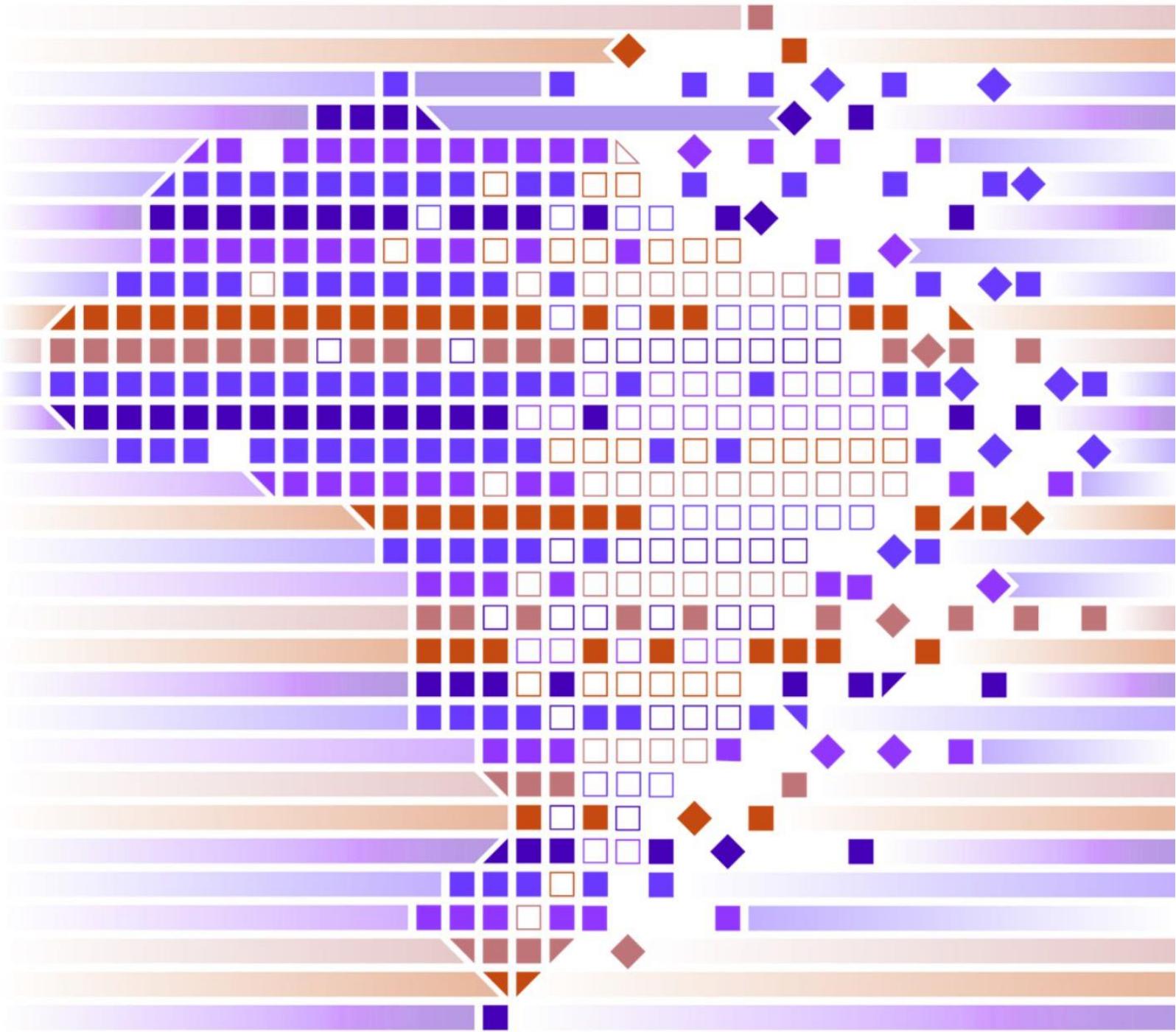


Plano de gestão de dados

Curso 4



A trilha Ciência Aberta abrange apostilas criadas pelo grupo RDPBrasil para treinamentos promovidos pela Rede Nacional de Pesquisa (RNP). O curso 4 intitula-se Planos de gestão de dados.

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
Planos de gestão de dados / 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.
66 f.

Curso 4 da trilha Ciência Aberta abrange apostilas criadas pelo grupo RDPBrasil para treinamentos promovidos pela Rede Nacional de Pesquisa (RNP).

1. Planos de gestão de dados. 2. Dados de pesquisa. 3. Gestão de dados de pesquisa. 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. Planos de gestão de dados. Porto Alegre: RDPBrasil, 2023.



ibict

RNP
ORGANIZAÇÃO SOCIAL DO MCTI

Curso 4

Plano de Gestão de Dados

MÓDULO I

Objetivos

Apresentar os movimentos de Ciência e Acesso Abertos, que levaram à concepção dos Planos de gestão de dados (PGDs), e demonstrar algumas ferramentas e orientações para o gerenciamento de dados de pesquisa.

Conceitos

Plano de Gestão de Dados; Ciência Aberta; Acesso Aberto.

1.1 PLANOS DE GESTÃO DE DADOS

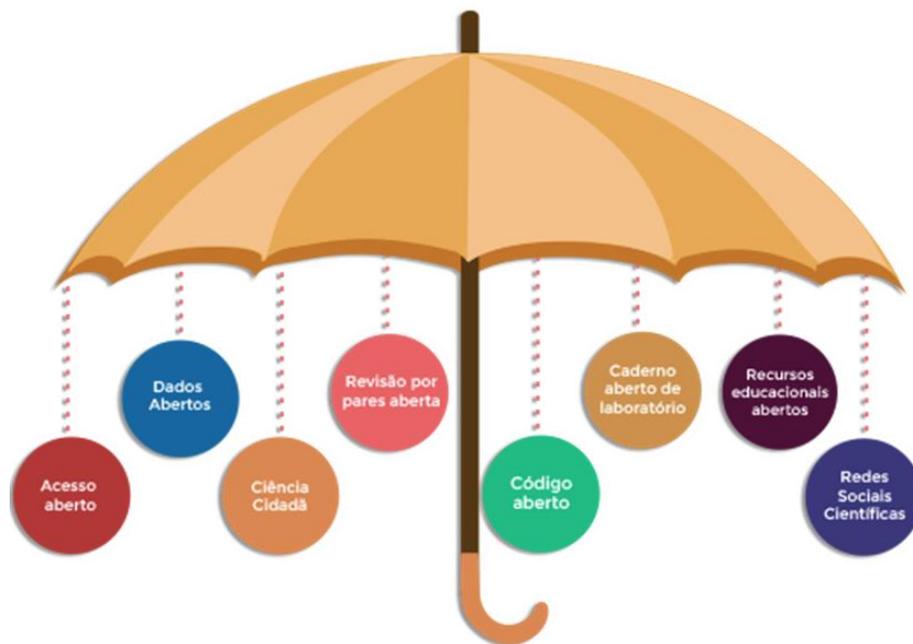
O movimento do acesso aberto foi constituído por uma série de iniciativas pioneiras que tiveram grande repercussão no Brasil, como o lançamento da SciELO em 1999 e a tradução do *Open Journal Systems* (OJS) pelo IBICT em 2003. Neste mesmo ano a *Declaração de Berlim sobre o Acesso Aberto ao Conhecimento em Ciências e Humanidades* ampliou o conceito de acesso aberto, dando espaço ao tratamento dos “[...] dados brutos e metadados, fontes originais, representações digitais de materiais pictóricos e gráficos, além de material multimídia acadêmico”, pois estes devem ser abertamente acessíveis e utilizáveis (PAMPEL; DALLMEIER-TIESSEN, 2014, p. 214).

Neste momento, um movimento mais amplo, denominado de “Ciência Aberta”, tem início. A Ciência Aberta busca tornar o conhecimento científico acessível a todos, promovendo a transparência e a colaboração em todas as etapas da pesquisa. Ela envolve a disponibilização pública de dados de pesquisa,

publicações científicas, métodos e resultados de pesquisa, bem como a promoção da colaboração entre pesquisadores, instituições e o público em geral. Ao promover a transparência e o acesso amplo, a Ciência Aberta busca acelerar o avanço do conhecimento, melhorar a qualidade da pesquisa e envolver a sociedade de maneira mais ampla na produção e no uso do conhecimento científico.

Para a Fiocruz, Ciência Aberta é um movimento que busca refletir sobre atividades, processos, métodos, avaliações e produções científicas, visando sua transparência, colaboração e abertura. Envolve diversos movimentos como o 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).

Figura 1 – “Guarda-chuva” da Ciência Aberta



Fonte: Fiocruz (20[--]).



SAIBA MAIS

[Compartilhamento de dados de pesquisa em repositórios \(RDP Brasil\)](#)

Dentro do ecossistema da Ciência Aberta, o movimento pela abertura dos dados de pesquisa vem ganhando força entre pesquisadores, agências de fomento, instituições de pesquisa e editores. 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.

São inúmeras iniciativas de abertura dos dados de pesquisa. No entanto, a comunidade reconhece vivenciar uma crise de reprodutibilidade: pesquisa realizada pela *Nature* em 2016 (BAKER, 2016) com 1.576 pesquisadores concluiu que mais de 70% falharam em reproduzir experimentos de outros cientistas e mais da metade falhou em reproduzir seus próprios experimentos. Nesse sentido, é fundamental que os pesquisadores se sensibilizem com a importância de disponibilizar informações sobre os dados, além de compartilhar o dado, de modo a possibilitar o reuso.

Os Planos de Gestão de Dados são usualmente elaborados a partir de ferramentas *online* que ajudam o pesquisador a descrever quais dados serão coletados, preservados e asseguram a qualidade, além de definir onde e por quanto tempo serão armazenados os dados de pesquisa:

Os Planos de Gerenciamento de Dados são documentos de texto em formato livre que descrevem os dados que são usados e produzidos durante as atividades de pesquisa. Eles especificam onde os dados serão arquivados, quais licenças e restrições se aplicam e a quem o crédito deve ser dado, etc. (MIKSA; WALK; NEIH, 2020, p. 1).

A UNIFESP (UNIVERSIDADE..., 2020, não paginado) descreve o *Data Management Plan* (DMP) ou Plano de Gestão de Dados (PGD), como “[...] um documento formal que descreve os dados produzidos durante um projeto de pesquisa e descrevem estratégias de gerenciamento de dados que serão implementadas durante e após a fase ativa do projeto de pesquisa.”. Mais especificamente, *Stanford University* (2015), *University College London* (2015) e *European Commission* (2022) definem Plano de Gestão de Dados como um documento que detalha os dados a serem adquiridos ou produzidos ao longo de um projeto de pesquisa, incluindo estratégias para gerenciamento, descrição, análise e armazenamento desses dados, bem como os métodos para compartilhamento e preservação dos mesmos após a conclusão do projeto.

Para promover a observância das boas práticas de pesquisa e gestão de dados, o *Digital Curation Centre* (2013) desenvolveu uma série de perguntas direcionadoras para a elaboração de um Plano de Gestão de Dados. Essas diretrizes abrangem aspectos como a seleção e método de coleta dos dados, documentação, metadados, considerações éticas e legais associadas aos dados, estratégias para gerenciamento e preservação a longo prazo, bem como questões relacionadas ao compartilhamento dos dados (DIGITAL..., 2013):



8 ORIENTAÇÕES

Digital Curation Centre

início orientações DCC



“**Dados administrativos**” descrevem os elementos essenciais que devem ser incluídos nessa seção, como a identificação dos dados, informações sobre o financiador da pesquisa, número de referência da bolsa (apenas para PGDs pós-concessão), nome do projeto, descrição do projeto, informações sobre o pesquisador principal, identificação do pesquisador (por exemplo, ORCID), informações de contato para dados do projeto, datas de criação e atualização do PGD. Além disso, o DCC destaca a importância de fornecer uma breve descrição do tipo de estudo realizado para que outros entendam os propósitos da coleta e criação de dados.



