

CARTILHA DE GESTÃO DE DADOS PARA PESQUISADORES

DADOS DE PESQUISA



Quem ama cuida

Luana Sales - IBICT
Luis Sayão - CNEN

Ilustrações: Joana Marques

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
COMISSÃO NACIONAL DE ENERGIA NUCLEAR

DADOS DE PESQUISA: QUEM AMA CUIDA

Luana Farias Sales

Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT

Luis Fernando Sayão

Comissão Nacional de Energia Nuclear – CNEN

Rio de janeiro, junho de 2019

Edição Preliminar

S123 Sales, Luana Farias

Dados de pesquisa: quem ama cuida / Luana Farias Sales e Luís Fernando Sayão ; Ilustração de: Joana Sales Marques. – Brasília, DF : Comissão Nacional de Energia Nuclear (Brasil) : IBICT, 2019.
XX f. ; il. color. (Série: Divulgação em Ciência Aberta) ; Rio de Janeiro, 2019.

ISBN

Inclui bibliografia:

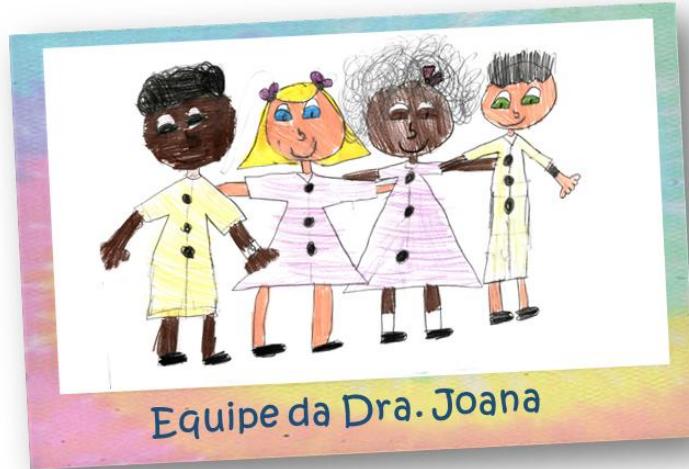
Cartilha

1. Curadoria de dados de pesquisa. 2. Pesquisa e Inovação. I. Sayão, Luís Fernando. II. Marques, Joana Sales. III. Comissão Nacional de Energia Nuclear. IV. Ministério de Ciência, Tecnologia, Inovação e Comunicação. V. Instituto Brasileiro de Informação em Ciência e Tecnologia. VI. Título.

CDD:

APRESENTAÇÃO

Olá meu nome é **Joana**! Trabalho em um laboratório coordenando uma equipe de pesquisadores



As nossas pesquisas produzem muitos dados valiosos para a equipe e também para outros pesquisadores.

AMO meus dados, mas às vezes me sinto afogada sem saber como lidar com eles. Porém, sei que precisam estar organizados e compreensíveis para valorizar o trabalho de pesquisa da minha equipe, pois **boas pesquisas exigem bons dados!!!**

Por esse motivo, estou aqui para compartilhar com vocês, meus colegas pesquisadores, algumas dicas de como tratar bem os seus dados de pesquisa.



1 INTRODUÇÃO

Meus dados são tão importantes como **meus artigos**. Então eu preciso aprender a fazer sua **gestão durante todo o transcorrer da minha pesquisa**. A gestão de dados vai permitir que os meus dados:

- ✓ Estejam **PLANEJADOS** e **ORGANIZADOS**;
- ✓ Estejam **DOCUMENTADOS** para que possam ser **COMPREENDIDOS** por mim e por outros pesquisadores e tenham sua **PROVENIÊNCIA** comprovada;
- ✓ Possam estar armazenados de forma **SEGURA** e **PRESERVADOS** para usos atuais e futuros;
- ✓ Possam ser **COMPARTILHADOS, ACESSADOS** e **REUSADOS** para novas pesquisas;
- ✓ Possam ser **PUBLICADOS, CITADOS** e que minha **AUTORIA** seja reconhecida;
- ✓ Estejam alinhados às **POLÍTICAS** das **agências de fomento** e da minha **instituição**.
- ✓ Respeitem os princípios **ÉTICOS** e **LEGAIS** pertinentes;
- ✓ Estejam **CONTEXTUALIZADOS** por meio de links com outros recursos.



