



# Princípios FAIR aliados ao acesso e reuso de dados

análise de conjuntos de dados sobre Covid-19  
presentes no repositório PubChem

**Tainá Regly** – IBICT

**Viviane Veiga** – Fiocruz

**Aline Alves** – Fiocruz

Pub  hem

# Introdução



O movimento de Ciência Aberta vem se tornando, cada vez mais, aceito pela comunidade científica que, por sua vez, é uma grande incentivadora do compartilhamento de dados de pesquisa justamente pelo seu potencial de reprodutibilidade de estudos e aceleração das descobertas.

Covid-19 e da investigação do seu tratamento, prevenção e vacinas, a pandemia resultou também na adoção de preceitos da Ciência Aberta, tal como a partilha de dados, a publicação de preprints e o acesso aberto às publicações.

Os repositórios de dados são um excelente ambiente que propicia esses benefícios, mais a visibilidade, compartilhamento e curadoria desses dados.

## PubChem

foi selecionado para fazer parte deste trabalho por deter essas características importantes para a Ciência Aberta e devido à sua capacidade de ser uma importante fonte de informação mantida pelo National Institutes of Health (NIH) que proporciona acesso aos usuários que desejam ter acesso a dados químicos.

## Princípios FAIR

encontráveis, acessíveis, interoperáveis e reusáveis

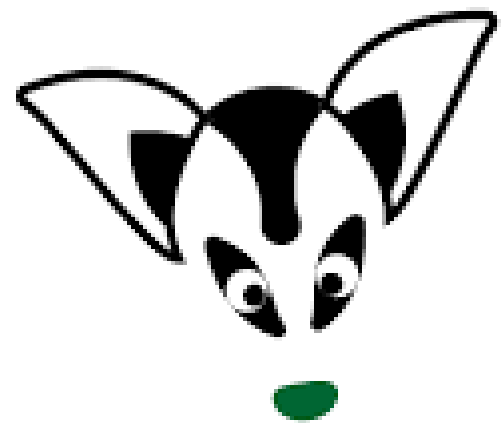
sua utilização impulsiona descobertas científicas mais velozes e fundamentadas, incluindo a leitura automatizada de máquinas tanto dos dados como dos metadados.



# Objetivo

Avaliar e verificar o nível de aderência de conjuntos de dados sobre o medicamento Cloroquina, oferecidos pelo PubChem, aos princípios FAIR.

# Método



**Fair Data** **BR**

GRUPO	CONJUNTO DE DADOS
Composto	Chloroquine
Substância	Chloroquine
Proteína	Chloroquine resistance transporter
Bioensaio	Resistance modification index, ratio of IC50 for chloroquine in presence of compound to IC50 for chloroquine alone for chloroquine-resistant Plasmodium falciparum W2
Patente	Chloroquine analogs and methods of preventing and treating plasmodial disease

# Encontrável

nota 7.60



- apesar de fazer uso de identificadores persistentes, os recursos digitais não poderem ser encontrados através de mecanismos de busca na web
- os metadados utilizados que são estruturados
- o conjunto de dados que está publicado em um repositório específico de domínio.

# Acessível

nota 8.33

- os dados são acessíveis ao público, inclusive através do identificador fornecido, podendo ser baixados sem a necessidade de cadastro
- uso de protocolos HTTP
- não há necessidade de intermediação de protocolos ou ferramentas especializadas e fazer uso de API de Web Service padrão
- O único ponto limitativo deste princípio está atrelado ao aspecto de que o repositório não deixa claro se os metadados estarão acessíveis mesmo quando os dados não mais o estiverem



# Interoperável

nota 7.75



- conjuntos de dados disponíveis em formatos preferenciais - XML e JSON
- uso de instrumentos de controle terminológico com identificador global, tal como ontologias e RDF
- os conjuntos de dados estão vinculados a outros dados a partir de identificadores
- O único ponto desfavorável em relação à interoperabilidade do repositório se refere ao fato de o esquema de metadados utilizado para descrever os dados não ser formalizado ou aprovado pela comunidade - esquema de metadados próprio

# Reusável

nota 8.20



- O PubChem não exemplifica de maneira clara se os dados são licenciados, deixando para os responsáveis pelos dados a incumbência de revelar a licença preferida, que nem sempre é indicada
- Foram encontradas menções a licenças que, por sua vez, são abertas
- Os conjuntos de dados possuem proveniência detalhada através do redirecionamento para a página onde estão contidos os dados originais e, em alguns casos, a proveniência é acompanhada pelo seu histórico de versões



# Considerações finais

- O repositório PubChem alcançou como média a nota 7.97
- acessibilidade obteve o melhor desempenho enquanto o da encontrabilidade o pior
- os conjuntos de dados estão, majoritariamente, alinhado aos princípios FAIR e seus dados estão aptos a serem compartilhados e reutilizados por pesquisadores
- modificações devem ser implementadas no repositório, tal como a formalização do esquema de metadados utilizado, o uso de licenciamento nesses campos, a inclusão de identificadores em todos os dados e a garantia de que os metadados estarão disponíveis mesmo quando o conjunto de dados não estiver mais acessível.



**Obrigada!**

**Tainá Regly**

taina.regly@gmail.com

**Viviane Veiga**

viviane.veiga@iciict.fiocruz.br

**Aline da Silva Alves**

aalves.fiocruz@gmail.com,

