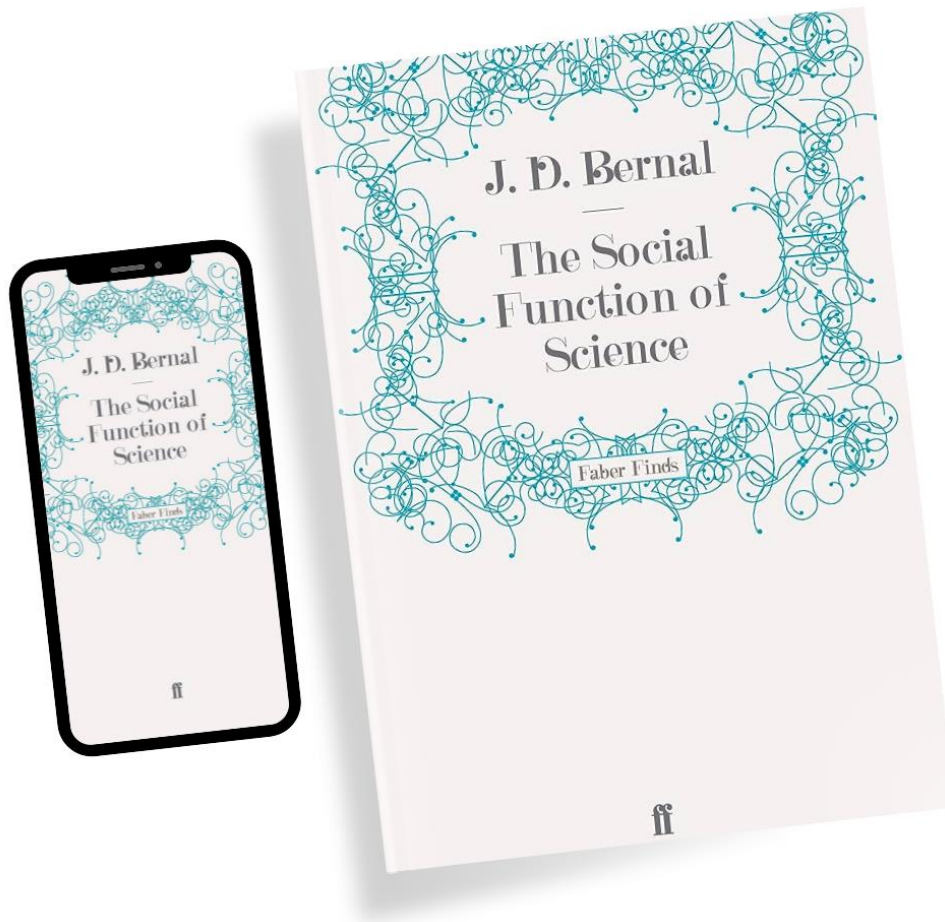
Conceitos

Ciência; Comunicação Científica; Dados de Pesquisa.

1.1 A CIÊNCIA E A COMUNICAÇÃO CIENTÍFICA

Em 1939, John Desmond Bernal publicou a obra *The Social Function of Science* (A função social da ciência), – Figura 1 – considerado possivelmente o primeiro texto da sociologia da ciência. Nesta obra, Bernal faz um resgate histórico da ciência, informando sua origem no século XVI, derivada de especulações de mágicos, padres e filósofos e das operações práticas dos artesãos. O autor afirma que a ciência está a serviço do homem e tornou-se parte integral da vida material e econômica. Foi ela que colocou nas mãos do homem o gosto pelas necessidades materiais satisfeitas, além das ideias que nos possibilitam entender, coordenar e satisfazer necessidades na esfera social e cultural.

Figura 1 – Capa de The Social Function of Science | John Desmond Bernal | 1939



A partir do lançamento do livro de Bernal, diversos outros cientistas se dedicaram a entender a ciência, a aplicação do método científico e a descoberta dos resultados, resultando em inúmeras obras categorizadas como ciência da ciência, sociologia da ciência ou filosofia da ciência. Estas ciências se dedicam a estudar a ciência e seu *modus operandi*, aspectos metafísicos, epistemológicos e metodológicos da produção de conhecimento.

O conhecimento científico é aquele obtido através do resultado de pesquisa científica, quando um fenômeno/objeto é analisado segundo uma metodologia científica, com regras definidas e controladas. A publicação da obra O Discurso do Método, de René Descartes, em 1637, trouxe o reconhecimento definitivo ao método científico, com base na observação e na interpretação através da

demonstração experimental. Foi sua implementação no século XVII que promoveu a institucionalização da ciência. O uso do método científico e a publicação dos resultados de pesquisa distingue o conhecimento científico do popular.

“A ciência busca, essencialmente, desvendar e compreender a natureza e seus fenômenos, através de métodos sistemáticos e seguros. No entanto, face à dinamicidade intrínseca à própria natureza, seus resultados são sempre provisórios. Isto é, esses sistemas explicativos não têm caráter permanente. Inserem-se num processo ininterrupto de investigação, o que faz da ciência uma instituição social, dinâmica, contínua, cumulativa” (TARGINO, 2000, p. 2).

A comunicação científica incorpora as atividades associadas à produção, disseminação e uso da informação, desde o momento em que o cientista concebe uma ideia para pesquisar até o momento em que a informação acerca dos resultados de pesquisa é aceita como parte do estoque universal de conhecimentos.

A comunicação científica varia de acordo com as práticas das comunidades científicas e é fundamental para que o cientista possa se informar sobre os resultados de pesquisa de seus pares; divulgar os resultados alcançados na sua própria pesquisa à sua comunidade; e apoiar a estrutura de recompensa acadêmica.

Percebe-se que o processo de pesquisa científica é cíclico e, portanto, todas as atividades realizadas pelos pesquisadores são encadeadas em uma série sucessiva, incluindo a coleta de dados, a aplicação do método científico e análise dos dados, a publicação dos resultados de pesquisa em artigos e *papers*, até que estes resultados dêem origem a novas ideias para novos projetos de pesquisa.

O contexto atual de fazer ciência, chamado *eScience*, pressupõe a colaboração independentemente da localização física, o acesso a instrumentos, dados, informações e recursos computacionais, e o acesso à bibliotecas digitais. Há uma tendência à publicação dos artigos que apresentam os resultados de pesquisa em

open journals (periódicos de acesso aberto), assim como ao arquivamento destes artigos em repositórios abertos e a demanda em torno do compartilhamento dos dados de pesquisa.

1.2 CONTEXTO DA GERAÇÃO, USO E COMPARTILHAMENTO DE DADOS

Os dados são criados ou produzidos de várias maneiras: por meio de observação, de visualização, de monitoramento e de sensores, na criação de metadados, por análise de comportamento, por cálculos matemáticos e estatísticos, etc. Com a consolidação da ciência aberta os dados começaram a ser reconhecidos como elementos importantes para a comunidade científica.

Os dados normalmente são gerados de duas maneiras: em formato digital ou material, como no caso dos projetos de pesquisa, documentos acadêmicos (que se tornam objetos digitais complexos), resultados de um experimento realizado em um laboratório, estudos empíricos, observação de um fenômeno cultural ou da erupção de um vulcão em determinado momento e lugar, por exemplo. Praticamente tudo o que é produzido e coletado no ambiente de pesquisa pode ser considerado dado, mas no caso dos dados em formato digital, não podem ser interpretados sem o devido suporte tecnológico.

Figura 2 - Gestão de dados



Fonte: Freepik (2023).

De acordo com Gabriel Junior *et al.* (2019, p. 90), o compartilhamento de dados:

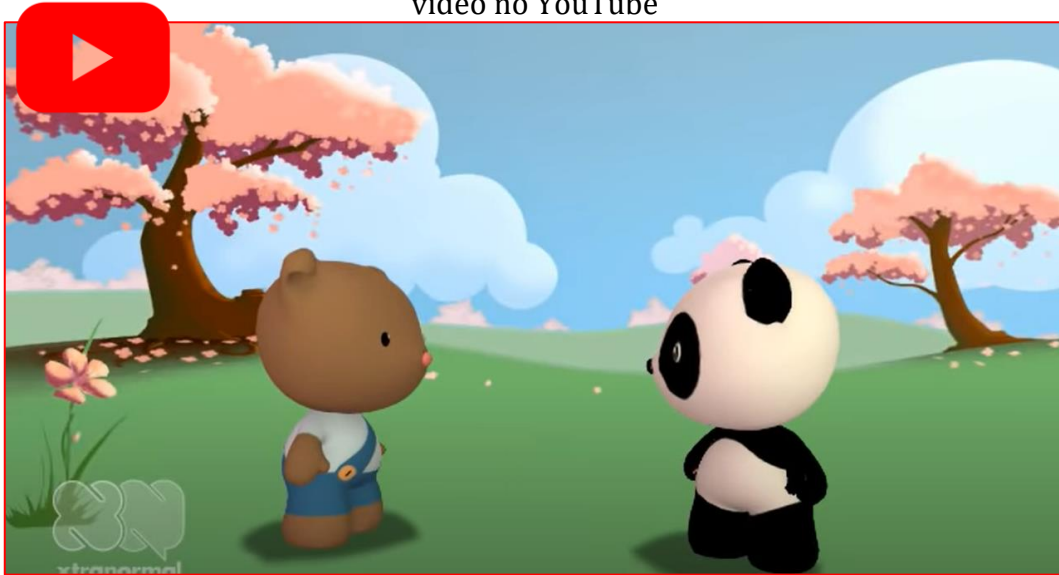
Consiste no ato de disponibilizar dados gerados ou coletados por uma pesquisa para reuso de outros pesquisadores. Os benefícios mais evidentes são a possibilidade da reprodução ou da verificação da pesquisa; a disponibilização dos resultados de pesquisas financiadas com fundos públicos; a continuidade das pesquisas dos questionamentos a respeito dos dados existentes e, conseqüentemente, a viabilização de avanços no estado da pesquisa e na inovação.