Uma **boa gestão de dados** leva em conta aspectos: técnicos, organizacionais, disciplinares, estruturais, legais, éticos e de sustentabilidade econômica. Hoje ela é uma parte essencial da **condução responsável dos projetos de pesquisa**. Além do mais vai se tornando também parte da **cultura da profissão de pesquisador**.

2 PLANEJE SEUS DADOS

A gestão de dados de pesquisa começa pelo **planejamento**. Esta é uma fase essencial que deve se **iniciar** quando a pesquisa ainda está sendo delineada e deve considerar como os seus dados serão **cuidados durante o desenvolvimento do projeto**, como eles serão **arquivados e preservados** depois que o **projeto estiver finalizado**, e ainda, como eles serão **compartilhados**, respeitando as condições do pesquisador, das instituições envolvidas e os padrões éticos e legais.

Muitas perguntas já podem ser formuladas e respondidas nesse momento:



É bom lembrar que este planejamento tem outro objetivo: servir de base para a elaboração de um **documento muito importante** exigido pela maioria das agências de fomento que vão financiar o seu projeto de pesquisa e às vezes pela sua instituição: trata-se do **PLANO DE GESTÃO DE DADOS (PGD)**

Plano de Gestão de Dados de Pesquisa

A ideia do PGD é planejar como gerar/coletar dados de alta qualidade e sustentáveis ao longo do tempo. A **estrutura do documento** varia de acordo com o padrão adotado pela sua instituição e/ou pela agência de fomento que vai financiar o seu projeto. Eis algumas informações que podem constar num PGD:



Assim, o PGD ajuda o seu projeto a se ajustar às **políticas mandatórias** dessas instituições e a implementar a gestão de dados de **acordo com as necessidades e propósitos da pesquisa**.

Ele é também um documento formal que explicita como os dados serão tratados **durante o desenvolvimento do projeto e depois que ele finalizar**. Portanto, o PGD é um documento “vivo”, que **vai mudando junto com as necessidades do projeto e de seus participantes**.

O Plano de Gestão de Dados é um elemento chave da boa gestão de dados (Comissão Europeia, 2016).

Para ajudar você desenvolver o seu PGD existem ferramentas automatizadas e exemplos de *templates* na Web, como por exemplo a Data Management Platform tool (**DMP tool**).

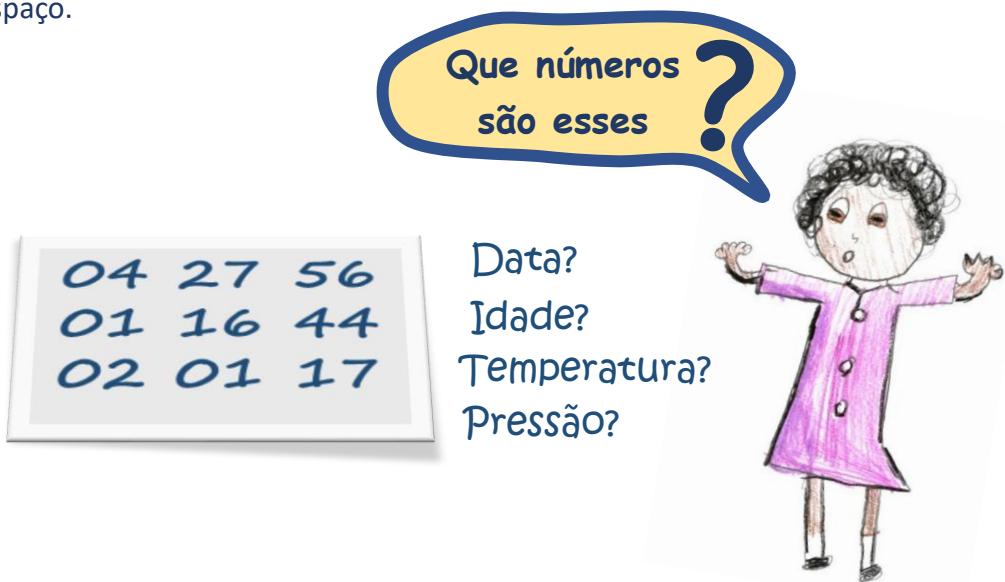
3 DOCUMENTE SEUS DADOS

Faça como Galileu Galilei, descreva seus dados para que todos os compreendam agora e no futuro

O caderno de notas de Galileu integrava seus **DADOS** (desenhos de Júpiter e suas luas), **METADADOS**-chaves (cronometragem de cada observação, condições meteorológicas, propriedades do telescópio) e **TEXTO** (descrição dos métodos, análises e conclusões). Isto permitiu que ele compreendesse o Sistema Solar como permitiu também que seus leitores compreendessem como ele chegou às suas descobertas.



