



Guia de preparação de dados de pesquisa

Maio de 2025
Versão 1.1



This is an Open Access document distributed under the terms of the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly.

1. Preparação de dados de pesquisa	2
2. Anonimização de dados	4
2.1 Anonimização de transcrições	5
2.2 Anonimização de imagens	5
3. Formato dos arquivos	6
4. Nomeação dos arquivos	8
5. Descrição dos dados	8
6. Arquivo README	10
7. Tamanho dos arquivos	11
8. Alterações no manuscrito	11
Referências	12
Anexo 1. Fluxograma SciELO Data	13
Anexo 2. Formatos adequados para preservação	14

1. Preparação de dados de pesquisa

A adoção de boas práticas na preparação de dados para depósito é um aspecto de fundamental importância para o acesso, compartilhamento e reuso dos dados de pesquisa. Preparar os dados de forma adequada envolve organização e documentação das informações de modo a promover o acesso e reuso delas por outros pesquisadores causando impacto significativo no progresso da pesquisa e na transparência científica, além de contribuir para que eles sigam os Princípios FAIR e sejam tão FAIR quanto possíveis.¹

Dados FAIR são dados que atendem a padrões de encontrabilidade (*Findable*), acessibilidade (*Accessible*), interoperabilidade (*Interoperable*) e reusabilidade (*Reusable*). A curadoria de dados, fundamentada nos princípios FAIR, visa promover a fácil localização dos dados e os tornar mais acessíveis, interoperáveis e reutilizáveis. Para alcançar esses objetivos, é essencial que os dados sejam armazenados em repositórios adequados, com metadados claros e consistentes, facilitando a sua descoberta e acesso.

Para melhor entender os Princípios FAIR e como tornar os seus dados o mais FAIR possível recomendamos o uso da ferramenta FAIR-Aware: <https://fairaware.dans.knaw.nl/>.

Após a publicação dos conjuntos de dados é possível verificar o nível de FAIR (FAIRness) através da ferramenta Fuji: <https://f-uji.net/>.

Além dos princípios FAIR, é importante considerar as diretrizes éticas e culturais relacionadas aos dados de pesquisa, em particular tratando-se de informações que envolvem comunidades, como as indígenas, ou populações vulneráveis e seus direitos. Foi nesse contexto que desenvolveu-se os princípios CARE² (*Collective Benefit, Access, Respect, Ethics* - Benefício

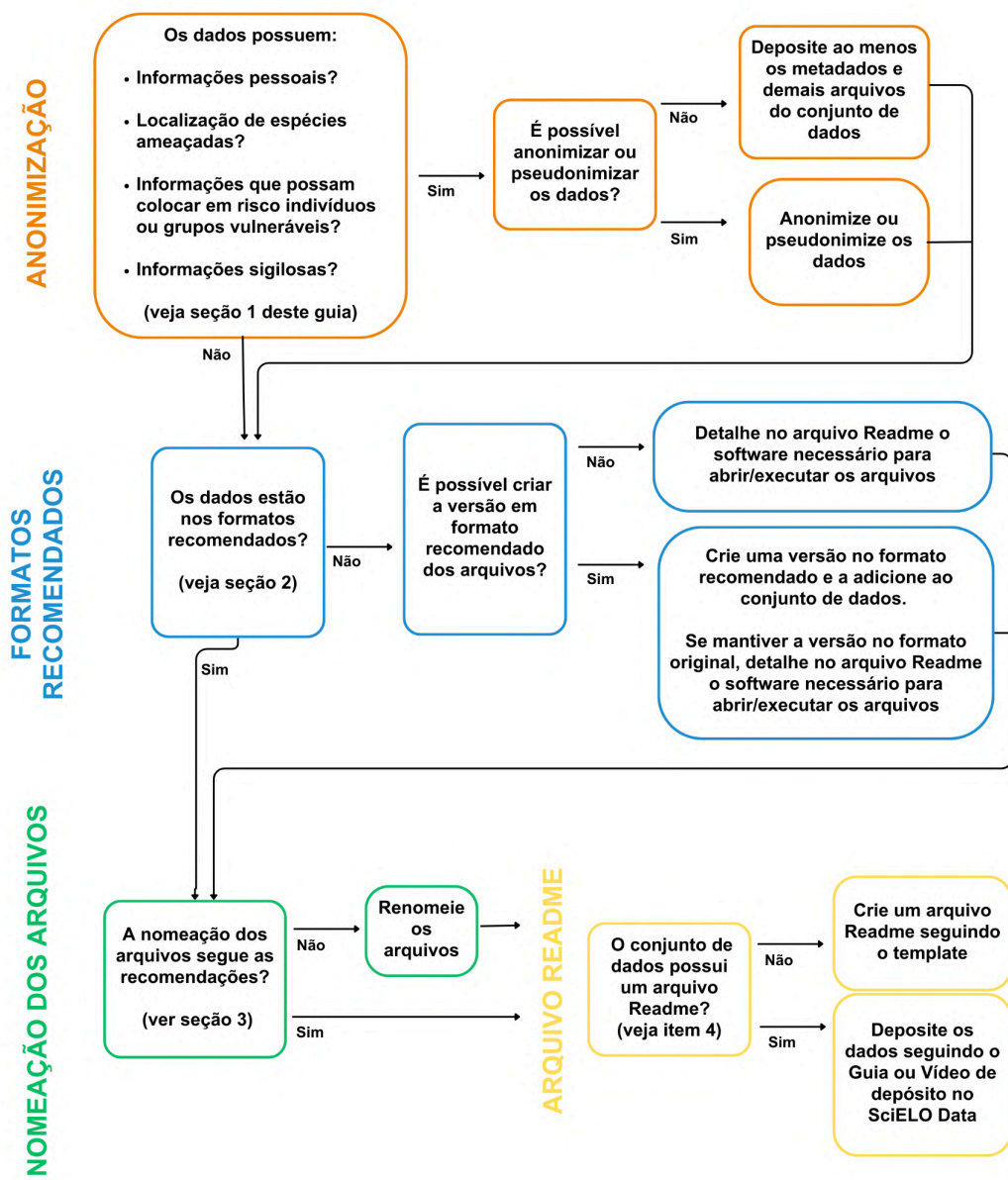
¹ Recomendamos o webinar "Cómo hacer que los datos sean FAIR? buenas prácticas para datos (abiertos) de investigación" disponível em: <https://www.youtube.com/watch?v=l14SwZxIRHY>.