Os benefícios que o compartilhamento de dados proporciona à comunidade científica são:

- o incentivo a compreensão de várias disciplinas e perspectivas quanto a determinado tema;
- a identificação de erros de interpretação e fraudes;
- evita o desperdício de recursos financeiros e a repetição de coleta de dados;

- o conhecimento de dados negativos (esses dados, na maioria das vezes, não são publicados e não geram um artigo, mas são fundamentais para o avanço da ciência pois constituem-se em aprendizado, mostram equívocos que não precisam ser repetidos e, com isso, proporcionam economia de recursos);
- promove benefícios para o pesquisador em relação às citações e reconhecimento por aquele estudo.

Figura 3 - Compartilhamento e gerenciamento de dados – um exemplo prático | vídeo no YouTube



Fonte: NYU Health Sciences Library (2013).

1.3 ABORDAGENS PARA REUSO DE DADOS

Uso dos dados: quando se discute sobre o uso dos dados é necessário compreender que há uma estrutura que representa a vida útil do dado, denominado de ciclo de vida dos dados. Ele delimita as fases de acesso e posterior uso (ou reúso) e foi promovida como um conceito para apoiar práticas de preservação digital e de curadoria tendo popularidade conforme a cultura de

compartilhamento de dados se tornasse parte da linguagem de pesquisa cotidiana, proporcionando estrutura e suporte para o uso dos dados, gerando mais pesquisas e financiamentos.

Um modelo internacional de referência para pesquisadores, bibliotecários e gestores de dados citado por Ball (2012) é o *UK Data Archive Data Lifecycle* desenvolvido pelo Arquivo de Dados do Reino Unido. O *UK Data Archive* fornece um modelo de ciclo de vida de dados como uma ajuda para os pesquisadores, considerando como o gerenciamento de dados se relaciona com o ciclo de vida de um projeto de pesquisa. Esse modelo define seis etapas de ações, como ilustrado na Figura 4.

Figura 4 - Modelo do Ciclo de Vida dos Dados de Pesquisa



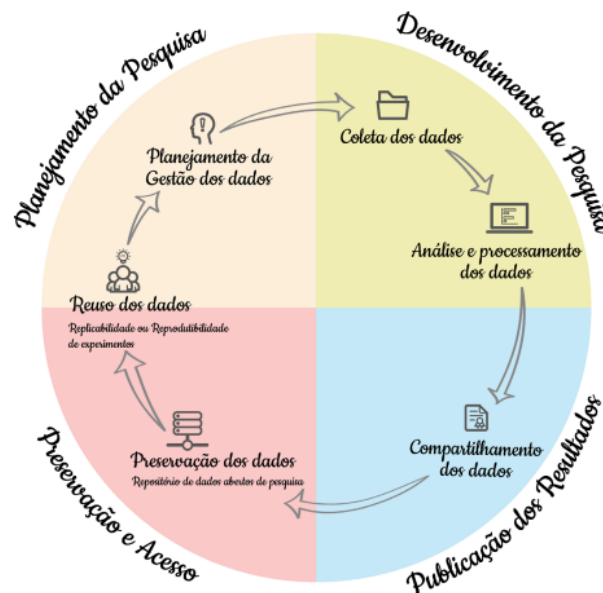
Fonte: UK Data Service (2019).

- **Planejamento:** compreende o planejamento da pesquisa, gestão e armazenamento dos dados, observando seus formatos, protocolos e modelos de processamento para explorar as fontes de dados existentes;
- **Coleta de dados:** etapa da pesquisa que consiste na coleta de dados;
- **Processamento e análise de dados:** equivale a fase de digitalizar, transcrever e traduzir dados, isto é, fazer a checagem dos dados;

- **Publicação e compartilhamento:** consiste em estabelecer questões de direitos autorais, criar documentação do usuário, formas de acesso apropriado aos dados para publicar e compartilhá-los;
- **Preservação dos dados:** compreende em migrar os dados para o melhor formato, armazená-los e fazer *backup*, criar documentação para preservação e curadoria;
- **Reutilização de dados:** consiste na descoberta dos dados por outros pesquisadores, possibilitando outras aplicações e estudos.

No contexto brasileiro o grupo de pesquisa RDP Brasil, em projeto da RNP, IBICT e CNPq, desenvolveu um modelo de ciclo de vida do Dado Aberto de Pesquisa levando em conta a realidade dos estudos e discussões sobre acesso aos dados de pesquisa no país, relacionando as ações inerentes ao ciclo de vida do dado às etapas de desenvolvimento da pesquisa científica. Conforme apresenta a Figura 5, a seguir, o modelo abrange um processo constante dividido em quadrantes que representam etapas da pesquisa científica.

Figura 5 - Modelo de Ciclo de Vida do Dado Aberto de Pesquisa



Fonte: Gabriel Junior *et al.* (2019).

De acordo o modelo da Figura 5 nota-se que ele tem quatro etapas do ciclo de vida de um dado com seis ações conforme descritas a seguir:

- **Planejamento da Pesquisa:** etapa inicial da pesquisa, que no contexto da ciência aberta pode abranger o reúso de dados, considerando a disponibilidade dos mesmos em repositórios de dados;
- **Desenvolvimento da pesquisa:** etapa que compreende a coleta, a análise e o processamento dos dados obtidos, sejam eles coletados na fonte ou reutilizados, ou seja, coletados por outros pesquisadores em pesquisas anteriores;
- **Publicação dos Resultados:** etapa tradicional da pesquisa científica, quando o pesquisador divulga seus resultados de pesquisa à comunidade científica. No âmbito da ciência aberta, é também o momento em que o processo de compartilhamento dos dados geralmente inicia;
- **Preservação e acesso:** etapa final que consiste na preservação das publicações e dos dados por meio dos repositórios de dados abertos de pesquisa possibilitando o reúso de dados, ação compartilhada com a etapa de planejamento da pesquisa, reiniciando-se, assim, o ciclo de vida do dado da pesquisa.

O reuso dos dados de pesquisa é uma prática valiosa para a comunidade científica, pois além de permitir a validação dos resultados de estudos, permite a produção de novos conhecimentos a partir dos dados já coletados. Alguns autores acreditam que os benefícios do compartilhamento de dados apenas podem ser avaliados por meio da sua reutilização (NIU, 2009), assim, os repositórios serviriam de impulsionadores para o compartilhamento, além de otimizar e maximizar recursos financeiros, físicos e financeiros.

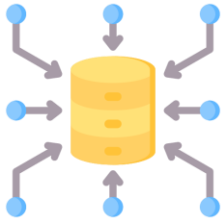
Curty (2019) sugere a seguinte classificação composta por cinco abordagens para o reuso de dados de pesquisa:

- reaproveitamento;

- agregação;
- integração;
- metanálise;
- reanálise.



REAPROVEITAMENTO: “[...] dados de um estudo são reutilizados integralmente, ou parcialmente selecionados e aproveitados, sem que sejam complementados ou integrados com dados provenientes de outras fontes. São submetidos a novas análises, de modo a responderem questões de pesquisas diferentes daquelas abordadas no estudo que originou os dados. O reaproveitamento é muito comum em estudos que reutilizam dados de levantamento com grandes amostras.”



AGREGAÇÃO: “O reúso por agregação reúne dados de um mesmo domínio provenientes de diferentes estudos/fontes para compor uma coleção de dados (*dataset*) mais completa. A partir da agregação o pesquisador busca compor uma coleção de dados, tomando como base variáveis ou construtos predefinidos, que permitam agregar casos, elementos (*data points*) ou indicadores que se encaixam ao estudo. Pesquisadores podem combinar dados de diferentes fontes preexistentes, inclusive com dados primários que tenham coletado.”



INTEGRAÇÃO: “O reúso por integração pode ser entendido como aquele que combina dados de diferentes domínios e diferentes tipos de estudos. Esse tipo de integração depende de variáveis que interliguem estudos separados. A variável de

integração pode ser um número de acesso, um nome de espécie ou uma localização geográfica, por exemplo.”