Uma parte importante da gestão de dados é assegurar que as coleções de dados possam ser **compreendidas e interpretadas** por você mesmo, por sua equipe e por outros pesquisadores. **Isto porque os dados não são como artigos e livros, eles não falam por si próprio, precisam de informações adicionais que os contextualizem para que se tornem informação e transmitam conhecimento ao longo do tempo e do espaço.**



Para serem compreendidos os dados precisam ser descritos e contextualizados!

Para que seus dados sejam encontrados, identificados, compreendidos, reusados e tenha sua **proveniência** comprovada, eles devem estar acompanhados de um conjunto de informações que descreva todos os seus aspectos.



A **descrição** que acompanha os dados pode estar em duas formas:

METADADOS – conjunto de informações padronizadas e estruturadas que descrevem os dados informando sobre PROVENIÊNCIA, ORIGEM, PROPÓSITO, AUTORES, INSTITUIÇÕES, DATAS, LICENÇAS, DETALHES TÉCNICOS. Os **metadados disciplinares** são voltados para áreas específicas descrevendo com mais precisão os dados. Os metadados são lidos por pessoas e também pelos computadores dos buscadores como o Google Acadêmico, dando mais visibilidade para sua pesquisa.

DOCUMENTAÇÃO – documentos que apoiam o entendimento sobre como os dados foram COLETADOS, GERADOS, PROCESSADOS e de como estão ESTRUTURADOS, ORGANIZADOS e NOMEADOS. Por exemplo: Projeto de pesquisa, caderno de laboratório e de campo, roteiro de entrevista, livro de códigos, arquivo “leia-me” e muito mais.

“

A boa documentação dos dados de pesquisa é essencial para que eles sejam gerenciados, identificados, publicados, encontrados, citados, reusados e tenham seu valor creditado aos seus autores



4 IDENTIFIQUE SEUS DADOS

para que eles sejam mais citados

As coleções de dados de pesquisa, assim como os artigos de periódicos, podem ser **identificadas** por esquemas de identificadores persistentes, padronizados e globalmente únicos.

A identificação apropriada vai permitir que seus dados sejam recuperados, citados e preservados e a autoria creditada a você e aos seus colaboradores.

Porém, a **identificação pessoal do pesquisador** torna ainda mais precisa a identificação das suas coleções de dados.

Então, **dois tipos de identificadores** são muito importantes:



IDENTIFICADOR PERSISTENTE PARA OS DADOS

Para que possam ser citadas corretamente, as suas coleções de dados precisam de um identificador persistente. O **DOI** (Digital Object Identifier) usado com frequência para os artigos acadêmicos, e pode ser aplicados também aos dados.

IDENTIFICADOR PERSISTENTE PARA O AUTOR

Uma forma de ligar os seus dados a você é criar um identificador persistente pessoal. O **ORCID** é comumente usado, permite que você seja distinguido de outros pesquisadores e conecta você as suas outras atividades de pesquisa.