“**Coleção de dados**” envolve a definição do que será coletado ou criado, juntamente com os métodos a serem utilizados. Para isso, é importante considerar o tipo, formato e volume de dados a serem coletados, bem como a capacidade de compartilhamento e acesso a longo prazo desses dados. Além disso, é relevante avaliar a possibilidade de reutilização de dados existentes. É essencial fornecer uma breve descrição dos dados, incluindo qualquer fonte de terceiros ou dados já existentes, especificando seu conteúdo, tipo e alcance. Também é importante justificar a escolha de formatos, levando em consideração as implicações relacionadas ao armazenamento, *backup* e acesso aos dados. A seção também aborda a maneira como os dados serão coletados ou

criados, destacando a importância de definir padrões ou metodologias a serem utilizados. Isso inclui a estruturação e nomeação de pastas e arquivos, o gerenciamento de versionamento e a adoção de processos de garantia de qualidade. O documento deve descrever como a coleta de dados será realizada, quais padrões da comunidade serão seguidos (se aplicável) e como os dados serão organizados durante o projeto. Isso pode envolver convenções de nomenclatura, controle de versão e estrutura de pastas. Também é fundamental explicar como a consistência e qualidade da coleta de dados serão controladas e documentadas, o que pode incluir processos como calibração, coleta de amostras ou medições repetidas, captura ou registro padronizado de dados, validação de entrada de dados, revisão por pares dos dados ou uso de vocabulários controlados.



“Documentação e metadados” aborda a documentação e os metadados que acompanharão os dados do projeto. Para garantir que os dados possam ser lidos e interpretados no futuro, é essencial considerar que informações serão necessárias. Isso envolve pensar em como essa documentação e metadados serão capturados ou criados, bem como quais padrões de metadados serão utilizados e por quê. O PGD deve descrever os tipos de documentação que acompanharão os dados, fornecendo informações básicas que ajudarão os usuários secundários a localizar e compreender os dados. Isso inclui detalhes sobre quem criou ou contribuiu para os dados, o título dos dados, a data de criação e as condições de acesso. A documentação também pode incluir informações sobre a metodologia utilizada, procedimentos analíticos, definições de variáveis, vocabulários, unidades de medida, suposições feitas e o formato e tipo de arquivo dos dados. É importante considerar como essas informações serão capturadas e onde serão registradas. Sempre que possível, é recomendável identificar e utilizar padrões de metadados existentes na comunidade, a fim de garantir a interoperabilidade e a compreensão futura dos dados.



A seção **"Ética e conformidade legal"** aborda tanto as questões éticas quanto às questões legais relacionadas aos dados do projeto. Para gerenciar questões éticas, é necessário considerar se foram obtidos consentimentos para a preservação e compartilhamento de dados, bem como a identidade dos participantes será protegida, por exemplo, por meio da anonimização. O manuseio seguro de dados sensíveis, tanto em termos de armazenamento quanto de transferência, também é fundamental. As questões éticas têm impacto no armazenamento, acesso e retenção de dados e podem envolver a anonimização dos dados, consulta a comitês de ética institucionais ou departamentais e acordos formais de consentimento. Além disso, é importante abordar questões legais relacionadas a direitos autorais e propriedade intelectual. Isso inclui identificar o proprietário dos dados, definir as licenças para uso e reutilização dos dados e considerar quaisquer restrições no reuso de dados de terceiros. Também é necessário abordar se o compartilhamento de dados será adiado ou restrito para fins de publicação ou busca de patentes. Em projetos com múltiplos parceiros, a questão da propriedade intelectual pode ser discutida em um acordo de consórcio. É importante estar ciente das políticas pertinentes do financiador, instituição, departamento ou grupo em relação a direitos autorais e propriedade intelectual, bem como obter permissões para reutilizar dados de terceiros, se necessário.



"Armazenamento e backup" especifica a frequência dos *backups*, a quantidade de cópias a serem feitas e a localização de armazenamento. É recomendável evitar armazenar dados exclusivamente em *laptops*, discos rígidos de computadores ou dispositivos de armazenamento externos, visto que isso apresenta riscos significativos. A preferência é pelo uso de armazenamento gerenciado e robusto fornecido por equipes de TI da universidade. Da mesma forma, é aconselhável utilizar serviços de *backup* automático fornecidos por serviços de TI em vez de depender de processos

manuais. Quando se opta por usar um serviço de terceiros, é importante garantir que isso esteja em conformidade com as políticas do financiador, da instituição, do departamento ou do grupo, evitando possíveis conflitos, como questões jurídicas ou proteção de dados sensíveis. Além disso, é necessário abordar questões relacionadas à segurança e ao acesso aos dados. Isso envolve a avaliação dos riscos para a segurança dos dados e o planejamento de medidas para gerenciá-los. O controle de acesso e a segurança dos dados devem ser explicados, bem como a garantia de que colaboradores possam acessar os dados de forma segura. Se os dados forem coletados ou criados em campo, é fundamental planejar a transferência segura desses dados para os sistemas seguros principais. Se os dados forem confidenciais (por exemplo, dados pessoais não disponíveis publicamente, informações confidenciais ou segredos comerciais), medidas de segurança apropriadas devem ser delineadas, e é importante mencionar quais padrões serão seguidos para garantir a segurança dos dados.



Na seção **“Seleção e preservação”** é determinado quais dados serão mantidos e por quanto tempo. Isso pode depender de obrigações legais, do valor potencial de reutilização, da viabilidade econômica de retenção e dos esforços adicionais necessários para preparar os dados para compartilhamento e preservação, incluindo possíveis conversões de formato de arquivo. O plano de preservação a longo prazo dos dados também é abordado, incluindo questões sobre onde os dados serão armazenados, em qual repositório ou arquivo. É importante considerar quais custos podem ser associados a esse armazenamento e se foram incluídos no orçamento do projeto. Além disso, deve-se planejar o tempo e o esforço necessários para preparar os dados para compartilhamento e preservação. Se os dados têm um valor a longo prazo, o PGD deve descrever como eles serão preservados e mantidos além do período de financiamento do projeto. Caso não se utilize um repositório estabelecido, o plano de gerenciamento de dados deve demonstrar que recursos e sistemas estarão

disponíveis para garantir a curadoria eficaz dos dados além do período de financiamento do projeto.



“**Compartilhamento de dados**” detalha a forma como os dados serão compartilhados, incluindo considerações sobre como os potenciais usuários terão conhecimento sobre os dados, com quem eles serão compartilhados e sob quais condições. Além disso, é importante definir o mecanismo de compartilhamento, o momento em que os dados estarão disponíveis e a busca por um identificador persistente para os dados. No Plano de Gestão de Dados, é essencial considerar onde, como e para quem os dados com valor a longo prazo serão disponibilizados. A escolha dos métodos de compartilhamento dependerá de diversos fatores, como o tipo, tamanho, complexidade e sensibilidade dos dados. É aconselhável mencionar exemplos anteriores de compartilhamento de dados, caso existam, para demonstrar um histórico eficaz de compartilhamento. Também é importante considerar como as pessoas poderão reconhecer o reuso dos dados. Quando se trata de restrições ao compartilhamento de dados, o PGD deve abordar as medidas que serão tomadas para superar ou minimizar essas restrições, identificando as razões para a necessidade de uso exclusivo dos dados e se será necessário um acordo de compartilhamento de dados ou equivalente. Restrições podem ser devido à confidencialidade, falta de acordos de consentimento ou direitos autorais, por exemplo. Deve-se analisar se um acordo de sigilo oferecerá proteção suficiente para dados confidenciais.



A seção de “**Responsabilidades e recursos**” informa as responsabilidades das partes envolvidas na gestão de dados e os recursos necessários para executar o plano. Isso inclui definir quem será responsável pela implementação do PGD e revisão do mesmo, bem como a divisão de responsabilidades para cada atividade de gerenciamento de dados,

particularmente em projetos de pesquisa colaborativos em que várias instituições estão envolvidas. Caso seja aplicável, deve-se considerar se a propriedade dos dados e as responsabilidades relacionadas ao gerenciamento de dados farão parte de acordos de consórcio ou contratos acordados entre os parceiros. É importante descrever claramente as funções e responsabilidades para todas as atividades, como captura de dados, produção de metadados, garantia de qualidade dos dados, armazenamento e *backup*, arquivamento de dados e compartilhamento de dados, e nomear os indivíduos sempre que possível. No que diz respeito aos recursos necessários, é fundamental considerar se será necessária especialização adicional, treinamento para a equipe existente, *hardware* ou *software* que não faz parte do conjunto de *softwares* disponibilizado pela instituição e se haverá cobranças por parte de repositórios de dados. Deve-se detalhar e justificar cuidadosamente quaisquer recursos específicos necessários para a execução bem-sucedida do plano, como *software*, *hardware*, *expertise* técnica, etc. Caso recursos dedicados sejam necessários, estes devem ser identificados e explicados de forma apropriada.

fim orientações DCC

Os Planos de Gestão de Dados fornecem uma estrutura abrangente, delineando a trajetória dos dados desde a fase de planejamento até a sua preservação, ao descrever tanto os dados quanto às ferramentas utilizadas nas pesquisas científicas. Conforme apontado por Aventurier e Alencar (2016), esses planos compreendem a descrição pormenorizada do projeto de pesquisa, informações sobre os autores e suas filiações, requisitos das agências de fomento, natureza dos dados gerados, diretrizes para distribuição e reutilização, bem como estratégias de conservação e arquivamento.

Documentar o ciclo de vida dos dados em um PGD é uma prática que resulta em uma gestão mais eficiente, colaborativa e confiável dos dados de pesquisa.

Muitos esforços estão voltados para dinamizar os Planos de Gestão de Dados. Conforme a pesquisa avança, é necessário adaptar o Plano de Gestão de Dados para abranger todas as fases do ciclo de vida dos dados.

1.2 FERRAMENTAS DE ELABORAÇÃO DE PGDS

Existem diversas ferramentas que facilitam a tarefa de elaborar um Plano de Gestão de Dados, orientando o pesquisador por meio de uma lista de perguntas pré-elaboradas ou de texto livre, normalmente criado no início da pesquisa.



O *DMPTool* é uma ferramenta de elaboração de Planos de Gestão de Dados criada em 2011 como uma colaboração entre a Universidade da Califórnia, a Universidade de Illinois e a Universidade de Drexel. Ela foi projetada para auxiliar pesquisadores, instituições e financiadores a desenvolver planos eficazes de gerenciamento de dados em projetos de pesquisa. O *DMPTool* é uma ferramenta gratuita para os usuários e oferece modelos e orientações para a criação de PGDs personalizados, ajudando os pesquisadores a cumprir os requisitos específicos de financiadores e instituições. A ferramenta permite a geração de documentos PGD prontos para apresentação, facilitando o compartilhamento e a conformidade com políticas de gerenciamento de dados.