METANÁLISE: “[...] caracteriza-se como aquele que combina análises de dados provenientes de múltiplos estudos independentes com perguntas de pesquisa e hipóteses idênticas ou muito semelhantes de modo a abordar questões além do escopo ou poder de uma única análise.”



REANÁLISE: “[...] voltado para verificação dos resultados obtidos pelo estudo que gerou os dados, por meio de nova análise dos dados, utilizando os mesmos métodos e técnicas empregados no estudo original. Por conseguinte, essa abordagem de reuso está diretamente atrelada à questão da reprodutibilidade em que dados são submetidos à verificação, com vistas à confirmação ou refutação das conclusões anteriores.”

5 ABORDAGENS QUE PODEM SER UTILIZADAS PARA CLASSIFICAR O REUSO DE DADOS

Aprendendo sobre Dados de Pesquisa, n. 12, 2022

Fonte: traduzido e adaptado de Open Library Series, n.6(2021) (U.S. Santa Barbara Library).

RDP BRASIL
Rede de Dados de Pesquisa
ufrgs.br/rdpbrasil

Quer saber mais?
ufrgs.br/rdpbrasil

Como os dados de pesquisa podem ser reutilizados?

Dados disponíveis publicamente podem ser reutilizados inúmeras vezes por quem criou, coletou ou por outras partes não envolvidas no estudo original.

Tome nota das abordagens mais comuns para a reutilização de dados

REANÁLISE

Procedimento metodológico que segue o mesmo processo analítico empregado pelo estudo original para verificação dos resultados (refuta ou confirma achados anteriores).

AGREGAÇÃO

Procedimento metodológico que reúne dados de várias fontes existentes para formar um novo conjunto de dados.

REAPROVEITAMENTO

Procedimento metodológico que explora conjuntos de dados existentes de diferentes maneiras para responder a novas perguntas de pesquisa.

INTEGRAÇÃO

Procedimento metodológico que combina dados de fontes de dados diferentes vinculadas por uma ou mais variáveis relacionadas.

META-ANÁLISE

No contexto dos dados de pesquisa, é o procedimento metodológico que mescla análises de dados advindos de várias fontes independentes.

licenciado por Creative Commons
Atribuição-NC-ND-CC-SA

UFRGS RNP ibict CNPq

Aprendendo sobre Dados de Pesquisa, n. 10, 2023

Fonte: traduzido e adaptado de Open Library Series, n.6(2021) (U.S. Santa Barbara Library).

RDP BRASIL
Rede de Dados de Pesquisa
ufrgs.br/rdpbrasil

Quer saber mais?
ufrgs.br/rdpbrasil

5 abordagens que podem ser utilizadas para classificar o reuso de dados

REAPROVEITAMENTO

é abordagem mais comum onde os dados são aproveitados parcialmente ou integralmente para outro propósito, sem que sejam complementados ou integrados com outros dados provenientes de outras fontes. Para o caso comum nas pesquisas brasileiras do uso de um subconjunto de dados filtrados do DataSUS;

AGREGAÇÃO

é a reunião de dados de um mesmo domínio proveniente de diferentes fontes. Por exemplo, a reunião de dados climáticos de diferentes estações de medição distribuídas nas regiões do país para realizar uma pesquisa climática a nível nacional;

INTEGRAÇÃO

é entendida como a combinação de dados de diferentes domínios e diferentes tipos de estudo a partir de variáveis que interliguem os estudos. Por exemplo, um estudo interessado em investigar o aumento das mortes relacionado à temperatura ambiente poderia utilizar dados do clima e das mortes, utilizando como variável de ligação o município;

META-ANÁLISE

é a combinação de análises de dados provenientes de estudos independentes com hipóteses e perguntas de pesquisa muito semelhantes ou idênticas. Neste caso, ao invés de reutilizar os dados primários de cada estudo, é feita uma análise das análises para extrair conclusões sobre o estado da arte em determinado tema de pesquisa;

REANÁLISE

é a verificação dos resultados obtidos pelo estudo original, através de nova análise dos dados, utilizando os mesmos métodos e técnicas utilizadas no estudo original.

FONTE:

CURTY, R. Abordagens de reuso e a questão da reutilização dos dados científicos [Approaches for data reuse and the issue of scientific data reusability]. *Linac em Revista*, v. 15, n. 2, p. 177-195, 2019. Disponível em: <https://doi.org/10.18637/linac.v15i2-4777>.

licenciado por Creative Commons
Atribuição-NC-ND-CC-SA

UFRGS RNP ibict CNPq

Como os dados de pesquisa podem ser reutilizados?

MÓDULO II

Objetivos

Aprofundar-se nos conceitos da Ciência Aberta, entendendo seus fenômenos.

Conceitos

Ciência Aberta; Acesso Aberto; *e-Science*, *Data Mining*; *Big Data*.

2 CIÊNCIA ABERTA

A primeira proposta de acesso aberto ao conhecimento surgiu em 1971 por iniciativa de Michael Hart e o Projeto Gutenberg (*Gutenberg Project*). O Projeto Gutenberg foi a primeira iniciativa que visava possibilitar o acesso livre e amplo a obras de domínio público ou sob a permissão dos autores dos documentos, pioneiro na digitalização de obras impressas. A visão de Hart lançou a pedra fundamental para o que seria no futuro o Movimento de Acesso Aberto. Em 1989, foi lançada por Stevan Harnad, a primeira revista de acesso aberto (VANZ; COSTA FILHO, 2019).

O “Movimento de Acesso Aberto” (*Open Access Movement – A Movement*) constitui uma série de iniciativas, formalizadas através de declarações que buscam fomentar projetos para o livre acesso aos conhecimentos gerados pela humanidade. O termo Open Access foi utilizado pela primeira vez na Reunião Budapest *Open Access Initiative* (BOAI) em 2002, e se relacionava com a disponibilização de publicações científicas.

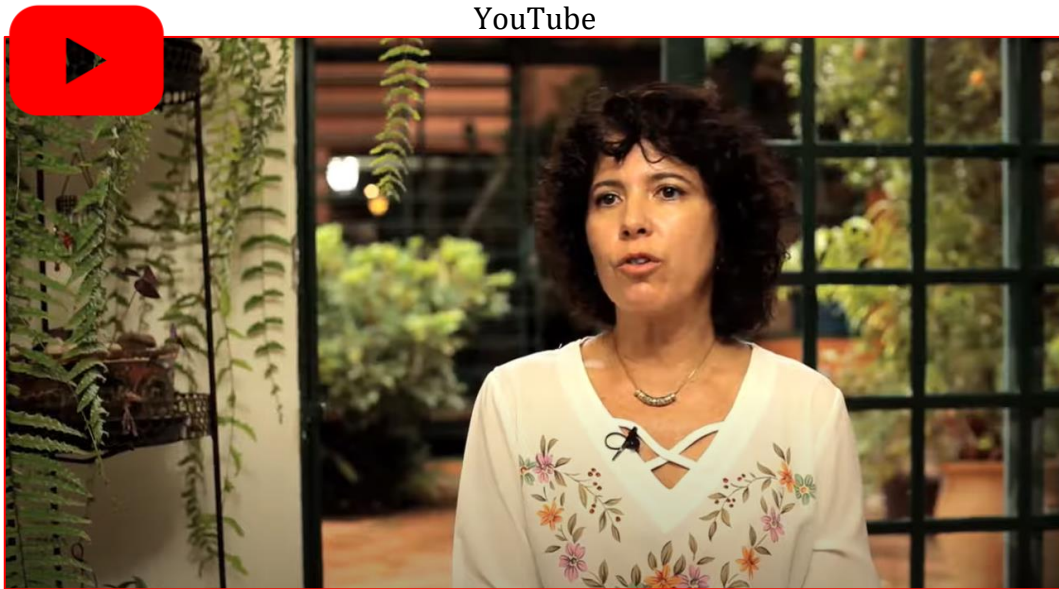
Figura 6 - pesquisa & ciência ABERTA



Fonte: Jamk University of Applied Sciences (201[-]).

No Brasil, o movimento de acesso aberto teve grande repercussão devido principalmente, a duas iniciativas: o lançamento da SciELO em 1999 e a tradução do software para gestão de revistas científicas, o *Open Journal System* (OJS) feita pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) em 2003.

Figura 7 - Ciência Aberta - um pouco da história - Sarita Albagli | vídeo no YouTube



Fonte: Ciência Aberta Ubatuba (2017).

A Ciência Aberta chega à sociedade com práticas e ferramentas que não somente divulgam e estimulam a participação social, como também propõem uma forma mais colaborativa de fazer ciência: “Trata-se de refletir sobre os desafios que essas mudanças trazem às dinâmicas [...], seus valores e práticas, e sobre os novos olhares que se impõem para melhor compreender e lidar com tais desafios.” (ALBAGLI; MACIEL; ADBO, 2015, p. 9).

Ciência aberta busca tornar o conhecimento científico acessível a todos, promovendo a transparência e a colaboração em todas as etapas da pesquisa. Isso inclui o acesso aberto a artigos científicos, dados e *software*, bem como o uso de práticas transparentes e colaborativas em todas as etapas do processo de pesquisa.