Lembrete: atribua também um **título descritivo** para a sua coleção de dados

5 CONTEXTUALIZE SEUS DADOS

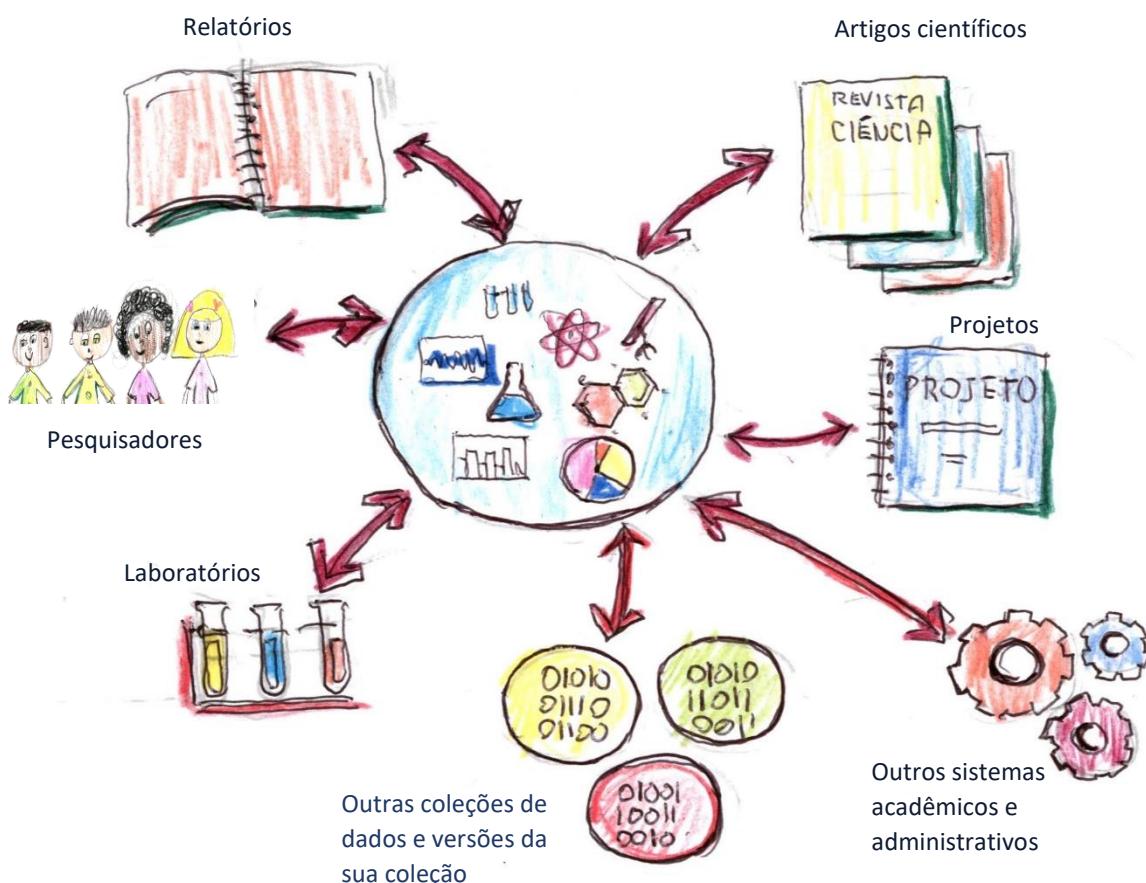
crie links com outros recursos

Além dos metadados e da documentação que ajudam a dar significado aos dados, é muito importante que eles estejam linkados com outros recursos informacionais criando uma “ecologia” de dados e informações.

Os seus dados devem estar minimamente linkados à literatura acadêmica escritas a partir deles, como seus artigos publicados em periódicos. Inversamente, os artigos devem referenciar os dados da mesma forma que você referencia outros recursos.

Mas os dados podem estar linkados também às suas próprias versões, a outros conjuntos de dados, a projetos, laboratórios, pesquisadores, relatórios, etc.

Essa contextualização naveável vai permitir que os dados sejam compreendidos com mais precisão e que possam ser localizados e acessados por várias vias, por exemplo, por meio de um *link* num artigo. E, consequentemente, vai permitir uma maior visibilidade e para a sua pesquisa e mais citação para seus artigos e dados.



6 PROTEJA SEUS DADOS



Os seus dados estão sujeitos a diversos riscos que os ameaçam todo o tempo: dano físico, perda, roubo, acesso indevido por *hackers*, por exemplo. Os dados em formatos digitais são objetos digitais que ainda sofrem com a obsolescência tecnológica!

Para manter a integridade e a confidencialidade dos seus dados, é necessário que eles sejam protegidos adequadamente durante o andamento do seu projeto, até que sejam arquivados em um ambiente confiável, como um repositório ou centro de dados.

Acesso Físico – controle o acesso de pessoas onde dados, computadores e mídias são mantidos.

Senhas – troque frequentemente as senhas para assegurar que somente os membros atuais do projeto possam acessar os dados.

Antivírus - mantenha versões atualizadas de proteção em todos os computadores.

Firewall – Se o seu computador está conectado à Internet, use *firewall*.

Armazenamento de dados confidenciais – não armazene dados confidenciais em computadores ligados a redes externas.

Envio de arquivos – dados pessoais ou confidenciais devem ser criptografados antes de serem enviados.

7 LICENCIE SEUS DADOS

Quando você escolhe um repositório para publicar seus dados, a eles é atribuído uma licença. Esta licença é um acordo legal entre você, que é o criador dos dados ou o depositante, e o repositório que vai cuidar dos seus dados.