O *DMPonline* é uma ferramenta de elaboração de Planos de Gestão de Dados (PGD) que foi desenvolvida em 2011 pelo *Digital Curation Centre* (DCC) no Reino Unido. O *DMPonline* é uma ferramenta gratuita para os usuários e é amplamente



utilizada por pesquisadores e instituições para criar planos de gerenciamento de dados eficazes. Ele oferece modelos e diretrizes que ajudam os pesquisadores a elaborar PGDs personalizados, atendendo aos requisitos específicos de financiadores e instituições. Além disso, o *DMPonline* permite que os usuários gerem documentos contendo PGDs que podem ser compartilhados e usados para cumprir políticas de gerenciamento de dados. A ferramenta desempenha um papel crucial na promoção de boas práticas de gerenciamento de dados e na conformidade com as diretrizes de financiamento de pesquisa, e continua a ser uma opção valiosa para a comunidade de pesquisa no Reino Unido e além.



O *Data Steward Wizard* é uma ferramenta de elaboração de Planos de Gestão de Dados (PGD) criada pela Universidade de Viena, na Áustria. O *Data Steward Wizard* é uma ferramenta gratuita para os usuários e foi desenvolvida para auxiliar pesquisadores a criar planos de gerenciamento de dados eficazes. Ele oferece modelos e orientações que ajudam os pesquisadores a desenvolver PGDs personalizados de acordo com os requisitos de suas instituições e financiadores. A ferramenta permite que os usuários gerem documentos PGD prontos para apresentação, facilitando o compartilhamento e a conformidade com políticas de gerenciamento de dados. O *Data Steward Wizard* é uma contribuição valiosa para a comunidade de pesquisa, fornecendo uma solução gratuita para a elaboração e implementação de PGDs eficazes.

 [Data Steward Wizard](#)



O **EasyDMP** é uma ferramenta de elaboração de Planos de Gestão de Dados (PGD) desenvolvida pela *Digital Science*, uma empresa de tecnologia que se concentra em soluções para a pesquisa. Ela foi criada para auxiliar pesquisadores, instituições e financiadores a criar PGDs eficazes. O *EasyDMP* oferece uma versão gratuita e uma versão paga, com recursos adicionais na versão paga. O *EasyDMP* permite que os usuários gerem documentos contendo seus PGDs, que podem ser usados para atender aos requisitos específicos de financiadores e instituições. A ferramenta oferece modelos e orientações para facilitar a elaboração de PGDs personalizados. Os PGDs criados com o *EasyDMP* são úteis para a conformidade com políticas de gerenciamento de dados e para comunicar eficazmente o plano de gerenciamento de dados da pesquisa.

 [EasyDMP](#)



O **RDMO** (*Research Data Management Organiser*) é uma ferramenta de elaboração de Planos de Gestão de Dados que foi desenvolvida para auxiliar pesquisadores e instituições a criar e gerenciar planos de dados eficazes. O RDMO é uma ferramenta de código aberto e, portanto, é gratuita para uso. Foi criada em 2014 como um projeto de código aberto e é mantida por uma comunidade de desenvolvedores e pesquisadores. O RDMO permite que os usuários elaborem PGDs personalizados de acordo com os requisitos específicos de suas instituições e financiadores. A

ferramenta oferece modelos e orientações para facilitar a elaboração de PGDs abrangentes e atende a uma variedade de padrões e diretrizes de gerenciamento de dados. Além disso, o RDMO ajuda os usuários a gerar documentos PGD prontos para apresentação, que podem ser usados para atender às políticas de gerenciamento de dados e compartilhados com partes interessadas relevantes. O RDMO é uma opção valiosa para aqueles que buscam uma ferramenta de PGD gratuita e de código aberto, oferecendo flexibilidade e recursos para apoiar o gerenciamento de dados em projetos de pesquisa.



O *FioDMP* é um *template* de Plano de Gestão de Dados desenvolvido pela Fundação Oswaldo Cruz em 2020 e gratuito. O PGD é um documento formal essencial para a gestão de dados em pesquisas. O *template* do PGD da Fiocruz contém uma série de perguntas e orientações destinadas a auxiliar os pesquisadores no processo de gerenciamento de seus dados ao longo de todo o ciclo de vida do projeto de pesquisa. Essas perguntas ajudam os pesquisadores a planejar, coletar, armazenar, compartilhar e preservar dados de maneira eficaz, garantindo a integridade e o cumprimento de políticas relevantes. O *FioDMP* é uma ferramenta valiosa que contribui para a promoção de boas práticas de gestão de dados na comunidade de pesquisa da Fiocruz.



SAIBA MAIS

[Planos de Gestão de Dados \(RDP Brasil\)](#)

MÓDULO II

Objetivos

Apresentar as sugestões do PGD-BR (IBICT) e da ferramenta DMPTool para a criação de Planos de gestão de dados (PGDs).

Conceitos

Plano de gestão de dados; PGD-BR; DMPTool.

2.1 PLANO DE GESTÃO DE DADOS CONFORME SUGESTÃO DO IBICT

O Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), é o órgão criado em 1954 e vinculado ao Ministério da Ciência, Tecnologia e Inovação (MCTI), que realiza estudos no campo da ciência da informação e temas relacionados. Devido ao seu âmbito de atuação, o IBICT lançou recentemente uma ferramenta para elaboração do Plano de Gestão de Dados, como veremos a seguir.



PGD-BR
IBICT

O PGD-BR¹ é uma ferramenta *online* de código aberto que oferece suporte aos pesquisadores brasileiros na elaboração de seus Planos de Gestão de Dados. O PGD-BR proporciona um processo orientado para a criação dos PGDs, alinhando-

¹ <https://pgd.ibict.br/>

se geralmente com os critérios estabelecidos pelas agências financiadoras. Além disso, o PGD-BR disponibiliza *links* diretos para os endereços eletrônicos das agências de fomento, recursos de ajuda para a resolução de dúvidas e informações sobre as melhores práticas na gestão de dados científicos.

O PGD-BR é uma ferramenta desenvolvida e oferecida pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Foi lançado em 2023 e é uma iniciativa voltada para a promoção de boas práticas no gerenciamento de dados de pesquisa no Brasil.

O interesse do IBICT em oferecer o PGD-BR está relacionado à importância crescente da gestão de dados de pesquisa. À medida que a pesquisa científica e tecnológica se torna mais orientada por dados, é fundamental garantir que os dados sejam coletados, armazenados, compartilhados e preservados de forma adequada. O PGD-BR visa auxiliar pesquisadores e instituições a elaborar Planos de Gestão de Dados eficazes, garantindo que os dados de pesquisa sejam tratados de acordo com as melhores práticas.

O PGD-BR permite que o usuário elabore seus PGDs e gere um documento PDF ou DOCX contendo o plano. Os PGDs criados com a ferramenta podem ser exportados e compartilhados conforme necessário. Além disso, oferece a opção de armazenar o plano em seu próprio banco de dados, o que pode ser útil para fins de referência e acompanhamento, embora a prática de gerenciamento dos dados em si dependa do pesquisador ou instituição que utiliza a ferramenta.

Modelo de plano de gestão de dados elaborado a partir do PGD-BR do IBICT

Após logar-se/inscrever-se na página da ferramenta *online* PGD-BR deve-se clicar em “Criar PGD” e preencher os campos de acordo com o projeto:

DETALHES DO PROJETO	
CAMPO	COMO PREENCHER
Qual o título do projeto de pesquisa que receberá o PGD atual?	Preencher com o título do projeto de pesquisa para o qual os dados serão criados
Resumo do projeto	Descrever resumidamente o projeto
Área de conhecimento de pesquisa científica	Escolher a área de conhecimento a partir da lista disponibilizada
Data de Início do Projeto de Pesquisa	Selecionar a data de início da conclusão do projeto
Data de Conclusão do Projeto de Pesquisa	Selecionar a data prevista de conclusão do projeto
Agência de Fomento	Se o projeto possui algum auxílio financeiro por parte de agência de financiamento, selecionar a agência neste campo

<p>Na aba Colaboradores do projeto, adicione o(s) colaborador(es) do projeto</p>	<p>Neste campo é possível convidar colaboradores para a elaboração do PGD</p>
---	---

<p>ESCREVER PGD</p>	
<p>1. INICIAR O PLANO</p>	
<p>CAMPO</p>	<p>COMO PREENCHER</p>
<p>Qual modelo vai utilizar?</p>	<p>Escolher entre Plano de Gestão de Dados FAIR e Plano de Gestão de Dados simplificado</p>

Princípios FAIR

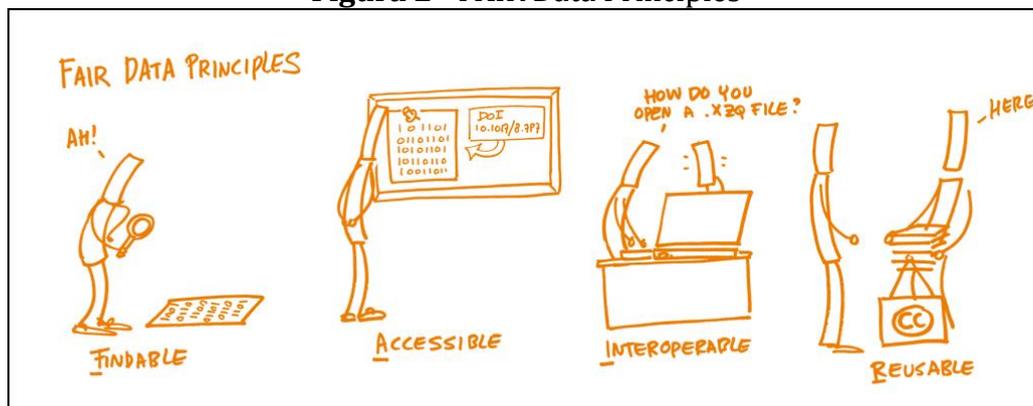
Os Princípios FAIR foram propostos para gestão de dados de pesquisa com o propósito de garantir a utilidade dos conjuntos de dados. FAIR é o acrônimo de *Findable, Accessible, Interoperable e Reusable* (WILKINSON *et al.*, 2016):

- **ENCONTRÁVEIS (FINDABLE):** os dados devem ser fáceis de encontrar para aqueles que precisam deles. Isso envolve a atribuição de identificadores exclusivos aos dados, a criação de metadados descritivos e a indexação adequada para que os dados possam ser localizados por meio de mecanismos de pesquisa;
- **ACESSÍVEIS (ACCESSIBLE):** os dados devem ser acessíveis a pessoas e sistemas autorizados. Isso implica em fornecer acesso aberto ou, quando necessário, garantir que os dados sejam acessíveis mediante solicitação e que os requisitos de acesso sejam claramente definidos;
- **INTEROPERÁVEIS (INTEROPERABLE):** os dados devem ser compatíveis com outros dados e sistemas, permitindo a integração e

a combinação com facilidade. Isso envolve a adoção de padrões de dados e metadados comuns, bem como a documentação clara dos formatos e estruturas de dados;

- **REUTILIZÁVEIS (REUSABLE):** os dados devem ser projetados e documentados de maneira a permitir sua reutilização eficaz. Isso inclui a criação de metadados detalhados, a documentação de metodologias de coleta e processamento de dados e a inclusão de informações sobre licenças e restrições de uso.

Figura 2 - FAIR Data Principles



Fonte: FOSTER (20[--]).

Portanto, os PGDs FAIR são planos de gerenciamento de dados que incorporam estratégias para garantir que os dados de pesquisa sejam tratados de maneira a atender aos princípios FAIR. Isso não apenas beneficia os pesquisadores na gestão de seus dados, mas também promove a transparência, a reprodutibilidade e o compartilhamento de dados, impulsionando a qualidade e o avanço da pesquisa científica e tecnológica.