A ciência aberta tem como objetivo maximizar o impacto e a eficiência da pesquisa científica, permitindo que cientistas de todo o mundo compartilhem informações e trabalhem juntos para resolver problemas globais. Ao tornar a pesquisa mais acessível e transparente, a ciência aberta também pode ajudar a promover a confiança na pesquisa científica e a aumentar a sua relevância para a sociedade em geral.

Para a Fiocruz, Ciência Aberta:

[...] um movimento que busca refletir sobre atividades, processos, métodos, avaliações e produções científicas, visando sua transparência, colaboração e abertura. Envolve diversos movimentos como Acesso Aberto, Dados Abertos, Recursos Educacionais Abertos, Código Aberto, Hardware Aberto, Caderno de Laboratório Aberto, Revisão por Pares Aberta, Redes Sociais Científicas e Ciência Cidadã. Cada um desses grandes temas formam o ecossistema da Ciência Aberta. (FIOCRUZ, 2022, não paginado).

Figura 8 - Ciência Aberta

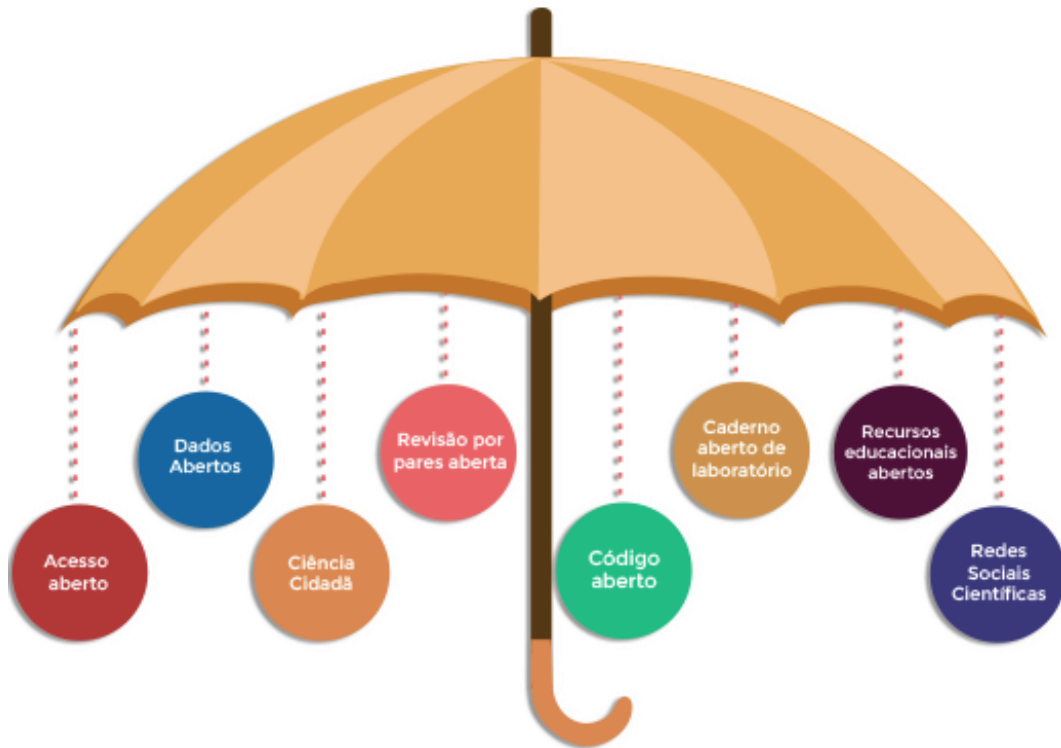


Fonte: UNESCO (2022).

A Ciência Aberta corrobora com a produção colaborativa, interativa e compartilhada da informação, refletindo em uma nova práxis científica. Portanto:

A Ciência Aberta é um movimento que incentiva a transparência da pesquisa científica desde a concepção da investigação até o uso de softwares abertos [...] e gestão de dados científicos, para que estes possam ser distribuídos, reutilizados e estar acessíveis a todos os níveis da sociedade, sem custos. (SILVA; SILVEIRA, 2019, p. 2).

Figura 9 - “Guarda-chuva” da Ciência Aberta



Fonte: FIOCRUZ (2019).

Dentro do “guarda-chuva” da Ciência Aberta, estão contemplados diversos conceitos, explicados abaixo:

Acesso aberto

Disponibilização na internet de literatura de caráter acadêmico ou científico, permitindo a qualquer usuário ler, baixar (fazer *download*), copiar, distribuir, imprimir, pesquisar ou referenciar (fornecer *link*) o texto integral dos documentos.

Dados abertos

Movimento de abertura dos dados visando sua transparência e processos colaborativos de coleta e reuso. Busca refletir sobre atividades, processos, métodos, avaliações e produções científicas.

Revisão
por pares
aberta

Metodologia que se vale da contribuição direta de cidadãos com recursos, cognição, dados e observações sobre objetos e fatos científicos.

Ciência
Cidadã

Cobre uma gama de práticas ou modelos de revisão por pares que visam aumentar a transparência, eficiência e responsabilidade do processo de revisão. A abertura da revisão por pares, na verdade, pode significar uma ampla variedade de interações entre revisores, editores e autores, resultando em muitas combinações.

Código
aberto

Linguagem que permite que todos os aplicativos desenvolvidos sob essa perspectiva sejam reescritos ou aprimorados de forma colaborativa. Assim, é possível criar novas funcionalidades ou corrigir *bugs* e inconsistências.

Caderno
aberto de
laboratório

São aplicações e/ou software que registram todas as etapas de uma pesquisa, contendo dados e informações essenciais para seu desenvolvimento.

Recursos
educacionais
abertos

São materiais, digitais ou não, disponíveis para reter, reutilizar, remixar, reformular e recompartilhar. Estão licenciados de maneira aberta.

Redes
sociais
científicas

São tipos específicos de redes sociais que representam as interações sociais e envolvem diferentes tipos de colaborações ou interações científicas oriundas do meio acadêmico.

2.1 e-SCIENCE

e-Science é um termo usado para descrever a prática de fazer ciência em um ambiente totalmente digital, usando tecnologias de informação e comunicação para conduzir e gerenciar a pesquisa. A *e-Science* é um elemento fundamental da Ciência Aberta, pois muitas das ferramentas e tecnologias necessárias para promover a abertura e a transparência na pesquisa científica são digitais (FABREGAS; KREMER; SCHILBACH, 2019).

A *e-Science* envolve o uso de ferramentas de computação, como bancos de dados, redes de computadores, sistemas de gerenciamento de informações científicas e análise de dados em larga escala, para coletar, gerenciar e analisar dados de pesquisa. A Ciência Aberta pressupõe uma prática científica colaborativa e contributiva na abertura dos dados de pesquisa (MEDEIROS; CAREGNATO, 2012). As práticas científicas e a infraestrutura para apoiar o fazer científico estão no escopo da *e-Science* (FERREIRA, 2018).

2.2 DATA MINING

É um processo de pesquisa para obter informações de alta qualidade com base em insights e padrões do corpus do texto. O processo consiste basicamente em 3 etapas: exploração, construção de modelo ou definição do padrão, e validação/verificação. Algumas diretrizes gerais para mineração de dados são:

- encontre corpus textual relevante;
- colete os dados sem burlar licenças e leis de direitos autorais;
- converta os dados para o formato desejado;
- use ferramentas que forneçam suporte ao *Data Mining*;

- cumpra os requisitos dos agentes financiadores e dos editores em relação ao compartilhamento de dados;
- siga as boas práticas para organizar, documentar e armazenar dados brutos e processados.

No contexto da Ciência Aberta, a mineração de dados pode ser utilizada para explorar conjuntos de dados abertos e identificar padrões que possam ser relevantes para a pesquisa científica (HAN; KAMBER; PEI, 2011).

A mineração de dados pode ser usada para extrair informações úteis de diversos tipos de dados abertos, como bancos de dados genômicos, dados climáticos, dados socioeconômicos, entre outros. Os pesquisadores podem usar técnicas de mineração de dados para encontrar relações complexas entre variáveis, identificar padrões de comportamento ou identificar anomalias que possam exigir mais investigação (FIA..., 2020).

MINERAÇÃO DE DADOS TEXTUAIS

Aprendendo sobre Dados de Pesquisa, n. 5, 2023

Fonte: traduzido e adaptado de *Data Literacy Series*, n.3/2022 (UC Santa Barbara Library)

RDPBRASIL
Rede de Dados de Pesquisa
ufrgs.br/rdpbrasil

Quer saber mais?



Mineração de dados textuais

Text Data Mining (TDM) é um processo de pesquisa para obter informações de alta qualidade com base em *insights* e padrões do *corpus* do texto.