A licença associada aos seus dados vai estabelecer o que um outro pesquisador, que tem a intenção de reusar os seus dados tem permissão para fazer.

Os pressupostos da ciência aberta preconizam licenças que permitem:

- ✓ Que os dados estejam disponíveis para a mais ampla audiência possível;
- ✓ Que possam ser aplicados a uma ampla variedade de reuso.



Entretanto, você pode aplicar licenças mais restritivas, pois em alguns casos as licenças mais abertas são inapropriadas. Por exemplo:

Coleções de dados contendo dados pessoais sensíveis cujo termo de consentimento não está claro;

Dados de projetos patenteáveis;

Quando não há permissão do detentor do *copyright*.

As licenças *Creative Commons* fornecem uma ampla faixa de licenças que podem ser aplicadas a dados, coleções de dados e base de dados.

Mas é importante lembrar que as licenças que você pode atribuir aos seus dados são determinadas ou limitadas pelo repositório que você escolher para publicar os seus dados.

8 TORNE SEUS DADOS ÉTICOS E LEGAIS

Coletar, usar e compartilhar dados no decorrer de uma pesquisa científica exige que padrões éticos e obrigações legais sejam respeitados. No âmbito da pesquisa, as leis mais relevantes são as concernentes à **proteção de dados** e à **propriedade intelectual**. Porém, tendo em vista o alcance das colaborações entre pesquisadores, as **leis internacionais** que porventura possam afetar um projeto de pesquisa devem também ser observadas.

A ÉTICA DOS DADOS DE PESQUISA

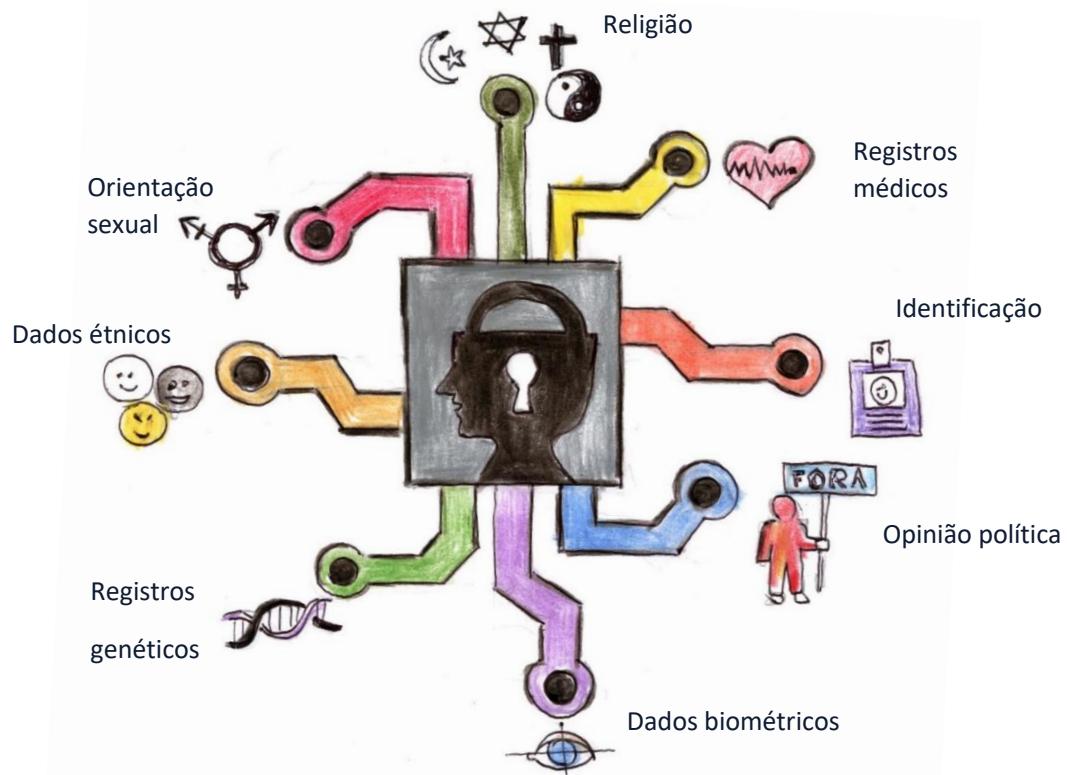
Especialmente quando a pesquisa envolve pessoas, espera-se que o pesquisador mantenha um comportamento pautado por um **rigoroso código de ética** que deve permear todo o ciclo da pesquisa, especialmente a fase de **compartilhamento de dados**. Neste contexto, dois tipos de dados são muito importantes:

DADOS PESSOAIS

São dados relacionados a indivíduos vivos que podem ser identificados a partir desses dados ou da combinação deles com outras informações.