SAIBA MAIS

[Seja FAIR com seus dados! \(RDP Brasil\)](#)

2. COLETA DE DADOS

CAMPO	COMO PREENCHER
Os dados serão:	Escolher entre as opções: coletar ou produzir, reutilizar de outras fontes ou outra opção
Caso os dados sejam reutilizados, indique o endereço eletrônico (URL, URI, etc.) ou identificador persistente (DOI, <i>handle</i>, etc.) que aponte a fonte dos dados ou metadados.	Indicar aqui o endereço eletrônico ou identificador persistente do conjunto de dados reutilizados, se for o caso
Do ponto de vista de sua natureza, quais tipos de dados serão coletados, produzidos ou reutilizados? Assinale quantas opções forem necessárias.	Informar se os dados serão imagem, desenho, caderno de laboratório, tabelas, fórmulas, maquete etc
Quais os formatos dos dados que serão coletados, produzidos ou reutilizados?	Informar se os arquivos dos dados serão csv, doc, mp4, xls, etc.
Qual o volume aproximado dos dados coletados, produzidos ou reutilizados?	Informar quantos bytes, Gigas, Megas serão utilizados para armazenar os dados
Descreva como os dados serão coletados ou criados?	Descrever aqui os procedimentos de coleta dos dados, por ex.: Os dados serão coletados a partir da pesquisa em bases bibliográficas nacionais...

Seu conjunto de dados contém dados pessoais ou sensíveis?

Selecionar sim ou não

A Lei Geral de Proteção de Dados (LGPD) entrou em vigor em setembro de 2020, com o objetivo de regulamentar o tratamento de dados pessoais no Brasil. Inspirada em regulamentações semelhantes, como o Regulamento Geral de Proteção de Dados (GDPR) da União Europeia, a LGPD estabelece diretrizes para a coleta, armazenamento, processamento e compartilhamento de informações pessoais, buscando proteger a privacidade e os direitos dos indivíduos. A lei impõe obrigações às organizações que lidam com dados pessoais e estabelece direitos para os titulares desses dados, como o direito de acesso, retificação e exclusão de informações pessoais. A LGPD tem impacto significativo em várias áreas, incluindo o setor empresarial, governamental e acadêmico, e promove uma cultura de segurança e privacidade de dados no Brasil.



SAIBA MAIS

[Boas práticas para dados abertos \(RDP Brasil\)](#)

[Atenção ao escolher os formatos dos arquivos! \(RDP Brasil\)](#)

3. DOCUMENTAÇÃO E METADADOS

CAMPO	COMO PREENCHER
Qual documentação irá acompanhar os dados?	Informar se o conjunto de dados será acompanhado por um caderno de laboratório, registro de fluxo de trabalho, anotações etc.

Qual padrão de **metadados** será utilizado?

Escrever se os metadados serão padronizados, por ex.: *Dublin Core*

O que são metadados?

Metadados são informações que fornecem contexto e descrição sobre outros dados. São dados sobre dados, e desempenham um papel fundamental na organização, identificação, recuperação e compreensão de informações. Os metadados podem incluir detalhes como título, autor, data de criação, formato de arquivo, tamanho, localização, e outras informações relevantes que ajudam a categorizar e facilitar a pesquisa e gerenciamento de recursos digitais.

Os metadados desempenham um papel fundamental na organização, gerenciamento, pesquisa e compreensão de informações. Os metadados ajudam a tornar os dados mais compreensíveis, acessíveis e úteis. Aqui estão alguns exemplos de metadados:

- **METADADOS DE DESCRIÇÃO:** esses metadados descrevem o conteúdo de um recurso. Por exemplo, o título, autor, resumo, palavras-chave, data de criação, tamanho do arquivo e informações similares;
- **METADADOS DE ESTRUTURA:** eles indicam como os diferentes componentes ou partes de um recurso se relacionam entre si. Por exemplo, em um livro, os metadados de estrutura podem incluir informações sobre capítulos, seções e páginas;
- **METADADOS ADMINISTRATIVOS:** são usados para gerenciar e administrar recursos. Isso pode incluir informações sobre a propriedade do recurso, datas de criação e modificação, histórico de versões, etc.

Os metadados facilitam a busca e a recuperação de informações, ajudam na preservação de recursos, garantem a conformidade com direitos autorais e políticas de uso, e contribuem para a organização eficaz de dados e conteúdo.

Formatos padronizados de metadados

Os formatos padronizados de metadados são estruturas ou esquemas de organização de dados que seguem regras e convenções predefinidas, a fim de facilitar a interoperabilidade, a troca e o gerenciamento eficaz de informações em diferentes sistemas e contextos. Alguns dos formatos de metadados padronizados mais comuns incluem:

- **DUBLIN CORE:** um padrão amplamente aceito que fornece um conjunto básico de elementos de metadados para descrever recursos digitais de forma simples e consistente;
- **MARC (MACHINE-READABLE CATALOGING):** usado principalmente em bibliotecas, o formato MARC é uma estrutura de metadados rica e altamente estruturada para a catalogação de materiais de biblioteca, incluindo livros e periódicos;
- **METS (METADATA ENCODING AND TRANSMISSION STANDARD):** o METS é um formato de metadados usado para descrever a estrutura e os relacionamentos entre componentes de objetos digitais complexos, como coleções digitais, arquivos de texto, imagens e multimídia;
- **MODS (METADATA OBJECT DESCRIPTION SCHEMA):** desenvolvido pela Biblioteca do Congresso dos Estados Unidos, o MODS é um formato de metadados que combina a simplicidade do *Dublin Core* com a riqueza do formato MARC, permitindo uma descrição detalhada de recursos digitais;

- **EAD (ENCODED ARCHIVAL DESCRIPTION):** usado em arquivologia, o EAD é um formato de metadados para a descrição de coleções de documentos arquivísticos e manuscritos, fornecendo informações sobre a estrutura e o conteúdo dessas coleções.

Esses formatos padronizados desempenham um papel essencial na facilitação da descrição, organização, pesquisa e intercâmbio de informações em uma variedade de contextos, incluindo bibliotecas, arquivos, museus, repositórios digitais, mecanismos de busca na web e muito mais. Eles ajudam a garantir a consistência e a qualidade dos metadados, tornando a recuperação e a gestão de recursos digitais mais eficazes e acessíveis.

Padrão de metadados *Dublin Core*

O *Dublin Core* é um conjunto de elementos de dados que foi desenvolvido para facilitar a descrição de recursos digitais, como documentos, imagens, vídeos e páginas da *Web*. Esses elementos de metadados são projetados para ajudar na catalogação, organização e recuperação de informações em ambientes digitais. O *Dublin Core* fornece um conjunto básico de termos que podem ser usados para descrever recursos de forma simples e consistente.

Os elementos de metadados do *Dublin Core* incluem informações como título, autor, data de criação, tipo de recurso, descrição, formato, idioma, cobertura geográfica e outros detalhes relevantes. Eles são frequentemente usados em bibliotecas, arquivos, museus, repositórios digitais e outros contextos nos quais a descrição precisa e a indexação de recursos digitais são necessárias.

O *Dublin Core* é um padrão aberto e amplamente aceito que facilita a interoperabilidade e a troca de metadados entre diferentes sistemas e instituições. Ele é utilizado para melhorar a pesquisa, a recuperação e a

organização de informações em ambientes digitais, tornando mais fácil localizar e acessar recursos digitais de interesse.



SAIBA MAIS

[Metadados em Dataverse \(RDP Brasil/IBICT\)](#)

4. ARMAZENAMENTO, BACKUP E PRESERVAÇÃO

CAMPO	COMO PREENCHER
Durante a pesquisa, onde o(s) conjunto(s) de dados será(ão) armazenado(s)?	Selecionar entre as opções se o armazenamento será na Nuvem da instituição, Nuvem privada, dispositivos externos de armazenamento etc.
Ao finalizar a pesquisa, onde o(s) conjunto(s) de dados serão armazenados (armazenamento definitivo)?	Selecionar se os dados serão armazenados em repositório institucional, temático, etc
Como será feito o <i>backup</i> dos dados?	Selecionar ou informar como será feito o <i>backup</i> dos dados
Como será gerenciado o acesso aos dados durante a pesquisa?	Selecionar ou escrever como os colaboradores da pesquisa terão acesso aos dados
Como será gerenciado o acesso aos dados ao final da pesquisa?	Selecionar ou descrever se o acesso aos dados de pesquisa se dará por

	meio de repositório de dados, se terá acesso ao público
Durante a pesquisa, se estiver criando ou coletando dados em campo, como você garantirá a transferência segura para seus principais sistemas?	Selecionar ou descrever se os colaboradores da pesquisa utilizarão dispositivos para coletar os dados, se os dados serão reunidos ao final da coleta etc

5. REQUISITOS ÉTICOS E LEGAIS

CAMPO	COMO PREENCHER
Quem possui os direitos autorais dos dados?	Selecionar ou descrever quem possui direito autoral sobre os dados
Informe a licença que será aplicada aos dados.	Informar a licença sob a qual os dados será depositados, a ferramenta oferece uma lista de opções além de campo de livre preenchimento
Como serão gerenciadas as questões éticas e legais?	Caso os dados sejam sensíveis ou possuam qualquer restrição, temporária ou não, descrever como os dados serão tratados
Como você protegerá a identidade dos participantes da pesquisa?	Selecionar se os dados serão anonimizados, criptografados etc. É possível ainda preencher o campo livre
Você obteve consentimento para o compartilhamento, a preservação e	Selecionar sim ou não, caso não se aplique, não responder

a reutilização dos dados por terceiros?



SAIBA MAIS

[Conheça seus direitos e use os dados corretamente! \(RDP Brasil\)](#)

6. COMPARTILHAMENTO DE DADOS	
CAMPO	COMO PREENCHER
Você planeja compartilhar os dados gerados/coletados na pesquisa?	Selecionar sim ou não
Quando você irá compartilhar os dados?	Informar se durante a pesquisa, após a conclusão da pesquisa
Onde você vai compartilhar os dados?	Selecionar ou descrever se os dados serão disponibilizados em repositórios de dados, institucionais etc.
Os dados ficarão embargados?	Selecionar sim ou não

7. RESPONSABILIDADE E RECURSOS

CAMPO	COMO PREENCHER
<p>Descreva as funções e responsabilidades de todas as atividades, por exemplo, captura de dados, produção de metadados, documentação, armazenamento e backup, arquivamento e compartilhamento de dados. Os responsáveis devem ser nomeados sempre que possível.</p>	<p>Descrever em campo de livre preenchimento a função de cada colaborador na coleta dos dados. Por ex.: A pesquisadora [Nome da pesquisadora] será responsável pela coleta, documentação, indicação dos metadados e envio para o repositório a ser definido.</p>
<p>Observação:</p> <p>Como o PGD deve, idealmente, ser elaborado na fase inicial, ainda em projeto de pesquisa, a aba Resultados da pesquisa deve ser preenchida ao final da referida pesquisa.</p> <p>Na aba Finalizar o plano o pesquisador pode definir a visibilidade do plano na página da ferramenta.</p> <p>Por fim, pode-se fazer o <i>download</i> do plano preenchido, optando pelos formatos pdf, csv, doc, html.</p>	

fim do modelo do PGD-BR/IBICT

O PGD-BR do IBICT foi construído e customizado com base na ferramenta *DMPTool*, apresentada a seguir.