Modelo básico para mineração de dados textuais



Diretrizes gerais para mineração de dados

- Encontre corpus textual relevante
- Colete os dados sem infringir licenças e leis de direitos autorais
- Como você pode obter os dados no formato desejado?
- Converta os dados para o formato desejado
- Utilize ferramentas que ofereçam suporte ao TDM (sintáticas ou semânticas)
- Cumpra os requisitos dos agentes financiadores e dos editores quanto ao compartilhamento de dados
- Siga as boas práticas para organizar, documentar e armazenar dados brutos e processados



licenciado por *Creative Commons*
Atribuição-NãoComercial-Compartilhável



2.3 BIG DATA

Com os avanços em tecnologia da informação, a capacidade das organizações, grupos e indivíduos de gerar dados tem se tornado cada vez maior, trazendo oportunidades ímpares e, ao mesmo tempo, impondo dificuldades à organização e preservação desses grandes volumes de dados. Este fenômeno é amplamente discutido na atualidade, sob a designação *Big Data*.

Big data pode ser entendida como o crescimento complexo e dinâmico de dados gerados por pessoas, máquinas e recursos tecnológicos (DANIEL, 2020). “Big data consiste em análise, tratamento e obtenção de informações contidas em grandes conjuntos de dados (SILVA; RIBEIRO, 2002).”

Figura 10 - O que é *Big Data* e para que serve? | vídeo no YouTube



Fonte: Faculdade Cásper Líbero (2018).

MÓDULO III

Objetivos

Visa entender todos os elementos que compõem o “Ecosistema dos Dados de Pesquisa”.

Conceitos

Ecosistema de Dados de Pesquisa.

3 ECOSSISTEMA DE DADOS DE PESQUISA

Um ecossistema de dados de pesquisa é um conjunto de infraestruturas, serviços e políticas que suportam a coleta, o gerenciamento, o compartilhamento e o reuso de dados de pesquisa. Para Oliveira e Lóscio (2019, p. 74): “[...] Ecosistema de Dados é composto por um conjunto de atores que se relacionam entre si por meio da troca de um conjunto de recursos distribuídos, heterogêneos, dinâmicos e em evolução.”. Esses ecossistemas são um elemento central da ciência aberta, pois permitem que os pesquisadores compartilhem seus dados com outros pesquisadores e também com o público em geral, aumentando a transparência e a reprodutibilidade da pesquisa.

Para Pollock:

Um ecossistema tem ciclos de dados, nos quais consumidores e intermediários de dados (e.g., desenvolvedores de aplicativos) podem compartilhar seus dados limpos, integrados e empacotados no ecossistema de forma reutilizável. Geralmente, esses dados limpos e integrados são mais valiosos do que a fonte original. (POLLOCK, 2011, não paginado).

Um ecossistema de pesquisa envolve a colaboração e interação entre diferentes atores que desempenham papéis específicos. Para Ponte, Mierzejewska e Klein (2017), alguns dos principais atores de um ecossistema de pesquisa incluem:

PESQUISADORES



São os principais produtores de conhecimento em um ecossistema de pesquisa. Eles conduzem experimentos, coletam e analisam dados, e publicam resultados em revistas científicas.

INSTITUIÇÕES DE PESQUISA



Essas instituições fornecem suporte e recursos para os pesquisadores, incluindo financiamento, instalações de laboratório, e acesso a equipamentos e tecnologia.

MERCADO EDITORIAL



São responsáveis pela revisão, edição e publicação de artigos científicos em revistas especializadas.

POLÍTICAS PÚBLICAS



Governos e outras organizações podem usar a pesquisa científica para informar políticas públicas, programas e projetos.

AGÊNCIAS DE FOMENTO



Organizações que fornecem financiamento para pesquisas. Isso pode incluir agências governamentais, fundações privadas, empresas e indivíduos.

REDES SOCIAIS ACADÊMICAS E NÃO ACADÊMICAS



São responsáveis pela disseminação de conhecimento científico para o público em geral. Isso pode incluir museus, centros de ciência, canais de TV, revistas e jornais.

SOCIEDADE CIVIL



Além de leitores, cidadãos comuns, políticas científicas, esferas institucionais de decisão política na Ciência, Tecnologia e Inovação etc.

3.1 TIPOS E CATEGORIAS DE DADOS DE PESQUISA

As tecnologias e as ferramentas digitais levam os dados de pesquisa a novas formas de aplicação e vêm sendo usados como recurso de validação das pesquisas científicas (BORGMAN, 2015).

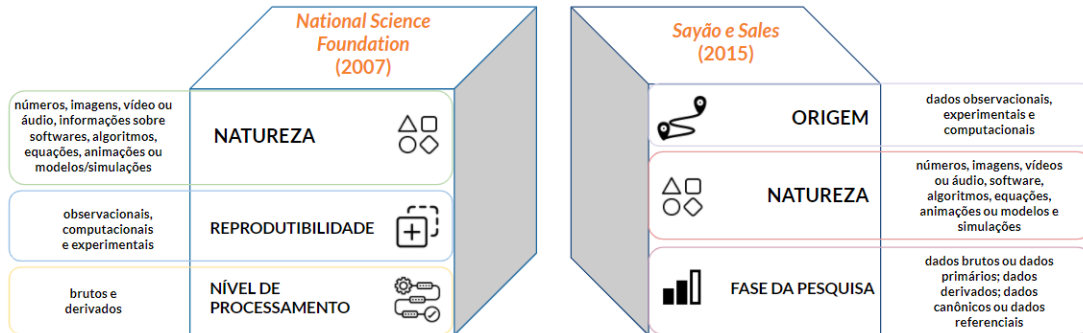
Um conceito amplamente utilizado quando se trata de dados de pesquisa, tanto no Brasil quanto internacionalmente, é o da OCDE (2007), que define dados de pesquisa como:

[...] 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 os resultados da pesquisa. Um conjunto de dados de pesquisa constitui uma representação sistemática e parcial do assunto que está sendo investigado. (OCDE, 2007, p. 13).

Para a Fiocruz (FUNDAÇÃO..., 2022, não paginado), dados de pesquisa: “São os objetos informacionais digitais comumente registrados e aceitos pela comunidade científica como necessários para validar os resultados de pesquisa, sendo dados produzidos para diversos fins e utilizados em pesquisas.”. Observa-se, entretanto, que a amplitude do entendimento de dados de pesquisas sugere um conceito complexo que se manifesta em múltiplas formas.

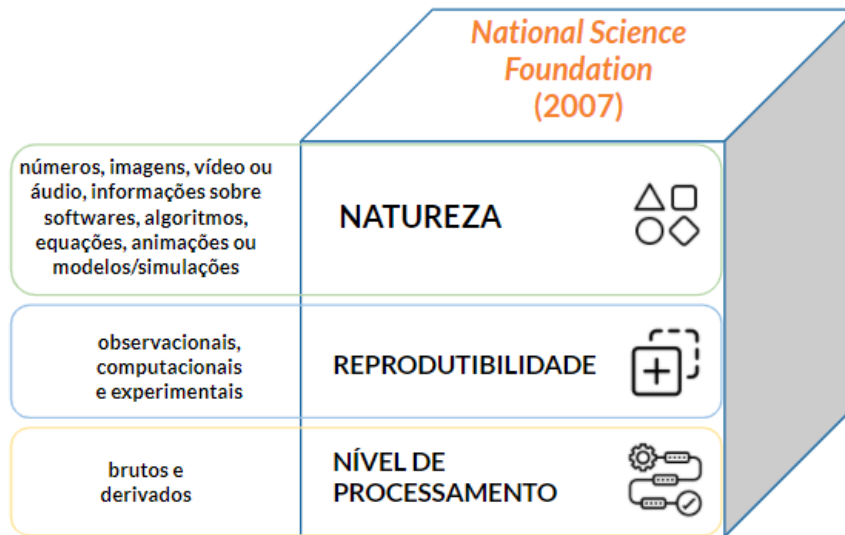
Uma característica importante é que os dados de pesquisa se distinguem de acordo com diferentes critérios e são classificados de acordo com o procedimento de coleta, caráter primário ou secundário, grau de estruturação, nível de cobertura e formato. A seguir apresentam-se algumas classificações feitas a partir do ponto de vista de diferentes autores e instituições.