DADOS PESSOAIS SENSÍVEIS

São dados sobre raça, origem étnica, opinião política, religião ou crenças filosóficas, orientação sexual, saúde, registros biomédicos e genéticos, etc.



TERMO DE CONSENTIMENTO INFORMADO

Consentimento informado se refere ao processo de comunicação por meio do qual um pesquisador repassa informações para que um indivíduo faça escolhas informadas sobre a sua participação voluntária numa pesquisa.

ANONIMIZAÇÃO DOS DADOS

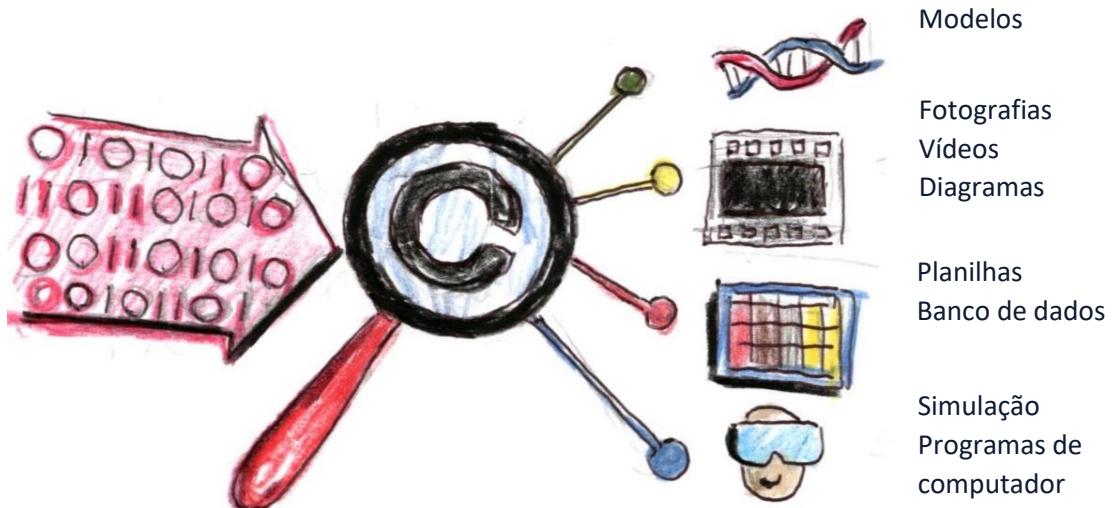
Compreende uma série de procedimentos a que os dados que contém informações pessoais são submetidos para que não revelem a identidade de indivíduos.

COPYRIGHT

Copyright é um direito de propriedade intelectual reconhecido internacionalmente e atribuído automaticamente a um autor de uma obra original, como uma pesquisa científica.

Entretanto, o **copyright não se aplica a fatos, ideias e conceitos**, mas na forma particular como eles são expressos. Por exemplo, no caso de dados de pesquisa em forma de **planilhas, banco de dados, programas de computador, modelos, relatórios etc.**

Portanto, *copyright* se aplica também a dados de pesquisa e é um item importante a ser considerado na criação, compartilhamento e reuso de dados.

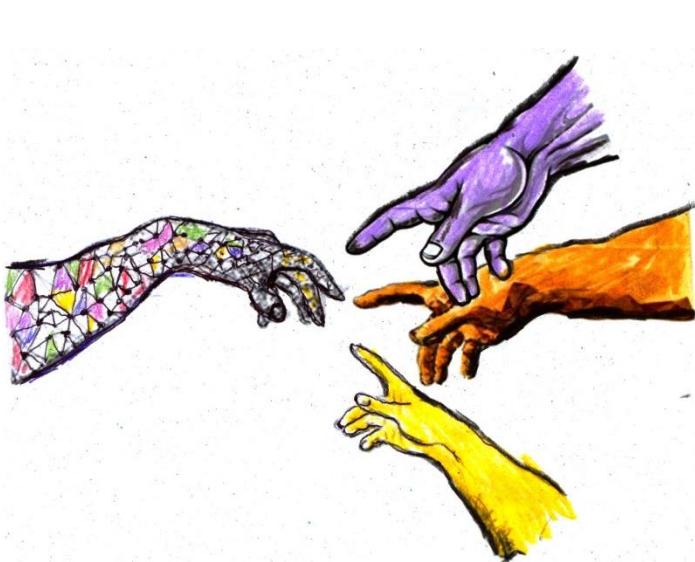


Assim, os conflitos entre a proteção e o compartilhamento de dados pessoais podem ser superados pela combinação do consentimento informado, anonimização, controle de acesso e *copyright*. Dessa forma, esses dados podem ser ética e legalmente usados, reusados e compartilhados.