2.2 FERRAMENTA DE ELABORAÇÃO DE PLANO DE GESTÃO DE DADOS DMPTOOL

O *DMPTool*² é uma ferramenta *online* projetada para auxiliar pesquisadores na elaboração de Planos de Gestão de Dados. A ferramenta oferece modelos e orientações para a criação de PGDs personalizados que atendam aos requisitos de instituições e agências de financiamento. O PGD é fundamental para garantir a conformidade com exigências de financiamento, promover boas práticas de gerenciamento de dados, facilitar a colaboração entre pesquisadores e permitir a revisão e atualização contínua dos planos à medida que o projeto de pesquisa avança.



DMPTool
*Universidades da Califórnia, de Illinois e de
Drexel*

Após logar-se na plataforma, o pesquisador deve responder às seguintes perguntas:

² <https://dmptool.org/plans>

Modelo de plano de gestão de dados elaborado a partir do DMPTool

1. DETALHES DO PROJETO

CAMPO	COMO PREENCHER
Título do projeto:	inserir o título da pesquisa que gerará os dados de pesquisa
Resumo do projeto:	inserir um breve resumo da pesquisa
Domínio de pesquisa:	escolher a partir da lista disponibilizada pela ferramenta
Início do projeto:	inserir a data de início do projeto
Fim do projeto:	inserir a data provável de fim do projeto
Órgão financiador:	ao começar a digitar o nome do órgão financiador a ferramenta irá sugerir opções

2. COLABORADORES

CAMPO	COMO PREENCHER
Adicionar colaborador:	inserir nome, ORCID, endereço de e-mail, afiliação e papel do pesquisador na coleta dos dados

3. *ESCREVER PLANO*

CAMPO	COMO PREENCHER
Quais dados você coletará ou criará?	escrever quais tipos de dados serão coletados
Como os dados serão coletados ou criados?	descrever como os dados serão coletados. Por exemplo: O procedimento de coleta de dados dar-se-á com a pesquisa em bases de dados bibliográficas...

4. *DOCUMENTAÇÃO E METADADOS*

CAMPO	COMO PREENCHER
Qual documentação e metadados acompanharão os dados?	informar se algum documento complementar irá acompanhar o conjunto de dados. Por exemplo: O conjunto de dados serão acompanhados por uma descrição da definição das variáveis escolhidas

5. *ÉTICA E CONFORMIDADE LEGAL*

CAMPO	COMO PREENCHER
Como você gerenciará quaisquer questões éticas?	informar como serão tratadas as questões éticas sobre os dados

Como você gerenciará questões de direitos autorais e direitos de propriedade intelectual?	informar de quem serão os direitos intelectuais e de propriedade dos dados gerados
--	--

6. ARMAZENAMENTO E BACKUP

CAMPO	COMO PREENCHER
Como os dados serão armazenados e copiados durante a pesquisa?	informar onde os dados serão armazenados durante a pesquisa
Como você gerenciará o acesso e a segurança?	informar como os colaboradores do projeto terão acesso aos dados. Por exemplo: Os colaboradores terão acesso via <i>drive</i>

7. SELEÇÃO E PRESERVAÇÃO

CAMPO	COMO PREENCHER
Quais dados têm valor de longo prazo e devem ser retidos, compartilhados e/ou preservados?	informar se os dados têm valor de longo prazo e devem ser retidos, compartilhados e/ou preservados
Qual é o plano de preservação de longo prazo para o conjunto de dados?	informar como os dados serão preservados a longo prazo. Por exemplo: Os dados serão depositados em formatos abertos

8. COMPARTILHAMENTO DE DADOS

CAMPO	COMO PREENCHER
Como você compartilhará os dados?	informar como os dados serão compartilhados. Por exemplo: Ao término os dados serão depositados em repositório de dados da área
São necessárias restrições ao compartilhamento de dados?	informar se há restrições para o compartilhamento

9. RESPONSABILIDADES E RECURSOS

CAMPO	COMO PREENCHER
Quem será responsável pelo gerenciamento de dados?	informar quem será responsável pela gestão dos dados. Por exemplo: A autora principal será a responsável pelo PGD, coleta e gestão
Quais recursos você precisará para entregar seu plano?	informar se algum outro recurso será requerido para acessar os dados, por exemplo, um <i>software</i>

10. RESULTADOS DA PESQUISA

Caso haja resultado no momento da elaboração do PGD, inserir neste momento.

11. FINALIZAR

Definir visibilidade do plano [escolher dentre as opções oferecidas]

Privado: visível para mim, para os colaboradores especificados e os administradores de minha organização

Organização: qualquer pessoa em minha organização pode ver

Público: qualquer pessoa pode ver

12. BAIXAR

É possível escolher entre os formatos pdf, csv, html, text, docx ou json.

É importante salvar as informações a cada conjunto de perguntas!

fim do modelo do DMPTool

MÓDULO III

Objetivos

Apresentar as sugestões para PGDs do CNPq (LattesData) e da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Conceitos

Plano de gestão de dados; LattesData; CNPq; FAPESP.

3.1 PLANO DE GESTÃO DE DADOS CONFORME SUGESTÃO DO CNPQ E FAPESP

O Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) é uma agência governamental brasileira que desempenha um papel fundamental na promoção e financiamento da pesquisa científica e tecnológica no Brasil. Em um esforço para fortalecer a governança e a transparência na pesquisa financiada por agências governamentais, o CNPq introduziu a exigência de Plano de Gestão de Dados como parte dos processos de concessão de recursos para pesquisas.

A exigência de PGDs pelo CNPq é uma resposta às crescentes demandas por transparência e responsabilidade no gerenciamento de dados de pesquisa. Os PGDs são documentos que descrevem como os dados gerados em projetos de pesquisa serão coletados, organizados, armazenados, compartilhados e preservados. Eles visam garantir a integridade, a reprodutibilidade e o acesso apropriado aos dados de pesquisa.

Essa exigência reflete a importância dada à gestão de dados de pesquisa de alta qualidade. Os PGDs não apenas ajudam os pesquisadores a planejar e organizar suas atividades de pesquisa, mas também incentivam o compartilhamento responsável de dados, promovendo a colaboração, a reutilização e a validação de resultados.

A exigência de PGDs pelo CNPq é uma iniciativa significativa que visa promover boas práticas de gerenciamento de dados, contribuindo para a qualidade e a transparência da pesquisa financiada no Brasil. Ela destaca a importância do gerenciamento adequado dos dados de pesquisa em um cenário de pesquisa cada vez mais globalizado e orientado por dados.

A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) é uma agência de fomento à pesquisa no Brasil e desempenha um papel fundamental no financiamento e apoio à pesquisa científica e tecnológica em São Paulo. A FAPESP tem adotado uma postura pró-ativa em relação à gestão de dados de pesquisa, reconhecendo a importância de dados de alta qualidade, acessíveis e reutilizáveis para o avanço da ciência.

A exigência de Planos de Gestão de Dados pela FAPESP é parte de sua política de apoio à pesquisa, e essa exigência é aplicada a projetos de pesquisa financiados pela fundação.

Essa exigência reflete o compromisso da FAPESP com a promoção de boas práticas de gerenciamento de dados de pesquisa. Os PGDs incentivam os pesquisadores a pensar estrategicamente sobre como seus dados serão tratados ao longo do ciclo de vida do projeto e a considerar as necessidades de compartilhamento e preservação a longo prazo. Além disso, a FAPESP tem promovido o desenvolvimento de infraestruturas e serviços de apoio ao gerenciamento de dados, ajudando a comunidade de pesquisa a cumprir essas exigências.

A exigência de PGDs pela FAPESP é uma iniciativa que promove a qualidade, a integridade e a reprodutibilidade da pesquisa financiada, contribuindo para a excelência da pesquisa científica e tecnológica em São Paulo e no Brasil como um todo. Ela ressalta a importância do gerenciamento adequado de dados de pesquisa em um ambiente cada vez mais orientado por dados e colaborativo.



LattesData
CNPq

Assim como a FAPESP, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) aderiu à solicitação de um PGD a todos os envolvidos na geração ou coleta de dados, do compromisso de depósito dos dados de pesquisa no repositório *LattesData*. O Plano inicia orientando o pesquisador a respeito do tipo de dados que será coletado ou criado. Veja:

Modelo de Plano de Gestão de Dados a partir do que o CNPq disponibiliza para os pesquisadores depositarem seus dados no LattesData

1. COLEÇÃO DE DADOS

1.1 Que dados serão coletados ou criados?

ORIENTAÇÃO

Descrição dos Dados

Descrever resumidamente os dados a serem coletados, produzidos ou reutilizados para a investigação da hipótese de pesquisa formulada. Justifique sua escolha de formato e considere as implicações do formato e

PERGUNTAS ORIENTADORAS

✓ Qual o tipo, o formato e o volume dos dados (por exemplo: dados tabulares em planilhas Excel com ocupação de espaço de aproximadamente 10 Mbytes)?

dos volumes de dados em termos de armazenamento, backup e acesso.

✓ *Os formatos e software escolhidos permitem o compartilhamento e o acesso de longo prazo aos dados?*

✓ *Há dados possíveis de serem reutilizados?*

[\[Preencher aqui\]](#)

1.2 Como os dados serão coletados ou criados?

ORIENTAÇÃO

Obtenção dos Dados

Relatar resumidamente como serão obtidos os dados coletados, produzidos ou reutilizados. Pode incluir processos como repetição de amostras ou medições, captura ou registro de dados padronizados, validação de entrada de dados, revisão por pares de dados ou representação com vocabulários controlados.

PERGUNTAS ORIENTADORAS

✓ *Quais padrões ou metodologias serão usados?*

✓ *Como serão estruturados e nomeados as pastas e arquivos?*

✓ *Como se lidará com o controle de versão?*

✓ *Que processos de garantia de qualidade serão adotados?*

[\[Preencher aqui\]](#)

Depois de preencher sobre os tipos dos dados, o pesquisador é instruído a escrever sobre alguma documentação que acompanhará os dados:

2. DOCUMENTAÇÃO E METADADOS

2.1 Que documentação e metadados irão acompanhar os dados?

ORIENTAÇÃO

Metadados

Indicar os metadados e os padrões adotados para ajudar os usuários secundários a entendê-los e reutilizá-los. Isso deve incluir pelo menos detalhes básicos que ajudarão as pessoas a encontrar os dados, incluindo encontrar quem os criou ou contribuiu com os dados, seu título, data de criação e em que condições eles podem ser acessados. A documentação também pode incluir detalhes sobre a metodologia usada, informações analíticas e procedimentais, definições de variáveis, vocabulários, unidades de medida, quaisquer suposições feitas e o formato e tipo de arquivo dos dados.