3.2 CLASSIFICAÇÃO DE DADOS



Classificação de dados de acordo com a *National Science Foundation* (2007):

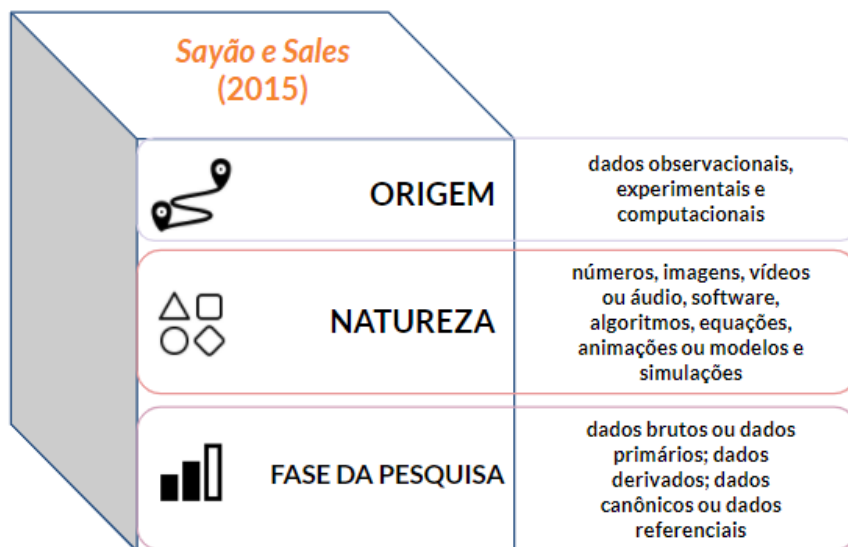
Figura 11 - Classificação de dados de acordo com a *National Science Foundation*



Fonte: adaptado de *National Science Foundation* (2007).

Classificação de dados de acordo com Sayão e Sales (2015):




Figura 12 - Classificação de dados de acordo com Sayão e Sales



Fonte: adaptado de Sayão e Sales (2015).

3.3 CLASSIFICAÇÃO DOS DADOS SEGUNDO A SUA ORIGEM

Quadro 1 - Classificação dos dados segundo a sua origem

TIPO DE DADO	CONCEITO	EXEMPLO
 <p>DADOS OBSERVACIONAIS</p>	<p>São dados obtidos por meio de observações diretas, que podem ser associadas a lugares e tempo específicos e não podem ser coletados uma segunda vez.</p>	<p>https://www.redape.dados.embrapa.br/dataset.xhtml?persistentId=doi:10.48432/SZDYZA</p>
 <p>DADOS COMPUTACIONAIS</p>	<p>São resultados da execução de modelos computacionais ou de simulações.</p>	<p>https://www.redape.dados.embrapa.br/dataset.xhtml?persistentId=doi:10.48432/HTKLJ4</p>
 <p>DADOS EXPERIMENTAIS</p>	<p>São provenientes de situações controladas em bancadas de laboratórios.</p>	<p>https://www.redape.dados.embrapa.br/dataset.xhtml?persistentId=doi:10.48432/KG02AD</p>



Fonte: National Science Foundation (2007).

3.4 CLASSIFICAÇÃO DOS DADOS SEGUNDO A SUA NATUREZA

Números, imagens, vídeos ou áudios, *software*, algoritmos, equações, animações ou modelos e simulações.

3.5 CLASSIFICAÇÃO DOS DADOS SEGUNDO A FASE DA PESQUISA

Quadro 2 - Classificação dos dados segundo a fase da pesquisa

TIPO DE DADO	CONCEITO	EXEMPLO
 <p>DADOS BRUTOS ou PRELIMINARES</p>	São dados que resultam diretamente da pesquisa científica.	https://arcadados.fiocruz.br/file.xhtml?persistentId=doi:10.35078/EPSXXL/5COPG7
 <p>DADOS DERIVADOS</p>	Resultam do processamento de dados brutos ou da combinação destes com outros dados.	https://arcadados.fiocruz.br/dataset.xhtml?persistentId=doi:10.35078/NNGKDX
 <p>DADOS CANÔNICOS ou DADOS REFERENCIAIS</p>	São coleções de dados consolidados e arquivados geralmente em grandes centros de dados, por exemplo, sequência genética, estrutura química, etc.	https://www.redape.dados.embrapa.br/dataset.xhtml?persistentId=doi:10.48432/J9YG00
 <p>DADOS DE REGISTROS</p>	São dados gerados por empresas públicas ou privadas e que também são úteis para a pesquisa científica, social e humanística.	https://arcadados.fiocruz.br/dataset.xhtml?persistentId=doi:10.35078/Q06VZB

Fonte: *National Science Foundation* (2007).

MÓDULO IV

Objetivos

Objetiva entender os componentes dos Repositórios de Dados, seus tipos e exemplos.

Conceitos

Repositórios de Dados.

4 REPOSITÓRIOS DE DADOS, TIPOS E EXEMPLOS

O termo repositório já é bastante conhecido pela comunidade científica. Ele se tornou popular com o lançamento dos repositórios de publicações, destinados ao armazenamento e preservação de documentos como artigos, livros, trabalhos de conclusão de curso, teses e dissertações, entre outros objetos digitais. Com o avanço do movimento da ciência aberta e demanda pelo compartilhamento de dados, surge o repositório destinado a este objeto digital – o dado de pesquisa.

Repositórios de dados são plataformas *online* que permitem que os pesquisadores armazenem, gerenciem e compartilhem seus dados de pesquisa com outros pesquisadores e o público em geral (VIDOTTI *et al.*, 2017).

Os Repositórios de Dados (RDs) são grandes infraestruturas de banco de dados que reúnem, organizam, gerenciam, preservam e disponibilizam conjuntos de dados para pesquisa.

Figura 13 - Repositórios de dados de pesquisa | vídeo no YouTube










Fonte: RDP Brasil / IBICT (2021).





4.1 BENEFÍCIOS DOS REPOSITÓRIOS DE DADOS DE PESQUISA

Os repositórios já demonstraram sua importância e benefício há alguns anos, para a disponibilização das publicações científicas. Sales (2014) apresenta alguns dos benefícios mais perceptíveis para os repositórios de dados:

Quadro 3 - Benefícios dos repositórios de dados

VISIBILIDADE DOS DADOS	COMPARTILHAMENTO DE DADOS 	CRÉDITO AO AUTOR DOS DADOS 	PRESERVAÇÃO DIGITAL 
<p>Amplia a visibilidade dos dados de pesquisa permitindo que eles sejam consultados e citados mais frequentemente (geralmente só é disseminada pelos canais formais a fração de dados – cerca de 10% - que está registrada nos artigos publicados).</p>	<p>Os repositórios pela sua capacidade de agregação e organização de recursos informacionais dispersos no tempo e no espaço, e como instrumento de socialização de comunidades e grupos pesquisadores ao redor desses recursos, tornam-se um dispositivo importante de troca de experiências e compartilhamento de dados.</p>	<p>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.</p>	<p>Oferece um ambiente tecnológico, gerencial e de padronização propício para a preservação de longo prazo dos dados de pesquisa de valor contínuo, especialmente para os dados observacionais.</p>

<p>MEMÓRIA CIENTÍFICA E TRANSPARÊNCIA</p> 	<p>SEGURANÇA DOS DADOS</p> 	<p>DISPONIBILIDADE</p> 	<p>CURADORIA DIGITAL</p> 
<p>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; na qualidade de registro das atividades de pesquisa das instituições, contribui também com os princípios de transparência, tão em voga nos tempos atuais.</p>	<p>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.</p>	<p>Permite que os dados estejam disponíveis on-line para serem acessados, baixados, visualizados e processados por pessoas ou por sistemas.</p>	<p>Proporciona um ambiente apropriado para os processos de avaliação, de adição de valor, reformatação, agregação e recriação de dados promovidos pela curadoria digital.</p>

<p>SERVIÇOS INOVADORES</p> 	<p>REUSO DOS DADOS</p> 	<p>REDES DE REPOSITÓRIOS</p> 	<p>INDICADOR DE QUALIDADE E PRODUTIVIDADE DA INSTITUIÇÃO</p> 
<p>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.</p>	<p>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.</p>	<p>Permite por meio de protocolos de interoperabilidade, como o OAI-PMH, a formação de redes de repositório de dados; abre a possibilidade de inserção dos repositórios de dados às redes interoperáveis definidas pelo padrão <i>Linked Data</i>.</p>	<p>As coleções de dados organizadas e arquivadas no repositório são evidências da qualidade e da relevância das atividades de pesquisa da instituição, atestando a sua produtividade e seu valor acadêmico.</p>

Fonte: Sales (2014).

4.2 TIPOS DE REPOSITÓRIOS

Repositórios Institucionais: são mantidos e gerenciados por uma instituição acadêmica, como universidades ou institutos de pesquisa, e são voltados para arquivar publicações e/ou dados que são, geralmente, provenientes unicamente das atividades acadêmicas dessas instituições.

Exemplos:

PUBLICAÇÃO CIENTÍFICA	REPOSITÓRIO DE DADOS
 <p style="text-align: center;">ALICE</p> <p style="text-align: center;">https://www.alice.cnptia.embrapa.br/alice/</p>	 <p style="text-align: center;">FULIR</p> <p style="text-align: center;">https://data.fulir.irb.br/</p>
 <p style="text-align: center;">ARCA</p> <p style="text-align: center;">https://www.arca.fiocruz.br/</p>	 <p style="text-align: center;">YODA</p> <p style="text-align: center;">https://www.uu.nl/en/research/yoda</p>

Multidisciplinares: repositórios institucionais também podem ser multidisciplinares ao armazenar dados provenientes das diversas vertentes de pesquisa da instituição.