²VIDOTTI, Silvana Aparecida Borsetti Gregorio; TORINO, Emanuelle; CONEGLIAN, Caio Saraiva. #SejaJUSTOeCUIDADOSO: princípios FAIR e CARE na gestão de dados de pesquisa. In: SALES, Luana Farias; VEIGA, Viviane dos Santos; HENNING, Patrícia; SAYÃO, Luís Fernando (org.). Princípios FAIR aplicados à gestão de dados de pesquisa. Rio de Janeiro: Ibict, 2021. p. 201 - 214. DOI: 10.22477/9786589167242.cap15

coletivo, acesso, respeito, ética) que buscam garantir que o uso de dados respeite as normas éticas e culturais.

A aplicação dos princípios CARE em conjunto com os princípios FAIR, ajuda a criar um ambiente de pesquisa mais justo, responsável e transparente, promovendo não apenas a reusabilidade dos dados, mas também seu uso ético e responsável.

Durante a preparação de seus dados de pesquisa para depósito em repositório, uma série de decisões e procedimentos devem ser realizados considerando as informações que estão nos dados, assim como o formato de seus arquivos e documentação relacionada. Abaixo segue um fluxograma de decisões que podem auxiliar no processo de adequação dos dados de pesquisa:



A versão em vídeo deste guia está disponível no Youtube: [Guia de depósito do SciELO Data](#). Para informações quanto ao fluxo de dados do SciELO Data, verifique o [Anexo 1](#).

2. Anonimização de dados

Devem ser obrigatoriamente anonimizados: Dados pessoais, sensíveis ou não³, informações que ultrapassem o direito de privacidade de pessoas envolvidas, ou as coloca em risco, assim como coordenadas de áreas protegidas, sob ameaça de extinção ou informações que infrinjam acordos comerciais, patentes ou pertençam a terceiros.

Reduza a presença de identificadores diretos nos arquivos que compõem o conjunto de dados para reduzir a precisão e detalhe das pessoas, locais ou informação que não possa ser identificada por meio de agregações como:

- Ano ou década de nascimento ao invés de data de nascimento precisa;
- Faixa etária ao invés de idade específica;
- Região ao invés de cidade;
- Localização urbana/rural ou geral (ex.: Zona Norte, Região Sul do município, prédio empresarial no centro da cidade, etc) ao invés do nome de locais;
- Ocupação ou área de especialização ao invés de cargo específico;
- Período de tempo ao invés de data ou horário específico.

Exemplo de anonimização de dados⁴:

Informação sem anonimização	Resposta sem anonimização
Nome	Juan Pérez
País de origem	Argentina
Idade	54
Anos de experiência	15
Modelo de aviões	Boeing 777/ 747
Data do último voo	05/01/2022

Informação anonimizada	Resposta anonimizada
-	-
Continente	América do Sul
Faixa Etária	50-60
Anos de experiência	10-20
Modelo de aviões	Comerciais
Data do último voo	01/2022

³ Dados pessoais: Podem ser considerados dados pessoais: nome e sobrenome; endereço residencial; endereço de e-mail (se contiver elementos que ajudem a identificar o dono, como nome e sobrenome); gênero; data de nascimento; número de documentos cadastrais, como RG, CPF e carteira de trabalho; dados de geolocalização de um telefone celular; número de telefone pessoal. Disponível em: <https://portal.fiocruz.br/noticia/entenda-melhor-lei-geral-de-protacao-de-dados-pessoais>. Acesso em 21 mar 2023.

Dado pessoal sensível: “dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural”. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em 30 jan 2023.