PERGUNTAS ORIENTADORAS

- ✓ *Que informações são necessárias para que os dados sejam lidos e interpretados no futuro?*
- ✓ *Como será capturada/criada a documentação e os metadados?*
- ✓ *Quais padrões de metadados você usará e por quê?*
- ✓ *Os dados de pesquisa se adequam ao princípio FAIR (Localizável, Acessível, Interoperável, Reutilizável)?*

[\[Preencher aqui\]](#)

O PGD do CNPq também destaca a importância dos dados estarem em conformidade legal, principalmente se utiliza-se algum dado pessoal que requer restrições de acesso, e a respeito dos direitos autorais. Confira:

3. ÉTICA E CONFORMIDADE LEGAL	
3.1 Como serão gerenciadas quaisquer questões éticas?	
<p>ORIENTAÇÃO</p> <p>Aspectos Éticos e Legais</p> <p>Relatar como serão observados os aspectos legais, éticos e de conduta, indicando como serão assegurados o cumprimento das leis de proteção de dados e de propriedade intelectual.</p>	<p>PERGUNTAS ORIENTADORAS</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> <i>Há consentimento para preservação e compartilhamento de dados?</i> <input checked="" type="checkbox"/> <i>Como será protegida a identidade dos participantes, se necessário? (exemplo: via anonimização)</i> <input checked="" type="checkbox"/> <i>Como os dados confidenciais serão tratados para garantir que sejam armazenados e transferidos com segurança?</i>
<p>[Preencher aqui]</p>	
3.2 Como serão gerenciadas as questões de direitos autorais e direitos de propriedade intelectual?	
<p>ORIENTAÇÃO</p> <p>Direitos autorais</p>	<p>PERGUNTAS ORIENTADORAS</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> <i>Quem é o dono dos dados?</i>

<p>Declarar quem será o proprietário dos direitos autorais de quaisquer dados que você coletar ou criar, junto com a(s) licença(s) para seu uso e reutilização. Considere também parceiros e as permissões para reutilizar dados de terceiros e quaisquer restrições necessárias ao compartilhamento de dados.</p>	<ul style="list-style-type: none"> ✓ <i>Como os dados serão licenciados para reutilização?</i> ✓ <i>Haverá período de embargo ou restrições de acesso aos dados e por quê? (exemplo: para publicação ou solicitação de patente)</i>
<p>[Preencher aqui]</p>	

Além disso, é de fundamental importância o pesquisador descrever no seu Plano de Gestão de Dados como eles serão armazenados e os procedimentos de segurança, incluindo backup, durante a pesquisa:

<h4>4. ARMAZENAMENTO E BACKUP</h4>	
<h5>4.1 Como os dados serão armazenados e preservados durante a pesquisa?</h5>	
<p>ORIENTAÇÃO</p> <p>Armazenamento</p> <p>Relatar brevemente como os dados serão armazenados e os procedimentos de segurança, incluindo <i>backup</i>, durante a pesquisa. Indique com que frequência os dados serão copiados e em quais locais.</p>	<p>PERGUNTAS ORIENTADORAS</p> <ul style="list-style-type: none"> ✓ <i>Você tem armazenamento suficiente ou precisará incluir encargos para serviços adicionais?</i> ✓ <i>Como será feito o backup dos dados?</i>

Armazenar dados em discos rígidos de computador ou dispositivos de armazenamento externo sozinho é muito arriscado. Normalmente é melhor usar serviços de backup automático fornecidos por serviços de TI. Se optar por usar um serviço de terceiros, deve-se garantir que não entre em conflito com qualquer financiador, políticas institucionais ou departamentais como em termos de jurisdição legal em que os dados serão mantidos ou a proteção de dados confidenciais.

- ✓ *Quem será responsável pelo backup e recuperação?*
- ✓ *Como os dados serão recuperados em caso de incidente?*

[\[Preencher aqui\]](#)

4.2 Como você gerenciará o acesso e a segurança?

ORIENTAÇÃO

Dados confidenciais

Delinear todas as medidas de segurança adequadas e observar quaisquer padrões formais que você cumprirá (exemplo: ISO 27001) em caso de dados confidenciais (exemplos: dados pessoais que ainda não sejam de domínio público, informações confidenciais ou segredos comerciais).

PERGUNTAS ORIENTADORAS

- ✓ *Quais são os riscos para a segurança dos dados e como eles serão gerenciados?*
- ✓ *Como será controlado o acesso para manter os dados seguros?*
- ✓ *Como você garantirá que os colaboradores possam acessar seus dados com segurança?*

	<p> <i>Se estiver criando ou coletando dados no campo, como você garantirá sua transferência segura para seus principais sistemas protegidos?</i></p>
<p>[Preencher aqui]</p>	

Planejar a respeito do armazenamento dos seus dados é pensar nas futuras pesquisas que se realizarão, mas simultâneo a isso, deve-se pensar na preservação a longo prazo desses dados e como serão selecionados para pesquisas futuras. O PGD do CNPq orienta os pesquisadores:

5. SELEÇÃO E PRESERVAÇÃO

5.1 Quais dados tem valor a longo prazo e devem ser retidos, compartilhados e/ou preservados?

ORIENTAÇÃO	PERGUNTAS ORIENTADORAS
<p>Preservação</p> <p>Como os dados para preservação serão selecionados e onde os dados serão preservados a longo prazo (exemplo: em um repositório de dados)? Lembre-se de considerar qualquer esforço adicional necessário para preparar os dados para compartilhamento e preservação, como alterar formatos de arquivo.</p>	<p> <i>Quais dados devem ser retidos para fins contratuais, legais ou regulamentares?</i></p> <p> <i>Como será decidido quais outros dados manter?</i></p> <p> <i>Quais são os usos de pesquisa previsíveis para os dados?</i></p>

Todos os dados coletados ou produzidos em projetos de pesquisa fomentados pelo CNPq serão depositados no repositório *LattesData*.

Por quanto tempo os dados serão retidos e preservados?

[\[Preencher aqui\]](#)

5.2 Qual é o plano de preservação de longo prazo para o conjunto de dados?

ORIENTAÇÃO

Plano de preservação

Considere como os conjuntos de dados com valor de longo prazo serão preservados. Caso o repositório utilizado para preservação não seja o *LattesData*, o PGD deve demonstrar que os recursos e sistemas estarão disponíveis para permitir que haja curadoria dos dados de forma eficaz além do prazo de validade da concessão. Além disso, o *dataset* deverá ter os seus metadados também depositados no *LattesData* referenciando o depósito no repositório escolhido.

PERGUNTAS ORIENTADORAS

Em qual repositório ou arquivo os dados, por exemplo, os dados serão armazenados?

Quais são os custos, se houver, com o repositório de dados ou arquivo selecionado?

Houve gasto de tempo e esforço para preparar os dados para compartilhamento/preservação?

[\[Preencher aqui\]](#)

6. COMPARTILHAMENTO DE DADOS

6.1 Como você compartilhará os dados?

ORIENTAÇÃO

Métodos para compartilhar dados

Como, quando e para quem os dados serão compartilhados? Se possível, mencione exemplos anteriores para mostrar um histórico de compartilhamento eficaz de dados. Os dados deverão ser compartilhados, no caso de beneficiários do CNPq em prazo não superior a 60 dias após o término do benefício.

PERGUNTAS ORIENTADORAS

- ✓ *Quando serão disponibilizados os dados?*
- ✓ *Se o registro de identificador DOI - Digital Object Identifier será feito para o conjunto de dados ou se haverá necessidade de atribuir individualmente identificadores arquivos, isso implica custos ao detentor do repositório LattesData.*
- ✓ *Caso o depósito não seja feito originalmente no LattesData, como será assegurado o registro de um identificador único e persistente (como um DOI - Digital Object Identifier) para cada conjunto de dados?*
- ✓ *Por quanto tempo os dados serão retidos e preservados?*

[Preencher aqui]

Direito de Acesso e Licença – Os dados coletados ou produzidos em projetos de pesquisa fomentados pelo CNPq e depositados no repositório LattesData serão

de Acesso Aberto e Licença CC/BY ou CC0 ou CC BY-SA ou CC BY- NC, com padrão de citação conforme estabelecido no LattesData.

6.2 É necessária alguma restrição ao compartilhamento de dados?

ORIENTAÇÃO

Restrições previstas

Descreva quaisquer dificuldades esperadas no compartilhamento de dados com valor reconhecido de longo prazo, juntamente com as causas e possíveis medidas para superá-las. Considere se um acordo de não divulgação forneceria proteção suficiente para dados confidenciais.

PERGUNTAS ORIENTADORAS

- ✓ *Que ação será tomada para superar ou minimizar as restrições?*
- ✓ *Por quanto tempo será preciso o uso exclusivo dos dados e por quê?*
- ✓ *Será necessário um acordo de compartilhamento de dados (ou equivalente)?*

[\[Preencher aqui\]](#)

7. RESPONSABILIDADES E RECURSOS

7.1 Quem será o responsável pela gestão dos dados?

ORIENTAÇÃO

Responsabilidade

Quem será o responsável pela gestão dos dados de pesquisa? Antes do depósito, no caso de beneficiários do CNPq, o responsável pela gestão dos dados será o

Coordenador do Projeto, e após o depósito no *LattesData*, caberá ao CNPq a gestão dos dados, conforme estabelecido no Termo de Depósito do *LattesData*.

[\[Preencher aqui\]](#)

7.2 Quais são os detalhes do projeto de autoria que precisam ser esclarecidos?

ORIENTAÇÃO

Projeto de pesquisa

Definir e descrever aspectos da pesquisa. Elementos principais como o título, resumo e palavras-chave precisam ser fornecidos.

PERGUNTAS ORIENTADORAS

- ✓ Qual é o título do projeto?
- ✓ Quais são as palavras-chave?
- ✓ Qual é a instituição executora?
- ✓ Qual é o início da vigência e o término da vigência?
- ✓ Quais foram os recursos financeiros e suas fontes (quantidade, alocação, distribuição e organização de recursos em geral)?
- ✓ Qual foi a chamada? Qual o número do processo?
- ✓ Qual a área do conhecimento (baseado nas áreas CNPq) que sua pesquisa se enquadra?
- ✓ Qual é a hipótese de pesquisa (objetivo dos dados)?

ORIENTAÇÃO	PERGUNTAS ORIENTADORAS
<p>Autoria</p> <p>Definir os dados pessoais essenciais sobre a autoria do conjunto de dados de pesquisa, tais como proponente, e-mail, identificador único de autor, instituição executora.</p>	<ul style="list-style-type: none"> <input checked="" type="checkbox"/> Qual o nome do responsável pelo conjunto de dados? <input checked="" type="checkbox"/> Qual o endereço de email para contato sobre o conjunto de dados? <input checked="" type="checkbox"/> Qual o identificador único do autor? (exemplo: ID Lattes, ORCID) <input checked="" type="checkbox"/> Qual o nome completo da instituição responsável pelos dados? <input checked="" type="checkbox"/> Qual o telefone para contato sobre o conjunto de dados?
<p>[Preencher aqui]</p>	

Por fim, há uma última observação no PGD do CNPq a respeito da “ciência” de todos os tópicos abordados:

Importante: O proponente declara que deu ciência, a todos os envolvidos na geração ou coleta de dados, do compromisso de depósito dos dados de pesquisa no repositório *LattesData* e que todos estão de acordo.

fim do modelo do LattesData





Modelo PGD

FAPESP

Nos últimos dois anos, 2022 e 2023, a FAPESP realizou chamadas de propostas onde o Plano de Gestão de Dados figurou como elemento mandatório, como a *Chamada FAPESP-NERC: Global Partnerships Seedcorn Fund – 2023*; *Auxílio à Organização Científica – 17ª Chamada para Escolas São Paulo de Ciência Avançada (ESPCA)*; *Chamada de Propostas BIOTA-FAPESP: Transformação*; *Chamada de Propostas Biodiversity on a Changing Planet 2023*; *Chamada FAPESP/CNPq Programa de Apoio à Fixação de Jovens Doutores no Brasil - 2022*; *Chamada de Propostas para Pesquisa Estratégica sobre a Internet (PIPE) - 2022*.