9 COMPARTILHE SEUS DADOS

Ame seus dados, mas não a ponto de escondê-los. Deixe que outros pesquisadores também os amem. **O compartilhamento de dados faz parte do processo científico!!!**

Seus dados de pesquisa podem ser reusados agora e no futuro por outros pesquisadores de sua área e de outras disciplinas, pois eles têm um valor que extrapola o seu valor original. São muitas as vantagens pra você e sua equipe, vejamos algumas:



Evita a **duplicação de esforços**

Incentiva o **debate científico**

Promove **novos usos** para os dados

Possibilita **novas formas de colaboração**

Possibilita a **validação das pesquisas**

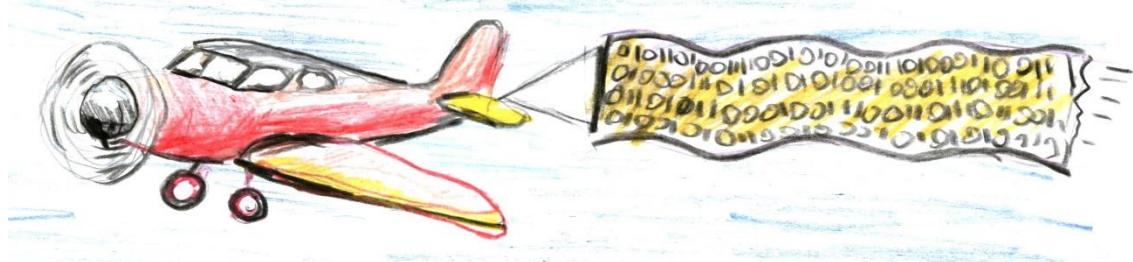
Compartilhar seus dados vai trazer mais **visibilidade** para a sua pesquisa, mais **citações** para os seus artigos, credita a sua **autoria** e dá mais **credibilidade** à sua **pesquisa!!!**

Mas você precisa **publicar os seus dados!!!** Como? Veremos adiante.



10 PUBLIQUE SEUS DADOS

Mostre a todo mundo o quanto você ama seus dados. Publique-os (mesmo que você não queira exibi-los ainda)



Você pode compartilhar seus dados na página Web do seu projeto de pesquisa ou na do seu laboratório. Mas essas páginas têm vida curta e não garantem a integridade e persistência dos seus dados.

Se você quer que suas coleções de dados possam ser **identificadas, documentadas, que mantenham-se íntegras, seguras, contextualizadas e preservadas, licenciadas e possam ser compartilhadas dando mais visibilidade à sua pesquisa**, o melhor é depositar num **repositório de dados de pesquisa**.

O ideal é **publicar** os seus dados em um repositório de dados da sua área de estudo, ou seja, um **repositório disciplinar**, que vai cuidar dos seus dados por longo prazo, mantendo-os íntegros e autênticos. Isso é verdade mesmo quando você não deseja torná-los públicos ainda. Os repositórios cuidam dos seus dados sem torná-los publicamente acessíveis pelo tempo necessário. Veja as vantagens:



Como encontrar um repositório para publicar meus dados?

Há diretórios como o **R3data** que indexam e classificam os repositórios de dados. Lá você pode encontrar um repositório perfeito **para publicar seus dados**, ou encontrar **coleções que você pode reusar**.

Mas se não houver um repositório disciplinar adequado aos seus dados você pode depositar nos repositórios multidisciplinares como o **Dryad**, **Dataverse**, **Figshare**, ou ainda no **repositório institucional** da organização onde você faz as suas pesquisas.



Lembre-se: muitos periódicos científicos exigem que você publique seus dados em um repositório de dados !!!

DATA JOURNALS !!!

