

Gerenciando Dados da Pesquisa

Agenda

Introdução

FAIR Principles

Onde publicar meus dados?

Referências e Dicas úteis

Introdução

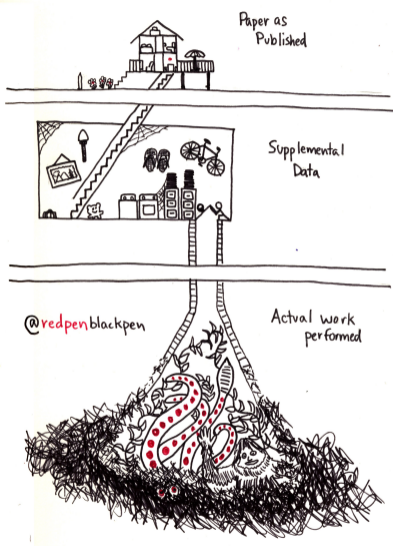


Figure 1: Don't mind the monsters in the basement. Disponível em: @redpenblackpen twitter

O que são dados?

Não existe um formato ou definição específica. Dados podem ser coletados ou criados, podem ser digitais ou físicos...

No contexto geral de uma pesquisa, o ciclo de vida é:

- ▶ **coletar** dados
- ▶ **processá-los** num formato de análise/exploração
- ▶ **publicá-los** por meio de artigos/monografia

Ao realizar uma gestão dos dados, evitamos que o ciclo de vida seja **caótico**

Imagine que você precise refazer a busca do seu mapeamento sistemático. Você possui todas as strings que usou ou deveria criar tudo novamente?

Dados... abertos?

Uma boa gestão de dados apoia a **Ciência Aberta**

Pessoas e **Sistemas** podem compreender, mas nem todos os dados de uma pesquisa podem ser abertos

- ▶ Informações pessoais (e-mail, nome, endereço)
- ▶ Dados que podem gerar inovação ou vantagem competitiva
- ▶ Lei Geral de Proteção de Dados
- ▶ ...

Compartilhar e gerenciar dados envolve: segurança, confidencialidade, ética e propriedade intelectual

FAIR Principles

Capacidade de encontrar, acessar, interoperar e reutilizar dados

[F]indable

- ▶ F1. (Meta)dados são atribuídos a um **identificador único e persistente**
- ▶ F2. Dados devem ser descritos utilizando metadados ricos
- ▶ F3. Os metadados incluem explicitamente o identificador dos dados que descrevem
- ▶ F4. Os metadados são registrados ou indexados por um **mecanismo de busca**

Ou, simplificando... seus dados podem ser encontrados pelo maior número de pessoas e de sistemas

[F]indable

- ▶ Buscar repositórios certificados e bem conhecidos
 - ▶ Buscar repositórios com apoio a identificador único
 - ▶ DOI
 - ▶
 - ▶ Verificar políticas e restrições de acesso do repositório (licenças abertas, e-mail...)
 - ▶ Buscar repositórios com bons critérios de classificação (assuntos, temas, disciplinas...)
 - ▶ **Entender o que a sua comunidade usa!**
-
- ▶ <https://www.re3data.org/>

[A]ccessible

Além de encontrar os dados precisamos auxiliar que outros outros usuários saibam como acessá-los

- ▶ A1. (Meta)dados são recuperáveis por seu identificador usando um protocolo **padrão** de comunicação
 - ▶ A1.1 O protocolo é aberto, gratuito e universalmente implementável
 - ▶ A1.2 O protocolo permite uma autenticação e autorização quando necessário
- ▶ A2. Os **metadados são acessíveis**, mesmo quando os dados não são mais acessível

[A]ccessible

Todo mundo pode acessar seus dados? Sob quais **condições**?

Como seus dados estão **documentados**?