⁴ Exemplo retirado de: Gestión de Datos de Investigación - Parte I. Disponível em: <https://www.youtube.com/watch?v=BM-lZ2XCCN0>. Acesso em 26/03/2025.

2.1 Anonimização de transcrições

Para anonimização de transcrições, não é suficiente retirar o nome da pessoa entrevistada. Analise cuidadosamente as respostas das entrevistas e retire informações que permitam identificações diretas ou indiretas, como:

Nome de locais citados que possam identificar onde a pessoa nasceu, mora ou trabalha;

- Número de telefone, número de documento, data de nascimento ou endereço;
- Cargo que ocupa ou trabalho específico que realiza que poderia identificar a pessoa ou afiliação institucional;
- Citações a nomes de outras pessoas (ex.: nome de professores ou colegas de trabalho).

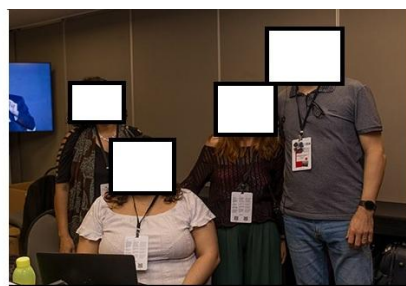
Evite apagar ou substituir itens sem indicar que houve edição realizada. Use pseudônimos ou substituições entre colchetes (ex.: “Ana mora em Brasília” -> “[E1] mora em [cidade do Centro-Oeste brasileiro]”) para indicar o local que houve alteração.

Ana mora em Brasília → [E1] mora em [cidade do Centro-Oeste brasileiro]

2.2 Anonimização de imagens

Imagens e vídeos também devem ser anonimizados. Desfoque ou cubra de forma a impedir a identificação de pessoas ou demais informações que não podem ser compartilhadas em acesso aberto (nomes de ruas, placas de carros, números de documentos, etc):

Exemplo de imagens anonimizadas:



Fonte: <https://25.scielo.org/fotos/>. Foto/Photo: Carla Formanek

O mesmo tratamento deve ser realizado com arquivos de áudio, adicionando atenção à possibilidade de identificação por meio da voz, sotaque e vícios de linguagem.

Em casos em que a prática de anonimização é impossível, busque usar pseudônimos ou, se ainda assim os participantes forem identificáveis, não deposite o arquivo em questão, apenas os demais arquivos que formarão o conjunto de dados (como as perguntas realizadas e análises das respostas), assim como um arquivo Readme ou outro arquivo de documentação do

conjunto de dados (veja seção 4), explicitando que nem todos os arquivos estão disponíveis e o motivo do impedimento do compartilhamento em acesso aberto.

Algumas ferramentas que podem auxiliar na anonimização são: [Amnesia do Openaire](#), [anonymoUUs da Utrecht University](#) e [Text Anonymisation helper tool do UK Data Service](#).

3. Formato dos arquivos

O SciELO Data aceita qualquer tipo de arquivo, porém é recomendado o depósito de formatos:

- Não proprietários;
- Abertos, com padrões internacionais documentados;
- Com codificação de caracteres padrão, de preferência Unicode (por exemplo, UTF-8).

Arquivos em formatos não proprietários não dependem de um software específico para serem abertos. Por isso, dê preferência aos formatos recomendados abaixo, a fim de facilitar o acesso aos arquivos por outras pessoas, assim como a preservação dos dados e o acesso a longo prazo.

Tabela 1. Formatos recomendados⁵

Tipo de documento	Formatos recomendados	Formatos não recomendados (mas aceitos)
Análises estatísticas	R (.r, .rdata) SPSS (.dat/.sps) STATA (.dat/.do)	SPSS Portable (.por) SPSS (.sav)
Arquivos compactados	.zip*	.rar
Dados tabulares	Comma Separated Values (.csv) Arquivo de texto (.txt) Excel (.xlsx)	Excel (.xls)
Dados textuais	Microsoft Word (docx) Arquivo de texto (.txt) HTML, XHTML ou XML, contendo referência para o DTD ou esquema de validação de formato (.html, .xhtml, .xml) OpenOffice (.odt, .ods ou .odp) PDF (.pdf)	Microsoft Word (.doc) Postscript (.ps) PDF cifrado (.pdf) WordPerfect (.wpd) DVI (.dvi)