A instituição não tem um modelo próprio para tais planos, pois estes variam conforme o domínio do conhecimento, os tipos de dados considerados e como os responsáveis pelo projeto pretendem disponibilizá-los. Algumas chamadas FAPESP poderão especificar o formato desejado do Plano. Conforme informações do site da FAPESP em outubro de 2023, um PGD submetido como anexo de uma proposta à FAPESP deve seguir as seguintes diretrizes:

Modelo de Plano de Gestão de Dados da FAPESP

Texto de até **duas páginas**, contendo as seguintes informações:

- a) **Descrição dos dados e metadados produzidos pelo projeto:** por exemplo, amostras, registros de coleta, formulários, modelos, resultados experimentais, software, gráficos, mapas, vídeos, planilhas, gravações de áudio, bancos de dados, material didático e outros;
- b) **Quando aplicável, restrições legais ou éticas para compartilhamento de tais dados:** políticas para garantir a

privacidade, confidencialidade, segurança, propriedade intelectual e outros;

- c) Política de preservação e compartilhamento:** por exemplo, compartilhamento imediato ou apenas após a aceitação da publicação associada. Período de carência (antes do compartilhamento) e período durante o qual os dados serão preservados e disponibilizados;
- d) Descrição de mecanismos, formatos e padrões para armazenar tais itens de forma a torná-los acessíveis por terceiros:** esta descrição pode incluir o uso de repositórios e serviços de outras instituições.

A FAPESP também tem um modelo de PGD para os Centro de Pesquisa, Inovação e Difusão (CEPID), Centro de Pesquisa em Engenharia (CPE), e Centro de Ciência para o Desenvolvimento (CCD). A FAPESP considera necessário que os dados resultantes de projetos por ela financiados sejam gerenciados e compartilhados de forma a garantir o maior benefício possível para o avanço científico, tecnológico, socioeconômico e cultural.

Um Plano de Gestão de Dados para propostas desses Centros é um texto de até quatro páginas, que deve responder a quatro perguntas básicas:

- 1. Quais dados digitais serão gerados pelo Centro?**
- 2. Como serão gerenciados internamente (cobrindo as etapas de geração, limpeza e armazenamento seguro e confiável)?**
- 3. Como serão preservados e compartilhados (interna e externamente), considerando questões éticas, legais, privacidade e segurança?**
- 4. Caso o Centro decida criar o próprio repositório institucional de dados, como planeja fazê-lo e gerenciá-lo?**

Para efeito do PGD, os dados coletados poderão ser tanto físicos como digitais (como por exemplo amostras de tecidos e cultivares, dentre outros). Para dados

produzidos pelas pesquisas do Centro, o PGD será restrito a dados digitais, como por exemplo registros de coleta, formulários eletrônicos, modelos computacionais, resultados experimentais e de simulação, séries temporais, *software*, gráficos, mapas, vídeos, planilhas, gravações de áudio, bancos de dados, material didático e outros. Notar que *software* (código ou executável) está contido na descrição ampla de dados digitais, embora se entenda que os procedimentos e natureza sejam distintos.

Também para efeito do PGD, considera-se “compartilhamento interno” aquele restrito aos membros do Centro e “compartilhamento externo” se refere a pessoas ou instituições que não estão associadas ao Centro.

O texto deverá ser organizado da seguinte forma, considerando as questões indicadas:

a) Sumário descritivo dos dados digitais a serem produzidos e gerenciados no Centro

Que dados (digitais ou não) serão coletados? Quais tipos e formatos de dados o Centro irá gerar? Qual a estimativa do volume dos dados a serem gerados? Haverá armazenamento e compartilhamento de dados brutos?

Software e outros objetos digitais – além de dados propriamente ditos, que se pretende gerar no Centro – *software, workflows*, protocolos, modelos, etc?

b) Gerenciamento dos objetos digitais armazenados – formatos, metadados e interoperabilidade

Que padrões de metadados serão utilizados para documentar os dados e outros objetos digitais produzidos (genéricos e/ou específicos a domínios do conhecimento). Quais formatos de dados e metadados serão adotados para viabilizar o compartilhamento interno e externo? Que providências

serão tomadas para permitir o reuso – por exemplo, documentação do *software* usado para gerar os dados, uso de padrões adotados amplamente, dentre outros. Os arquivos serão identificados por DOI ou outro tipo de identificador?

c) Compartilhamento interno e externo

Os dados produzidos pelo Centro serão, em sua integralidade, tornados abertos e disponíveis livremente visando permitir o reuso mais amplo possível? Identificar os tipos de dados que não podem ser compartilhados livremente e apresentar justificativa para tal. Haverá previsão do uso de licenças de uso (por exemplo, CC-BY). Que mecanismos estão previstos para compartilhamento interno? E para compartilhamento externo?

Haverá políticas diferenciadas dentro do Centro contemplando dados que possam ser compartilhados livremente e outros que tenham restrições de disseminação? O Centro pretende estabelecer níveis diferentes de acesso para compartilhamento interno? E para compartilhamento externo?

d) Gerenciamento dos dados – responsabilidades, preservação e segurança

Quem será responsável pelo gerenciamento dos dados gerados no Centro? Haverá um comitê gestor de dados? Como se pretende decidir quais dados serão preservados, e por quanto tempo, e quais descartados? Quais os mecanismos que serão usados para preservação durante a existência do Centro? Qual o planejamento para preservação a longo prazo? Que cuidados se pretende tomar para impedir acesso indevido ou destruição dos dados? Como a obediência ao plano será monitorada e como se pretende atualizá-lo ao longo da vida do Centro?

e) Restrições éticas, de confidencialidade, legais e outras

Haverá algum tipo de restrição ética ou legal que possa impactar o compartilhamento amplo dos dados produzidos? Haverá necessidade de, por exemplo, termos de consentimento e autorização de comitês de ética? Quais as medidas que serão tomadas para conformidade com a LGPD? Havendo restrições de compartilhamento, quais os mecanismos previstos para permitir o eventual reuso (por exemplo, acordos de confidencialidade)?

Haverá embargo para a abertura dos dados produzidos? Se sim, justificar e indicar o período de embargo.

f) Repositórios

O Centro pretende criar seu próprio repositório de dados? Se sim, descrever brevemente como planeja criá-lo e gerenciá-lo, incluindo necessidade de equipamentos e pessoal. Adicionalmente, descrever as medidas para permitir que os dados nele depositados sejam compartilhados interna e externamente, conforme os planos de compartilhamento mencionados nos itens anteriores, garantidas as restrições éticas e legais.

Se não pretende ter seu próprio repositório, em quais repositórios institucionais confiáveis pretende armazenar os dados gerados?

fim do modelo da FAPESP

MÓDULO IV

Objetivos

Apresentar os aspectos relativos à documentação de dados e da pesquisa, bem como demonstrar a importância de formatos abertos e não proprietários para a reprodutibilidade e replicabilidade de pesquisas.

Conceitos

Documentação de dados e pesquisa; Formatos abertos e não proprietários; Reprodutibilidade e replicabilidade de pesquisas.

4.1 DOCUMENTAÇÃO DO DADO E DA PESQUISA

A documentação de dados de pesquisa é um componente essencial de qualquer projeto de pesquisa, especialmente em contextos científicos e acadêmicos. Ela consiste na criação de registros detalhados que descrevem e explicam todos os aspectos relacionados aos dados coletados ou gerados durante uma pesquisa. A documentação de dados de pesquisa desempenha um papel fundamental na garantia da qualidade, na reprodutibilidade e no compartilhamento de dados, permitindo que outros pesquisadores compreendam, utilizem e avaliem os dados de maneira eficaz.

A documentação de dados de pesquisa não apenas beneficia o próprio pesquisador, ajudando a manter a consistência e a integridade dos dados, mas também facilita o compartilhamento responsável de dados com a comunidade acadêmica. Além disso, a documentação adequada pode ser fundamental para a revisão por pares, a replicação de estudos e a validação dos resultados da

pesquisa. Portanto, é uma prática importante para garantir a integridade e a credibilidade da pesquisa científica.

Conforme Sayão e Sales (2015), a documentação exaustiva dos dados é fundamental para compreender o significado dos dados, em qualquer tempo. A ausência de uma descrição detalhada que abranja o contexto tecnológico dos arquivos de dados, o ambiente de criação ou coleta, as medidas, os detalhes espaciais e temporais, os instrumentos, parâmetros, unidades, qualidade e proveniência dos dados, torna improvável que esses dados possam ser identificados, interpretados, gerenciados e utilizados de maneira eficaz, incluindo sua reutilização. Os metadados desempenham um papel fundamental nesse aspecto, pois constituem a documentação dos dados, permitindo que os dados sejam autoexplicativos tanto no presente quanto no futuro.

Sayão e Sales (2015) explicam que um elemento de extrema relevância na gestão de dados de pesquisa reside na necessidade de assegurar a compreensibilidade e interpretabilidade dos dados, tanto no presente quanto no futuro, por parte de qualquer utilizador. Isso implica na elaboração de uma descrição precisa e minuciosa dos dados, juntamente com anotações suplementares e informações contextuais que permitam aos dados transmitir informações e conhecimentos através do tempo e do espaço. Esse requisito é concretizado através da documentação que deve acompanhar os dados, conhecida como “documentação dos dados”.

A documentação que acompanha os dados descreve o processo de aquisição ou geração desses recursos, esclarece o significado dos dados, descreve seu conteúdo e estrutura, e registra qualquer manipulação à qual tenham sido submetidos. A documentação dos dados é considerada uma das melhores práticas no que diz respeito à criação, organização e gestão de dados, além de constituir uma estratégia fundamental para a preservação digital dos mesmos.

Assim, para garantir que seus dados possam ser identificados, localizados, acessados, utilizados e reutilizados de maneira apropriada por pesquisadores ou outras partes interessadas, é imperativo que esses dados sejam acompanhados por uma documentação completa que abranja todos os seus aspectos. A presença

de uma documentação abrangente aumenta significativamente as chances de que os dados sejam descobertos na *internet*, citados por terceiros e que seu valor seja adequadamente atribuído aos autores.

Uma parcela crucial da documentação é composta por metadados. A inclusão de metadados simplifica a busca e o uso dos dados ao longo do tempo. Os metadados são um subconjunto estruturado e padronizado da documentação dos dados, contendo elementos de informação claramente definidos, como "título", "autor", "resumo" e "fonte", que contribuem para fornecer contexto e esclarecer a origem dos dados, ou seja, sua procedência e histórico, tanto para indivíduos como para sistemas. Os metadados apresentam informações sobre a origem, propósito, período de referência, localização geográfica, autor, condições de acesso e termos de uso das coleções de dados, entre outros detalhes.

Deste modo, a elaboração de uma documentação de qualidade primordialmente implica a atribuição de metadados adequados aos dados. É essencial identificar as exigências relativas ao nível de descrição e atribuição de metadados desde o início do projeto e revisá-las ao longo do ciclo de vida dos dados. Este é o cerne da gestão eficaz de dados. No entanto, é importante ter em mente que o esforço investido na documentação dos dados depende da durabilidade pretendida e do grau de compartilhamento almejado.

Uma estratégia útil é começar por descrever o projeto de pesquisa, o que auxiliará na contextualização da pesquisa e justificará a coleta de dados necessária. Em seguida, é necessário descrever os dados em si, abordando todas as suas características técnicas e estruturais.

Vale ressaltar que, quando um protocolo de pesquisa é utilizado, grande parte da documentação necessária já está disponível. Se forem utilizados instrumentos, devem ser registrados os processos de calibração de equipamentos e quaisquer ajustes necessários para a captura dos dados, bem como outras informações relevantes, como parâmetros, unidades e valores codificados, por exemplo.

O exemplo de registro no caderno de laboratório talvez represente a forma mais rigorosa de documentação da pesquisa, portanto, considera-se incluir uma versão digital deste caderno na documentação que acompanha os dados.