Registrar dados de acesso, uso ou reuso dos dados

- ▶ Isso pode gerar mais informações sobre como a comunidade interage com o seu projeto

Boas práticas:

- ▶ Criar um arquivo de orientação sobre os dados (*readme* ou *leiam*)
- ▶ Usar uma estrutura de diretórios coesa e consistente

[I]nteroperable

Os dados geralmente precisam ser integrados a outros dados. Ou ainda, precisam interoperar com softwares, aplicativos, pessoas, fluxos de trabalho...

- ▶ I1. Os (meta)dados usam uma **linguagem** formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.
- ▶ I2. (Meta)dados usam **vocabulários** que seguem os princípios FAIR
- ▶ I3. (Meta)dados incluem referências qualificadas a outros (meta)dados

[I]nteroperable

Usar padrões de metadados **bem estabelecidos** dentro da sua comunidade e que são **relevantes** para a sua pesquisa

- ▶ Dublin Core
- ▶ IEEE LOM
- ▶ OBBA
- ▶ ...

Aplicar termos de vocabulários ou padrões **técnicos** ou **científicos**

- ▶ ACM Computing Classification System: <https://dl.acm.org/ccs>
- ▶ Padrões IEEE
- ▶ ...

Usar formatos abertos ao invés de formatos proprietários

- ▶ **Cuidado durante o processo de conversão**

[R]eusable

Meta(dados) devem ser bem descritos para que possam ser replicados e/ou combinados em diferentes configurações

- ▶ R1. Meta(dados) são descritos de forma rica, com uma pluralidade de atributos precisos e relevantes
 - ▶ R1.1. (Meta)dados são liberados com uma **licença de uso** de dados clara e acessível
 - ▶ R1.2. (Meta)dados são associados à proveniência detalhada
 - ▶ R1.3. (Meta)dados atendem aos padrões da comunidade relevantes para o domínio

[R]eusable

Adotar **licenças abertas** ao disponibilizar os materiais

- ▶ Creative Commons
- ▶ Evitar regras autorais ou licenças pouco conhecidas
- ▶ ...

Citar os dados de modo padronizado no seu artigo/monografia

- ▶ Eleazar Morales Díaz, & Susana Vila Melero. (2021). *Katas (programming exercises) and related statistics (1.0.0)* [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.4664585>

Onde publicar meus dados?

Buscar conhecer **como e onde a comunidade divulga e trabalha**

- ▶ Quem apoiou/criou? Iniciativas governamentais, privada...
- ▶ Tamanho e tempo do acervo

Observar as funcionalidades e suporte:

- ▶ Apoio a comunidades/disciplinas da sua pesquisa
- ▶ Links permanentes
- ▶ Limites de armazenamento de tempo e de espaço
- ▶ Apoio ao **versionamento**
- ▶ DOI e ORCID
- ▶ ...

Onde publicar meus dados?

Evitar publicações em site pessoal, locais pouco conhecidos ou com problemas de estrutura

- ▶ **Drives** pessoais
- ▶ Considerar sempre os **princípios FAIR**

Usar repositórios de dados:

- ▶ Zenodo: <https://zenodo.org/>
- ▶ Mendeley Data: <https://data.mendeley.com/>
- ▶ ...

Seus dados podem gerar publicações em periódico:

- ▶ Data in Brief: <https://www.journals.elsevier.com/data-in-brief>

Exemplos

A dataset for emotional reactions and family resilience during COVID-19 isolation period among Indonesian families

- ▶ Dataset: [link](#)
- ▶ Data in Brief: [link](#)
- ▶ Estatística: [link](#)

How do developers utilize source code from stack overflow?

- ▶ Dataset: [link](#)
- ▶ Estudo: [link](#)

Referências e links úteis

- ▶ The FAIR Guiding Principles for scientific data management and stewardship:
<https://www.nature.com/articles/sdata201618>
- ▶ Elaboração e esclarecimento de dúvidas sobre PLANOS DE GESTÃO DE DADOS:
<https://www.youtube.com/watch?v=iLwcLVpXjuU>
- ▶ FAIR Principle: <https://www.go-fair.org/fair-principles/>
- ▶ Curso de Gestão de dados da Elsevier:
<https://researcheracademy.elsevier.com/research-preparation/research-data-management>