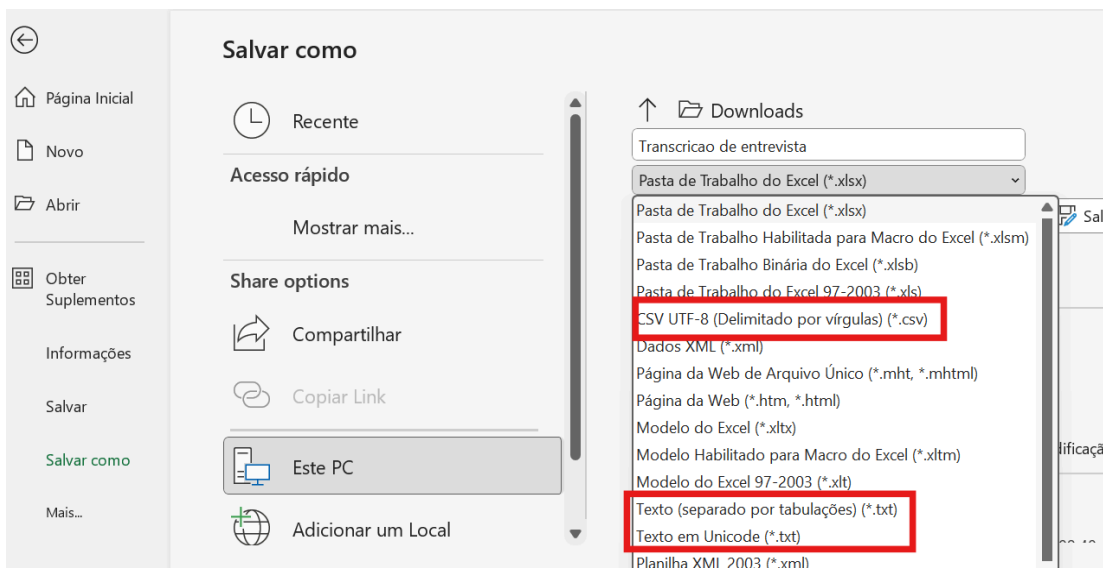
⁵ Fonte: Formatos adequados para preservação. [S.l.]: Arquivo.pt, [s.d.]. Disponível em: <https://sobre.arquivo.pt/pt/colabore/recomendacoes/formatos-adequados-para-preservacao/>. Acesso em: 30 abr. 2025.

Imagens	PNG (.png) JPEG2000 usando compressão sem perda de informação (.jp2) TIFF sem compressão SVG (.svg)	Macromedia Flash (*.swf) PhotoShop (.psd) JPEG 2000 Part 2 (.jpf, .jpx) MrSID (.sid) TIFF em formato Planar (.tiff) FlashPix (.fpx) RAW Encapsulated Postscript (.eps)
Áudio	AIFF com Pulse-code modulation (.aif, .aiff) WAV com Pulse-code modulation (.wav, bwf) Ogg Vorbis (.ogg, .oga)	RealNetworks 'Real Audio' (.ra, .rm, .ram) Windows Media Audio (.wma) WAV com compressão (.wav) AIFC com compressão (.aifc) NeXT SND(.snd)
Vídeo	QuickTime Movie sem compressão (.mov) AVI sem compressão (.avi) Motion JPEG 2000 (ISO/IEC 15444-4) (mj2) Motion JPEG (.avi, .mov)	Windows Media Video (.wmv) AVI com compressão (.avi) QuickTime Movie com compressão (.mov) RealNetworks 'Real Video' (.rv, .rm)

*Arquivos compactados com a extensão .zip serão descompactados após upload dos arquivos.

Uma vez depositados arquivos com formatos não recomendados, ou não listados como Alta ou Média adequação para preservação conforme o [Anexo 2](#), não há garantias que os arquivos serão devidamente preservados a longo prazo.

A conversão dos formatos dos arquivos na maioria das vezes pode ser realizada por meio do próprio programa utilizado para criar os documentos. Para arquivos criados com o Pacote Office, por exemplo, basta abrir o arquivo, clicar em “Salvar como” e selecionar algum formato aberto correspondente, como os exemplos destacados na imagem abaixo.



Se realizar a conversão dos arquivos de um formato para outro, informe no Readme ou documentação de apoio depositada com o conjunto de dados.

4. Nomeação dos arquivos

A adoção de boas práticas para nomeação de arquivos evita que estes sejam sobrescritos e facilita a localização e o reuso por outros pesquisadores.

- Use nomes auto explicativos/ descritivos, com no máximo 100 caracteres (ex.: “entrevista-crianca-1”);
- Não use acentuação;
- Use caracteres alfanuméricos, underline ou hífen (ex.: entrevista-crianca_1);
- Evite espaços (ex.: entrevista crianca 1), pontos (entrevista.crianca.1), iniciais maiúsculas (EntrevistaCrianca1) e caracteres especiais (tais como \ / ? : * " > < | : # % " { } | ^ ` ~ @ & ; ° æ Æ ø Ø å Å ä Ä ö Ö);
- Use o formato YYYY-MM-DD (entrevista_crianca_1_2021-01-07) ou YYYYMMDD (entrevista_crianca_1_20210107) para datas;
- Inclua o número da versão na nomeação quando apropriado (entrevista-crianca-1-001.csv, entrevista-crianca-1-002.csv, ..., entrevista-crianca-1-010.csv, ...);
- Use a mesma nomeação para arquivos com o mesmo conteúdo, mas em formatos diferentes (entrevista-crianca-1-dados.doc e entrevista-crianca-1.txt).

5. Descrição dos dados

Para que os dados depositados possam ser interpretados corretamente e reutilizados, tanto por você no futuro quanto por outros pesquisadores, é essencial que eles sejam descritos da forma mais detalhada e compreensível possível.