Elaborar a documentação dos dados a partir do pressuposto de que quem vai utilizar que não possui familiaridade com o projeto de pesquisa (muitas vezes nem o conhece), tampouco conhece o ambiente de pesquisa e a metodologia adotada. Entender que o contexto do projeto de pesquisa é desconhecido é fundamental para a vida dos dados de pesquisa. Ademais, uma vez que os dados serão arquivados para uso futuro, a documentação deve ser redigida de modo a instruir os utilizadores que estarão acessando os dados daqui muitos anos. O tema vem sendo muito discutido na atualidade. Veja a matéria publicada no *Blog da SciELO*, acerca da criação da editoria de replicabilidade na revista Dados:

Figura 3 – Artigo “Revista DADOS cria editoria especializada em replicabilidade”.



Fonte: SciELO (2023).

4.2 FORMATOS ABERTOS E NÃO PROPRIETÁRIOS

"Formatos abertos" e "não proprietários" referem-se a tipos de formatos de dados que são acessíveis, interoperáveis e não estão vinculados a uma empresa ou entidade específica, o que permite sua utilização, modificação e distribuição por qualquer pessoa sem restrições significativas.

FORMATOS ABERTOS: são aqueles cujas especificações estão disponíveis para o público em geral, sem restrições de acesso ou uso. Isso significa que qualquer pessoa pode entender a estrutura do formato, seu funcionamento e como os dados estão organizados dentro dele. Geralmente, as especificações dos formatos abertos são documentadas e publicadas, possibilitando a criação de *software* compatível e interoperável. Exemplos de formatos abertos incluem:



CSV

Comma-Separated Values

Um formato de arquivo simples que armazena dados em forma de tabela, onde os valores são separados por vírgulas.



JSON

JavaScript Object Notation

Um formato de troca de dados que é fácil de ler e escrever, utilizado comumente em APIs da Web.



XML

Extensible Markup Language

Um formato que define regras para a codificação de documentos de forma legível por humanos e por máquinas.

FORMATOS NÃO PROPRIETÁRIOS: são aqueles que não são controlados por uma entidade específica e não estão sujeitos a restrições de propriedade

intelectual ou licenciamento. Isso significa que o formato pode ser utilizado livremente por qualquer pessoa, sem a necessidade de pagar *royalties* ou obter permissões especiais. Exemplos de formatos não proprietários incluem:



PNG

Portable Network Graphics



OGG

Ogg Vorbis

Um formato de imagem amplamente utilizado para armazenar imagens de forma comprimida sem perda de qualidade, sem restrições de patentes. Um formato de compressão de áudio que é livre de patentes e pode ser usado para armazenar e reproduzir música e outros sons.

A utilização de formatos abertos e não proprietários é fundamental em contextos de pesquisa, pois promove a transparência, a colaboração e a preservação dos dados, permitindo que outros pesquisadores possam acessar, entender e usar os dados de forma eficiente e sem barreiras legais ou técnicas. Além disso, ajuda a evitar dependências excessivas de fornecedores específicos e garante a longevidade e acessibilidade dos dados ao longo do tempo.



SAIBA MAIS

[Por que os formatos abertos são importantes? \(RDP Brasil\)](#)

4.3 PLANO DE GESTÃO DE DADOS E DOCUMENTAÇÃO DA PESQUISA COMO GARANTIA PARA A REPRODUTIBILIDADE E REPLICABILIDADE DAS PESQUISAS

Spinak (2023) argumenta que o empreendimento científico depende da comunidade científica testar e validar continuamente as afirmações, destacando a necessidade crucial de reconhecer e comunicar incertezas nos resultados. A

reprodutibilidade e replicabilidade nos dados, especialmente em trabalhos científicos que utilizam intensivamente a computação, ganharam destaque com o "movimento de pesquisa reprodutível" iniciado por Jon Claerbout nos anos 1990, fundamentado na suposição de que a análise repetida dos mesmos dados com os mesmos métodos resultaria em conclusões idênticas. Ou seja, a reprodutibilidade é entendida, conforme Spinak (2023), como a capacidade de alcançar resultados idênticos em um estudo independente, seguindo procedimentos o mais similares possível ao experimento original.

Os autores Schaefer, Campos e Candido (2023) apontam que o movimento de Ciência Aberta, entre suas diversas facetas, diz respeito à divulgação e reprodução de dados, além de outros aspectos. Os autores destacam a necessidade de transparência nos processos de coleta, tratamento e análise de dados para permitir que terceiros reproduzam os resultados. Nesse sentido, a replicabilidade, que significa clareza e transparência nas pesquisas, é crucial para evitar erros, aprimorar a avaliação de pesquisas, fortalecer o desenvolvimento científico e aumentar o impacto do trabalho.

Diante do exposto, o plano de gestão de dados e a documentação minuciosa são elementos fundamentais para a garantia da reprodutibilidade e replicabilidade nas pesquisas científicas. A transparência assegurada por um PGD bem elaborado e uma documentação detalhada é fundamental para elucidar de forma clara e aberta os processos de coleta, tratamento e análise dos dados. Essa transparência não apenas fomenta a confiança entre os pares acadêmicos, mas também permite uma avaliação criteriosa dos métodos e resultados, promovendo a solidez e credibilidade da pesquisa.

Além disso, a replicabilidade e reutilização dos estudos são facilitadas pela documentação detalhada dos métodos utilizados, dos dados coletados e das análises conduzidas. Essa prática não somente valida os resultados iniciais, mas também possibilita que outros pesquisadores explorem, reutilizem e expandam as descobertas originais, impulsionando o avanço contínuo do conhecimento na área.

Outro ponto crucial é a detecção de erros nos processos de coleta e análise de dados, viabilizada por um PGD bem estruturado e por uma documentação minuciosa. Essa capacidade de identificar e corrigir equívocos contribui para a promoção de boas práticas científicas, prevenindo a disseminação de resultados falsos ou imprecisos.

Além disso, a consideração ética é fundamental, especialmente no que diz respeito à proteção de dados sensíveis. Um PGD bem delineado e uma documentação cuidadosa garantem a anonimização adequada dos dados, preservando a privacidade dos participantes da pesquisa e mantendo a integridade ética do estudo.

A disponibilização de dados e métodos também impulsiona a inovação e colaboração entre os pesquisadores, permitindo que outros explorem novas abordagens e contribuam com diferentes perspectivas. Isso estimula a criatividade e acelera o avanço do conhecimento, fortalecendo a credibilidade e o impacto das descobertas.

Portanto, a documentação da pesquisa e a criação de um plano de gestão de dados são fundamentais para a integridade e confiabilidade de qualquer estudo científico. Através de uma documentação clara e abrangente, os pesquisadores garantem que cada etapa do processo seja registrada, desde a coleta inicial até a análise final dos dados. Essa transparência não só facilita a compreensão do trabalho por parte de outros pesquisadores, mas também permite a replicação dos estudos, fortalecendo a validade e a confiabilidade dos resultados apresentados.

Além disso, a documentação detalhada contribui para a consistência e continuidade do projeto, assegurando que novos membros da equipe possam entender facilmente o trabalho realizado e dar continuidade ao projeto, caso necessário. Também desempenha um papel vital na preservação e proteção dos dados, delineando como eles são armazenados, organizados e compartilhados. Isso não apenas evita perdas acidentais de dados, mas também garante conformidade com regulamentos de privacidade e ética.

Outro aspecto importante é que a documentação eficaz facilita a análise e interpretação dos resultados, ajudando os pesquisadores a compreender corretamente as descobertas e a identificar eventuais erros ou vieses. Além disso, a documentação clara e um plano de gestão de dados bem elaborado promovem a colaboração entre equipes e instituições, facilitando a troca de conhecimentos e a comunicação entre os pesquisadores.

REFERÊNCIAS

AVENTURIER, P.; ALENCAR, M. C. F. Os desafios dos dados de pesquisa abertos. **Revista Eletrônica de Comunicação & Inovação em Saúde**, v. 10, n. 3, jul./set. 2016. p. 1-19.

DIGITAL CURATION CENTRE. **Checklist for a data management plan. v4.0**. Edinburgh: DCC, 2013. Disponível em: https://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Ch_ecklist_2013.pdf. Acesso em: 15 nov. 2022.

EUROPEAN COMMISSION. **Data management**. Disponível em: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm. Acesso em: 10 out. 2023.

FORMENTON, D. *et al.* Los estándares de metadatos como recursos tecnológicos para la garantía de la preservación digital. **Biblios**, 2017, n. 68, p. 82-95. DOI: <http://dx.doi.org/10.5195/biblios.2017.414>

FUNDAÇÃO OSWALDO CRUZ. **O que são formatos abertos**. Disponível em: <https://campusvirtual.fiocruz.br/portal/guiarea/OpenFormats.html#:~:text=Os%20formatos%20propriet%C3%A1rios%20representam%20a,podem%20ser%20conhecidas%20por%20todos>. Acesso em: 23 nov. 2023.

MIKSA, T.; WALK, P.; NEISH, P. **RDA DMP common standard for machine-actionable data management plans**. 2020. Disponível em: <https://doi.org/10.15497/rda00039>

OPEN KNOWLEDGE. **Formatos de arquivo**. Disponível em: https://opendatahandbook.org/guide/pt_BR/appendices/file-formats/. Acesso em: 23 nov. 2023.

SANCHEZ, F. A.; SILVA, N. B. P.; VECHIATO, F. L. Padrões de metadados para representação e organização da informação em repositórios de dados de pesquisa. **Informação & Tecnologia**, v. 5, n. 1, 2019. DOI: <https://doi.org/10.22478/ufpb.2358-3908.2018v5n1.38350>

SAYÃO, L.; SALES, L. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. 2015. Disponível em: <https://www.aben.com.br/Arquivos/420/420.pdf>. Acesso em: 03 nov. 2023.

SCHAEFER, B.; CAMPOS, L. A.; CANDIDO, M. R. Revista DADOS cria editoria especializada em replicabilidade [online]. **SciELO em Perspectiva**, 2023. Disponível em: <https://blog.scielo.org/blog/2023/10/20/revista-dados-cria-editoria-especializada-em-replicabilidade/>. Acesso em: 18 nov. 2023.

SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C.. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ciência da Informação**, v. 29, n. 1, p. 93–102, jan. 2000.

SPINAK, E. Reprodução e replicação na pesquisa científica – parte 1 [online]. **SciELO em Perspectiva**, 2023. Disponível em: <https://blog.scielo.org/blog/2023/05/19/reproducao-e-replicacao-na-pesquisa-cientifica-parte-1/>. Acesso em: 18 nov. 2023.

STANFORD UNIVERSITY. Libraries. **Data management plans**. Disponível em: <https://library.stanford.edu/research/data-management-services/data-management-plans>. Acesso em: 10 out. 2023.

UNIVERSIDADE FEDERAL DE SÃO PAULO. Escola de Filosofia, Letras e ciências Humanas. **Gestão de dados de pesquisa**. 11 nov. 2020. Disponível em: [https://www.unifesp.br/campus/gua/dados-de-pesquisa/plano-de-gestao-de-dados#:~:text=0%20Data%20Management%20Plan%20\(DMP,ativa%20do%20projeto%20de%20pesquisa](https://www.unifesp.br/campus/gua/dados-de-pesquisa/plano-de-gestao-de-dados#:~:text=0%20Data%20Management%20Plan%20(DMP,ativa%20do%20projeto%20de%20pesquisa). Acesso em: 10 out. 2023.

UNIVERSITY COLLEGE LONDON. **Writing a data management plan**. Disponível em: <https://www.ucl.ac.uk/library/open-science-research-support/research-data-management/policies/writing-data-management-plan>. Acesso em: 10 out. 2023.

WILKINSON, M. D. *et al.* The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, v. 3, p. 160018, mar. 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 03 nov. 2023.