Você pode publicar ainda um **Data Paper**, ou seja, um artigo descrevendo os seus dados num **Data Journal**. O seu artigo faz um *link* com os dados depositados num repositório. É mais divulgação e visibilidade para a sua pesquisa!



Para finalizar, FINAL FELIZ!!!

Seguindo essas orientações, seus dados poderão ser mais facilmente

Identificados, localizados e acessados;

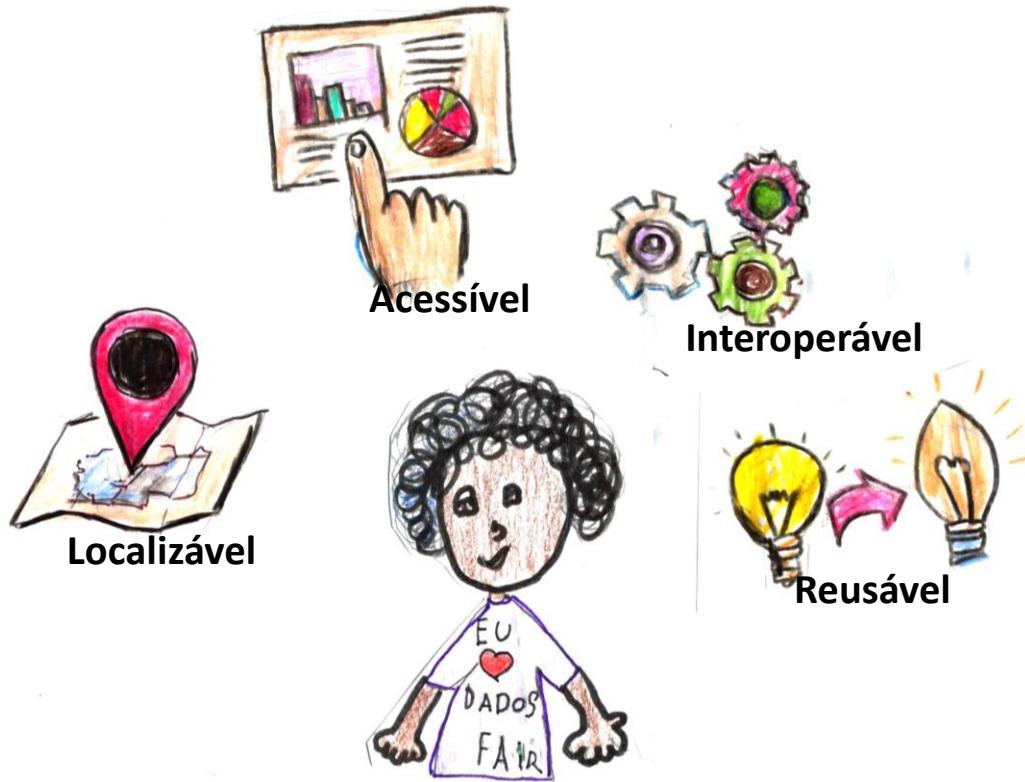
Eles estarão inteligentemente disponíveis e poderão ser compreendidos e reusados por outros pesquisadores, agora e no futuro;

Poderão estar linkados com outros materiais de pesquisa como artigos, teses e outros conjuntos de dados, criando uma contextualização em torno dos seus dados.

Estarão seguros e digitalmente preservados;

E respeitarão os princípios éticos, legais e os direitos a eles associados, como *copyright*.

E, PORTANTO, TERÃO MAIOR **VALOR** PARA VOCÊ E PARA OUTROS PESQUISADORES. ELES ESTARÃO TAMBEM MAIS PRÓXIMOS DOS **PRINCÍPIOS FAIR**, sigla em inglês para **Findable, Accessible, Interoperable, Reusable**, adotados internacionalmente.



Referências Citadas e Obras Consultadas

CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES. Training Working Group. **Data Management Expert Guide**. Bergen, Norway: CESSDA ERIC, 2017-2018. Disponível em: <<https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide>>. Acesso em: 09 maio 2019.

CORTI, Louise et al. **Managing and data sharing**: best practice for researchers. Colchester : Uk Data Archive, 2011. Disponível em: <<https://ukdataservice.ac.uk/media/622416/trainingresourcespack.pdf>>. Acesso em: 09 maio 2019.

EUROPEAN COMMISSION. **Guidelines on FAIR Data Management in Horizon 2020**. July 2016. Disponível em: <https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf>. Acesso em: 09 maio 2019.

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