Essa descrição detalhada deve ser fornecida através do preenchimento dos campos durante o depósito e de um arquivo README, que funciona como um guia para o usuário e deve ser depositado junto aos arquivos de dados.

Alternativamente, outras documentações podem ser enviadas como cadernos de laboratório e dicionários de dados (que fornecem informações detalhadas sobre tabelas, colunas, tipos de dados, restrições e outras características dos dados), que possuam informações descritivas e contextuais sobre os dados depositados. O [Data Dictionary Creator](#) é um app que auxilia na criação de um dicionário de dados a partir dos arquivos da pesquisa.

Exemplo de dicionário de dados

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1	Tabela: Nome da tabela no banco de dados (ex: Tratamentos, Estudos).																					
2	Nome da Coluna: Nome da coluna que está na tabela (ex: ID_Paciente, ID_Tratamento).																					
3	Tipo de Dado: Tipo de dado que a coluna irá armazenar (ex: INT para números inteiros, VARCHAR(255) para textos com até 255 caracteres, DATE para datas).																					
4	Descrição: Uma breve explicação sobre o que a coluna representa (ex: Faixa etária do paciente, Duração do tratamento em dias).																					
5	Restrições: Regras que se aplicam à coluna. A restrição pode incluir "Não nulo", "Chave primária", "Chave estrangeira" e "Opcional".																					
6																						
7	Tabela, Nome da Coluna, Tipo de Dado, Descrição, Restrições																					
8	Pacientes, ID_Paciente, INT, Identificador único do paciente, Chave primária, Não nulo																					
9	Pacientes, Nome, VARCHAR(255), Nome completo do paciente, Não nulo																					
10	Pacientes, Idade, INT, Idade do paciente, Não nulo																					
11	Pacientes, Sexo, VARCHAR(10), Sexo do paciente (masculino ou feminino), Não nulo																					
12	Tratamentos, ID_Tratamento, INT, Identificador único do tratamento, Chave primária, Não nulo																					
13	Tratamentos, Nome_Tratamento, VARCHAR(255), Nome do tratamento administrado, Não nulo																					
14	Tratamentos, Tipo_Tratamento, VARCHAR(100), Tipo de tratamento (ex: farmacológico, fisioterápico), Não nulo																					
15	Tratamentos, Duração, INT, Duração do tratamento em dias, Não nulo																					
16	Estudos, ID_Estudo, INT, Identificador único do estudo, Chave primária, Não nulo																					

Exemplo de caderno de Laboratório

Exemplo de Caderno de Laboratório (Anônimo)

Página 1: Dados do Estudo

Data: 10/2023

Responsável: Dra. Maria Oliveira

Título do Estudo: Avaliação da eficácia de tratamento cognitivo-comportamental em pacientes com transtorno de ansiedade generalizada

Objetivo: Avaliar a redução dos sintomas de ansiedade após 5 semanas de tratamento cognitivo-comportamental em pacientes diagnosticados com transtorno de ansiedade generalizada.

Página 2: Participantes do Estudo

Número de Participantes: 5

Códigos dos Participantes:

- Paciente 01
- Paciente 02
- Paciente 03
- Paciente 04
- Paciente 05

Critérios de Inclusão:

- Diagnóstico de transtorno de ansiedade generalizada (conforme DSM-5).
- Idade entre 18 e 55 anos.
- Aceitação do termo de consentimento informado.

Critérios de Exclusão:

- História de transtornos psicóticos.
- Uso recente de medicamentos psicotrópicos (menos de 3 meses).

Página 3: Métodos

Tratamento Aplicado:

- **Tipo de Tratamento:** Terapia cognitivo-comportamental (TCC), 1 sessão semanal de 60 minutos durante 8 semanas.
- **Objetivos do Tratamento:** Redução dos sintomas de ansiedade, melhoria nas estratégias de enfrentamento e alteração de padrões cognitivos disfuncionais.

Instrumentos de Avaliação:

1. **Escala de Ansiedade:** Utilizada para medir a gravidade da ansiedade antes e após o tratamento.
2. **Questionário de Qualidade:** Avalia a qualidade de vida dos participantes.
3. **Entrevistas Semiestruturadas:** Conduzidas no início e no final do estudo para avaliar a experiência do paciente com o tratamento.

Página 4: Cronograma de Sessões

Semana 1: Introdução à TCC, identificação de pensamentos disfuncionais.

Semana 2: Estratégias de enfrentamento do estresse.

Semana 3: Modificação de pensamentos catastróficos.

Semana 4: Técnicas de relaxamento e controle da ansiedade.

Semana 5: Exposição gradual aos medos.

6. Arquivo README

O arquivo README, em específico, é um arquivo cuja **presença é obrigatória nos conjuntos de dados depositados no SciELO Data**. Ele deve ser escrito como um texto simples com codificação de caracteres Unicode UTF-8 (.txt) ou como PDF se for necessário ilustrar ou formatar a descrição dos dados.