Exemplos:

PUBLICAÇÃO CIENTÍFICA	REPOSITÓRIO DE DADOS
 <p data-bbox="472 651 560 683">LUME</p> <p data-bbox="320 706 715 742"><i>Repositório Digital da UFRGS</i></p> <p data-bbox="320 761 715 798">https://www.lume.ufrgs.br/</p> 	 <p data-bbox="1027 651 1102 683">PAIN</p> <p data-bbox="852 706 1278 798"><i>Pain and Interoception Imaging Network</i></p> <p data-bbox="799 821 1331 891">https://www.painrepository.org/repositories/</p> 




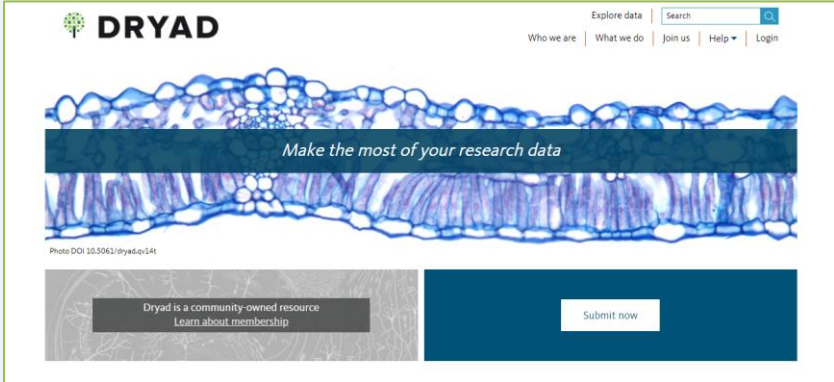
Repositórios Disciplinares: dedicados ao arquivamento de domínios específicos de pesquisa como por exemplo, física de partículas ou ciências ambientais.

Exemplos:

PUBLICAÇÃO CIENTÍFICA	REPOSITÓRIO DE DADOS
 <p>CERN <i>Document Server</i> https://cds.cern.ch/</p> 	 <p>GEOINFO <i>Infraestrutura de dados espaciais da EMBRAPA</i> http://geoinfo.cnps.embrapa.br/</p> 

Repositórios Multidisciplinares: reúnem coleções de dados coletados ou gerados por atividades de pesquisa em várias áreas de conhecimento.

Exemplos:

PUBLICAÇÃO CIENTÍFICA	REPOSITÓRIO DE DADOS
 <p>BASE <i>Bielefeld Academic Search Engine</i></p> <p>https://www.base-search.net/</p>	 <p>HARVARD Dataverse Support</p> <p>Harvard <i>Dataverse Support</i></p> <p>https://support.dataverse.harvard.edu u/</p>
 <p>DRYAD</p> <p>DRYAD</p> <p>https://datadryad.org/stash</p> 	

Repositórios Orientados por projetos: são voltados para coleções de dados resultantes de projetos de pesquisa ou resolução de problemas específicos.

Exemplo:




Como escolher um repositório de dados?

Aprendido sobre Dados de Pesquisa, n. 3, 2023

Fonte: traduzido e adaptado de *Data Literacy Series*, n.5/2021 (UJ Santa Barbara Library)

RDPBRASIL
Rede de Dados de Pesquisa


Quer saber mais?
ufrgs.br/rdpbrasil



Como escolher um Repositório de Dados?


Os Repositórios de Dados (RDs) são grandes infraestruturas de banco de dados que reúnem, organizam, gerenciam, preservam e disponibilizam conjuntos de dados para pesquisa.

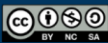
10 ITENS A CONSIDERAR AO ESCOLHER UM REPOSITÓRIO DE DADOS


 **SERÁ QUE O REPOSITÓRIO...**

- 1 Foi recomendado por sua instituição, editoras e/ou agências de financiamento?
- 2 Atribui identificadores persistentes e únicos aos conjuntos de dados?
- 3 Fornece uma página de destino para os conjuntos de dados, com metadados que ajudem outros pesquisadores a encontrar dados, os tornem compreensíveis, visíveis e promovam a reutilização dos dados?
- 4 Acompanha como os dados têm sido utilizados, fornecendo estatísticas de acesso e *download*?
- 5 Segue um modelo sustentável e é certificado como confiável, com um compromisso explícito de preservar e disponibilizar dados a longo prazo?
- 6 Segue padrões que correspondem às suas necessidades particulares de dados (por exemplo, formatos aceitos, padrões de metadados, capacidade de tamanho)?
- 7 Fornece orientação sobre como citar os dados que foram depositados?
- 8 Oferece termos, licenças e condições claras que satisfaçam os requisitos legais (por exemplo, proteção de dados e restrições)?
- 9 Oferece serviços de curadoria e revisão de dados e documentação antes da publicação dos dados?
- 10 Cobra por seus serviços?

Ainda inseguro? Entre em contato conosco.

 Nós podemos ajudá-lo a encontrar e selecionar o repositório de dados mais apropriado para sua pesquisa.


licenciado por Creative Commons
Atribuição- NãoComercial- Compartilhavel



REFERÊNCIAS

Módulo I

ALMEIDA, Fernanda Gomes. **Suporte à gestão de dados de pesquisa: uma ampliação dos serviços oferecidos pelas bibliotecas**. 2019. 249 f. Tese (doutorado em Gestão & Organização do Conhecimento) – Programa de Pós-Graduação em Gestão & Organização do Conhecimento, Escola de Ciência da Informação, Universidade Federal de Minas Gerais, 2019. Disponível em: <https://repositorio.ufmg.br/handle/1843/31596>. Acesso em: 22 maio 2023.

BALL, Alex. **Review of data management lifecycle models**. Bath: University of Bath, 2012. Disponível em: <https://purehost.bath.ac.uk/ws/portalfiles/portal/206543/redm1rep120110ab10.pdf> Acesso em: 9 maio 2023.

CÓRDULA, Flavio Ribeiro; ARAÚJO, Wagner Junqueira de. **O compartilhamento de dados científicos na era do e-science**. In: DIAS, Guilherme Ataíde; OLIVEIRA, Bernardina Maria Juvenal Freire de. (Orgs.). **Dados científicos: perspectivas e desafios**. João Pessoa: Editora UFPB, 2019. p. 189-207. Disponível em: https://www.researchgate.net/profile/Wagner-Araujo/publication/339566217_O_compartilhamento_de_dados_cientificos_na_era_do_e-science/links/5e8bd149a6fdcca789fbdfc0/O-compartilhamento-de-dados-cientificos-na-era-do-e-science.pdf. Acesso em: 9 maio 2023.

CURTY, R. Abordagens de reuso e a questão da reusabilidade dos dados científicos. **Liinc Em Revista**, v. 15, n. 2, 2019. Disponível em: <https://doi.org/10.18617/liinc.v15i2.4777>. Acesso em: 16 abr. 2023.

GABRIEL JÚNIOR, René Faustino et al. Acesso aberto a dados de pesquisa no Brasil: mapeamento de repositórios, práticas e percepções dos pesquisadores e tecnologias. **Ci.Inf.**, Brasília, DF, v.48 n.3 (Supl.), p.87-101, set./dez. 2019. Disponível em: <https://lume.ufrgs.br/handle/10183/212266>. Acesso em: 3 maio 2023.

MEDEIROS, J. da S.; CAREGNATO, S. E. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, [S. l.], v. 8, n. 2, 2012. Disponível em: <https://revista.ibict.br/liinc/article/view/3368>. Acesso em: 22 maio 2023.

NIU, J. **Perceived Documentation Quality of Social Science Data**. University of Michigan, Ann Arbor, 2009. 132 f. Thesis (Doctor of Philosophy - Information) University of Michigan, Horace H. Rackham School of Graduate Studies, 2009. Disponível em: <https://deepblue.lib.umich.edu/handle/2027.42/63871?show=full>. Acesso em: 28 abr. 2023.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação & Informação**, Londrina, v. 21, n. 2, p. 116-142, maio/ago. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940>. Acesso em: 22 maio 2023.

SAYÃO, Luis Fernando; SALES, Luana. Afinal, o que é dado de pesquisa? **Biblos**: Rio Grande, v. 34, n. 02, p. 32-51, jul./dez. 2020. Disponível em: <https://periodicos.furg.br/biblos/article/view/11875>. Acesso em: 27 mar. 2023.

TARGINO, M. G. Comunicação científica: uma revisão de seus elementos básicos. **Informação e Sociedade Estudos**, João Pessoa, v. 10, n. 2, p. 37-85, 2000. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/326>.

UK DATA SERVICE. Research Data Lifecycle. 2019. Disponível em: <https://ukdataservice.ac.uk/learning-hub/research-data-management/>. Acesso em: 18 abr. 2023.

VEIGA, Viviane Santos de Oliveira; SILVA, Cícera Henrique da; BORGES, Maria Manuel. **Modelo de fatores que influenciam no comportamento de compartilhamento de dados de pesquisa (MFDados)**. In: BORGES, Maria Manuel; CASADO, Elias Sanz (org.). Sob a lente da Ciência Aberta: olhares de Portugal, Espanha e Brasil. Coimbra: Imprensa da Universidade de Coimbra. 2021. p. 153-187. Disponível em: <https://www.arca.fiocruz.br/handle/icict/46052>. Acesso em: 22 maio 2023.