O arquivo README deve conter pelo menos as seguintes informações:

- Título do conjunto de dados;
- Informações de contato (nome, instituição e e-mail) do pesquisador correspondente / principal ou da pessoa responsável pela coleta dos dados;
- Data de coleta dos dados (data única ou intervalo de tempo);
- Visão geral dos dados e arquivos (breve descrição dos dados que cada arquivo contém, data de criação de cada arquivo e como eles se relacionam entre si etc.);
- Descrição dos métodos de coleta ou geração dos dados;
- Descrição dos métodos usados para o processamento dos dados;
- Informações específicas dos dados (lista de variáveis, unidades de medida, definições de códigos ou símbolos, calibragem de equipamentos etc.).

O SciELO Data possui seu próprio template README, que abrange os tópicos acima e pode servir como base aos autores, e está [disponível para download](#).

Ao criar seu conjunto de dados, insira o máximo de informações e descrições dos dados possíveis, como por exemplo um Readme conforme o template SciELO Data, junto a um dicionário de dados especificando as colunas, linhas e variáveis de tabelas presentes.

Exemplos de outros templates para o arquivo README:

Geral	<ul style="list-style-type: none">• https://drive.google.com/drive/folders/1LNxx4YKxN1a2DF10bbVoRW-ioZwNz-9M• https://drive.google.com/file/d/167cJdaRy4sxQWA5qEEp2cgfWqKV-smZV/view• https://cornell.app.box.com/v/ReadmeTemplate
Ciências sociais	<ul style="list-style-type: none">• https://social-science-data-editors.github.io/template_README/template-README.html
Código de software	<ul style="list-style-type: none">• https://drive.google.com/file/d/1VIDF489DDr044Uta8z1G7EuLj-Wm_UtG/view

Para exemplos de outras informações que podem ser acrescentadas ao arquivo README consultar:

- [Guide to writing "readme" style metadata](#)

Para mais informações sobre a preparação de dados consulte também:

Códigos e dados de computadores	<ul style="list-style-type: none"> • Experiences on reproducibility of paper experiments • Research Code
Dados de Ciências Sociais	<ul style="list-style-type: none"> • Guide to Social Science Data Preparation and Archiving
Dados tabulares	<ul style="list-style-type: none"> • Preparing tabular data for description and archiving

7. Tamanho dos arquivos

O tamanho limite dos arquivos individuais é 2GB. Para adicionar arquivos acima deste limite, entre em contato com data@scielo.org.

8. Alterações no manuscrito

Ao depositar dados de pesquisa relacionados a um artigo, é necessário que o DOI do conjunto de dados seja incluído na seção Disponibilidade de Dados (ou equivalente), conforme o padrão do periódico. Caso esteja publicando um preprint, certifique-se de adicionar essa seção ao documento.

Como exemplo de texto, sugerimos:

“Todo o conjunto de dados anonimizados que dá suporte aos resultados deste estudo foi disponibilizado no SciELO Data e pode ser acessado em [**URL ou DOI do conjunto de dados**].


É também possível fazer a citação de seu próprio conjunto de dados no corpo do texto (exemplos de padrão de citação de conjuntos de dados estão disponíveis no [Guia de citação de dados de pesquisa](#)), o que incide na inserção do conjunto de dados na lista de referências do artigo seguindo o padrão normativo de referências determinado por seu periódico.

Para ter acesso à citação completa e DOI do conjunto de dados, após realizar o depósito no SciELO Data, vá à página inicial do conjunto de dados. As informações estarão no quadro azul abaixo do título do conjunto. O DOI só ficará ativo após a curadoria realizada pela equipe editorial e publicação do conjunto de dados.

[SciELO Data](#) > [SciELO Preprints](#) >

Dados de replicação para: Consumo alimentar de usuários da rede de Atenção Primária à Saúde, portadores de doenças crônicas não transmissíveis em Ponta Grossa, Paraná: estudo transversal, 2023

Versão 1.1



Batistel szczerepa, Sunali; Pollyanna Kássia de Oliveira Borges; Erido Vicente Muller; Denilson de Castro Teixeira; Clisia Mara Carreira; Eduardo Bauml Campagnoli; Sofia Lira Chiodi, 2024, "Dados de replicação para: Consumo alimentar de usuários da rede de Atenção Primária à Saúde, portadores de doenças crônicas não transmissíveis em Ponta Grossa, Paraná: estudo transversal, 2023", <https://doi.org/10.48331/scielodata.XEPS3V>, SciELO Data, V1

[Acessar conjunto de dados](#)

Contato com o proprietário [Compartilhar](#)

[Citar o cj. de dados](#)

[Aprenda sobre Padrões de citações de dados](#)

Referências

Arquivo.PT. Formatos adequados para preservação. [S.l.]: Arquivo.pt, [s.d.]. Disponível em: <https://sobre.arquivo.pt/pt/colabore/recomendacoes/formatos-adequados-para-preservacao/>. Acesso em: 30 abr. 2025.

Cornell University. Guide to writing "readme" style metadata. *Cornell University* [online]. [viewed 12 February 2021]. Available from: <https://data.research.cornell.edu/content/readme>.

DataverseNO. Prepare your data. *DataverseNO* [online]. [viewed 12 February 2021]. Available from: <https://site.uit.no/dataverseno/deposit/prepare/>.

Nanyang Technological University. DR-NTU (Data) User Guides and Policies. *Nanyang Technological University* [online]. [viewed 12 February 2021]. Available from: <https://libguides.ntu.edu.sg/drntudataguidespolicies/depositor#s-lg-box-21651979>.

Syracuse University. De-identification. Qualitative Data Repository, 2023. Disponível em: <https://qdr.syr.edu/guidance/human-participants/deidentification>. Acesso em: 30 abr. 2025.

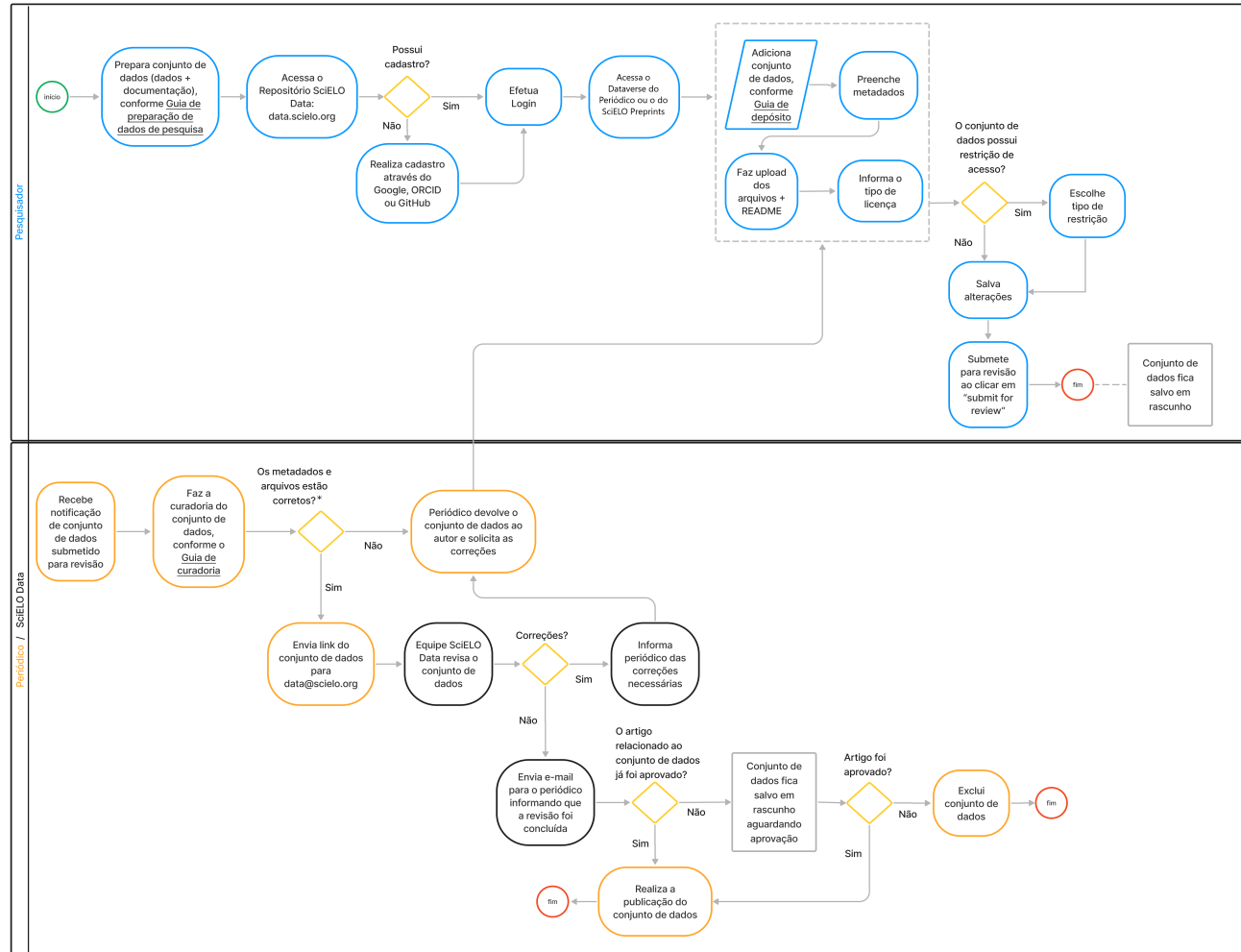
UC Santa Barbara Library. The Dos and Don'ts of file naming. *UC Santa Barbara* [online]. [viewed 30 March 2021]. Available from: <https://www.library.ucsb.edu/sites/default/files/dls-n01-2021-filenaming.pdf>.

University of Illinois at Urbana-Champaign. Dataset Documentation. *Illinois Data Bank* [online]. [viewed 12 February 2021]. Available from: https://databank.illinois.edu/help#dataset_documentation.

Como citar este documento

SciELO. *Guia de preparação de dados de pesquisa* [online]. SciELO, 2025 [cited DD Month YYYY]. Available from: _____.