Módulo II

ALBAGLI, S.; MACIEL, M. L.; ADBO, A. H. (org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: Unirio, 2015. Disponível em: [https://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20\(5\).pdf](https://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20(5).pdf). Acesso em: 021 ago. 2022.

BERTIN, Patrícia Rocha Bello; VISOLI, Marcos Cezar; DRUCKER, Debora Pignatari. A gestão de dados de pesquisa no contexto da e-Science: benefícios, desafios e oportunidades para organizações de P&D. **Ponto de Acesso**, Salvador, v. 11, n. 2, p. 34-48, ago. 2017. Disponível em: <https://periodicos.ufba.br/index.php/revistaici/article/view/21449/15200>. Acesso em: 22 maio 2023.

DANIEL, B. K. Big Data e ciência de dados: uma revisão crítica de questões para a pesquisa educacional. **PerCursos**, Florianópolis, v. 21, n. 45, p. 80-103, jan./abr. 2020. Disponível em: <https://periodicos.ufrn.br/informacao/article/view/22452/13544>. Acesso em: 28 abr. 2023.

DE MAURO, A.; GRECO, M.; GRIMALDI, M. A Formal Definition of Big Data based on its Essential Features. **Library Review**, v. 65, n. 3, p.122-135, Apr. 2016. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/LR-06-2015-0061/full/html>. Acesso em: 28 abr. 2023.

FABREGAS, R.; KREMER, M.; SCHILBACH, F. Realizing the potential of digital development: the case of agricultural advice. **Science**, v. 366, n. 6471, 2019. Disponível em: 10.1126/science.aay303. Acesso em: 07 abr. 2023.

FERREIRA, V. B. **e-Science e políticas públicas**: ciência, tecnologia e inovação no Brasil. Salvador: EDUFBA, 2018. v. 1. 256p. Disponível em: <https://books.scielo.org/id/bc84k/pdf/ferreira-9788523218652-03.pdf>. Acesso em: 07 abr. 2023.

FIA BUSINESS SCHOOL. Data Mining: o que é, para que serve e tipos de técnicas. 2020. Disponível em: <https://fia.com.br/blog/data-mining/>. Acesso em: 07 abr. 2023.

FUNDAÇÃO OSWALDO CRUZ. **Ciência aberta**. Disponível em: <https://portal.fiocruz.br/glossario-da-ciencia-aberta>. Acesso em: 20 mar. 2023.

FUNDAÇÃO OSWALDO CRUZ. **O que é Ciência Aberta**. 2019. Disponível em: <https://bit.ly/2G80wXc>. Acesso em: 23 abr. 2023.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining**: Concepts and Techniques. San Francisco: Morgan Kaufmann Publishers. 3rd edition. 2011.

Laboratório de Suporte à Decisão. **Data Mining**. Disponível em: http://www.ufopa.edu.br/lsd/index.php?option=com_content&view=article&id=8:beginners&catid=19&Itemid=260. Acesso em: 22 maio 2023.

LORENZI, Larissa. **Principais diferenças entre Big Data, Data Science e Data Analytics**. Blog da Indicium - tudo sobre data science e analytics. Florianópolis, SC, 18 nov. 2020. Disponível em: https://blog.indicium.tech/diferencas-big-data-data-science-data-analytics/amp/?utm_source=Google&utm_medium=cpc&utm_term=&utm_campaign=19229929630&utm_content=&gclid=Cj0KCQjwpPKiBhDvARIsACn-gzC0zfdZYupsmj1qPr8-z3qahc5PLR-FE XFkbWwcsQPLPLhN7HKTMaAiFuEALw_wcB. Acesso em: 22 maio 2023.

MEDEIROS, J.; CAREGNATO, S. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, v. 8, n. 10, 2012. Disponível em: <https://doi.org/10.18617/liinc.v8i2.488>. Acesso em: 07 abr. 2023.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. OECD Principles and Guidelines for Access to Research Data from Public Funding. Paris: OECD, 2007. 24 p. Disponível em: <http://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 05 dez. 2022.

SILVA, F. C. C.; SILVEIRA, L. O ecossistema da Ciência Aberta. **Transinformação**, Campinas, v. 31, p. e190001, 2019.

TOOLKIT DE LA UNESCO PUEDE AYUDAR A ACELERAR LA TRANSICIÓN A LA CIENCIA ABIERTA MUNDIAL. *In*: Blog de la biblioteca de Traducción y Documentación de la Universidad de Salamanca.

UNIVERSO ABIERTO. Salamanca, Espanha, 11 abr. 2023. Disponível em: <https://universoabierto.org/2023/04/11/toolkit-de-la-unesco-puede-ayudar-a-acelerar-la-transicion-a-la-ciencia-abierta-mundial/>. Acesso em: 22 maio 2023.

UNESCO. **Recomendação da UNESCO sobre Ciência Aberta**. 2022. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000379949_por. Acesso em: 23 abr. 2023.

VANZ, S A. S.; SILVA FILHO, R. O protagonismo das revistas na comunicação científica: histórico e evolução. Capítulo no prelo. 2019.

Módulo III

BORGMAN, C. **Big data, little data, no data: scholarship in the networked world.** Cambridge, Massachusetts: MIT Press. 2015.

BORGMAN, C. L. **Research Data: who will share what, with whom, when and why.** (RatSWD Paper n. 161. Oct. 2010). Disponível em: http://sydney.edu.au/research/data_policy/resources/ANDS_Borgman_2010_research_data.pdf. Acesso em: 19 maio 2013.

FUNDAÇÃO OSWALDO CRUZ. **Ciência aberta.** Disponível em: <https://portal.fiocruz.br/glossario-da-ciencia-aberta>. Acesso em: 14 nov. 2022.

NATIONAL SCIENCE FOUNDATION. **Long-Lived Digital Data Collections: enabling research and education in the 21st century.** (2005). Disponível em: <http://www.nsf.gov/pubs/2005/nsb0540/>. Acesso em: 22 jan. 2023.

NATIONAL SCIENCE FOUNDATION - NSF. **Cyberinfrastructure vision for 21st century Discovery.** March 2007. Disponível em: http://escience.caltech.edu/workshop/CI_Vision_March07.pdf. Acesso em: 23 jan. 2023.

OLIVEIRA, M. I. S.; LÓSCIO, B. F. **Ecosistema de dados na web: da teoria ao desafio.** Fortaleza, CE, 2019. Disponível em: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/download/62/273/515-1?inline=1> Acesso em: 20 mar. 2023.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. OECD **Principles and Guidelines for Access to Research Data from Public Funding.** Paris: OECD, 2007. 24 p. Disponível em: <http://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 05 dez. 2022.

POLLOCK, R. **Building the (open) data ecosystem.** 2011. Disponível em: <https://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/>. Acesso em: 20 mar. 2023.

PONTE, D.; MIERZEJEWSKA, B. I.; KLEIN, S. The transformation of the academic publishing market: multiple perspectives on innovation. **Electronic Markets**, v. 27, n. 2, p. 97-100, 2017.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. Disponível em:
http://www.cnen.gov.br/images/CIN/PDFs/GUIA_DE_DADOS_DE_PESQUISA.pdf. Acesso em: 22 jan. 2023.

Módulo IV

AMORIM NETO, M. R. **Métrica em Repositórios**. 2021. Disponível em:
https://www.arca.fiocruz.br/bitstream/handle/icict/50567/Curso_Metrica_Repositorios.pdf;jsessionid=946B2842011A1B0E4354C7E34A998883?sequenc e=2. Acesso em: 16 abr. 2023.

OPEN Access USP. **Plataformas de repositório de dados: um primer**. 2021. Disponível em: <https://www.acessoaberto.usp.br/plataformas-de-repositorio-de-dados-um-primer/>. Acesso em: 16 abr. 2023.

RDP BRASIL. **Como escolher um repositório de dados?** 2023. Série: Aprendendo sobre Dados de Pesquisa, n. 3, 2023. Disponível em:
<https://www.ufrgs.br/rdpbrasil/wp/wp-content/uploads/2022/08/n.-3-2023.png>. Acesso em: 22 maio 2023.

SALES, L. F. **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.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, v. 21, n. 2, p. 90-115, 2016. DOI: 10.5433/1981-8920.2016v21n2p90 Acesso em: 16 abr. 2023.

VIDOTTI, S. A. B. G.; CONEGLIAN, C. S.; ROA-MARTÍNEZ, S. M.; ARAKAKI, F. A.; BRANDT, M. B.; FERREIRA, A. M. J. F. C. Repositório de dados de pesquisa para grupo de pesquisa: um projeto piloto. **Informação & Tecnologia**, v. 4, n. 2, p. 221-242, 2017. Disponível em:
<http://hdl.handle.net/20.500.11959/brapci/101623>. Acesso em: 07 abr. 2023.