Anexo 1. Fluxograma SciELO Data



***Apenas para os conjuntos de dados depositados no Dataverse do SciELO Preprints:** a curadoria é feita pela Equipe SciELO Data. Se houver necessidade de edição ou correções, a equipe SciELO entrará em contato com o autor depositante por e-mail para solicitar as correções. O conjunto de dados só será publicado após a aprovação do preprint e a realização das correções.

Anexo 2. Formatos adequados para preservação

Alguns formatos de arquivo são considerados mais adequados para preservação a longo prazo, pois não dependem de softwares específicos para serem abertos e acessados. Para saber se os formatos dos seus arquivos são classificados como de **Alta**, **Média** ou **Baixa** adequação para preservação, consulte a lista abaixo. Sempre que possível, dê preferência para salvar seus arquivos em formatos de Alta adequação para preservação.

Texto

Alta adequação para preservação

- HTML, XHTML ou XML, contendo referência para o DTD ou esquema de validação de formato (.html, .xhtml, .xml)
- Textos planos usando codificação de caracteres UTF-8, USASCII ou UTF-16 com Byte Order Mark (.txt)
- PDF/A-1 segundo a norma ISO 19005-1 (.pdf)
- Open Document Text (.odt)

Média adequação para preservação

- HTML, XHTML ou XML, sem referência para o DTD ou esquema de validação de formato (.html, .xhtml, .xml)
- Cascading Style Sheets (.css)
- Texto plano com codificação de caracteres em ISO-8859-1 (.txt)
- PDF com fontes embebidas (.pdf)
- Rich Text Format 1.x (.rtf)
- HTML 4.x incluindo declaração de DOCTYPE (.html)
- Open Office Text Document (.sxw)
- Office Open XML (.docx)
- DTD (.dtd)
- SGML (.sgml)

Baixa adequação para preservação

- Microsoft Word (.doc)
- Postscript (.ps)
- PDF cifrado (.pdf)
- WordPerfect (.wpd)
- DVI (.dvi)

Imagem

Alta adequação para preservação

- PNG (.png)
- JPEG2000 usando compressão sem perda de informação (.jp2)
- TIFF sem compressão
- SVG (.svg)

Média adequação para preservação

- JPEG2000 usando compressão com perdas de informação (.jp2)
- GIF (.gif)
- JPEG/JFIF (.jpg)
- TIFF com compressão
- BMP (.bmp)
- Digital Negative (.dng)
- Computer Graphic Metafile e WebCGM (.cgm)

Baixa adequação para preservação

- Macromedia Flash (*.swf)
- PhotoShop (.psd)
- JPEG 2000 Part 2 (.jpf, .jpx)
- MrSID (.sid)
- TIFF em formato Planar (.tiff)
- FlashPix (.fpx)
- RAW
- Encapsulated Postscript (.eps)

Áudio

Alta adequação para preservação

- AIFF com Pulse-code modulation (.aif, .aiff)
- WAV com Pulse-code modulation (.wav, bwf)
- Ogg Vorbis (.ogg, .oga)

Média adequação para preservação

- MP3 (MPEG-1/2, Layer 3) (.mp3)
- Free Lossless Audio Codec (.flac)

- SUN Audio sem compressão (.au)
- Standard MIDI (.mid, midi)
- Advance Audio Coding (.mp4, .m4a, .aac)

Baixa adequação para preservação

- RealNetworks 'Real Audio' (.ra, .rm, .ram)
- Windows Media Audio (.wma)
- WAV com compressão (.wav)
- AIFC com compressão (.aifc)
- NeXT SND(.snd)

Vídeo

Alta adequação para preservação

- QuickTime Movie sem compressão (.mov)
- AVI sem compressão (.avi)
- Motion JPEG 2000 (ISO/IEC 15444-4) (mj2)
- Motion JPEG (.avi, .mov)

Média adequação para preservação

- MPEG-1, MPEG-2 (.mpg, .mpeg)
- MPEG-4 (.mp4)
- Ogg Theora (.ogg, .ogm, .ogv)

Baixa adequação para preservação

- Windows Media Video (.wmv)
- AVI com compressão (.avi)
- QuickTime Movie com compressão (.mov)
- RealNetworks 'Real Video' (.rv, .rm)

Outros formatos

Alta adequação para preservação

- Comma Separated Values (.csv)
- SQL DDL

Média adequação para preservação

- OpenOffice (.sxc/.ods, .sxi/.odp))
- OOXML segundo a norma ISO/IEC DIS 29500 (.xlsx, .pptx)

Baixa adequação para preservação

- Microsoft Excel (.xls)
- Microsoft PowerPoint (.ppt)
- Microsoft Access (.mdb)
- Microsoft Visio(.vsd)

Fonte: Formatos adequados para preservação. [S.l.]: Arquivo.pt, [s.d.]. Disponível em: <https://sobre.arquivo.pt/pt/colabore/recomendacoes/formatos-adequados-para-preservacao/>. Acesso em: 30 abr. 2025.