



# METADADOS PARA REPRESENTAÇÃO DE DADOS EM COVID-19: um estudo exploratório

Autores: Anderson Silva de Araujo

Viviane Santos de Oliveira Veiga

Isabella Henrique Lima Pereira

Mylena Cristhina Araujo de Oliveira



## RESUMO

Apresentar os padrões de metadados utilizados pelos repositórios de dados de pesquisa que disponibilizam conjuntos de dados em Covid-19 em acesso aberto.

13<sup>a</sup> conferência lusófona  
**CIÊNCIA ABERTA**  
Maputo - Moçambique  
10 a 12 outubro 2022



**DADOS EM COVID-19**

13<sup>a</sup> conferência lusófona  
**CIÊNCIA ABERTA**  
Maputo - Moçambique  
10 a 12 outubro 2022

F  
Findable



A  
Accessible



I  
Interoperable



R  
Reusable



Localizáveis, Acessíveis,  
Interoperáveis e Reutilizáveis.

13<sup>a</sup> conferência lusófona  
**CIÊNCIA ABERTA**  
Maputo - Moçambique  
10 a 12 outubro 2022



Virus Outbreak Data Network (VODAN)





Para que conjuntos de dados sejam FAIR, dados e metadados precisam estar alinhados a estes princípios.



A função dos metadados é garantir a padronização dos recursos informacionais, pautados em esquemas e regras internacionais na tentativa de facilitar a identificação, a busca, a localização, a recuperação, a preservação, o uso e o reuso.



Esquema de metadados é uma lista de propriedades principais de metadados escolhidas para uma identificação consistente de um recurso para fins de citação e recuperação, juntamente com instruções de uso recomendado (DataCite Metadata Working Group, 2017).



## METODOLOGIA

Foram realizados levantamentos bibliográficos e documental. Para o levantamento bibliográfico recorreu-se às bases *Web of Science e PubMed*, nas quais foram aplicados os seguintes termos: covid; interoperability; covid-19; interoperable; sars vírus; metadata standards; sars-cov-2; metadata; sars; medical records; coronavírus; coronavirus disease.



## METODOLOGIA

Ao analisar literatura levantada sobre metadados e covid-19, identificou-se a presença de 1 (um) padrão de metadados em artigo indexado na Web of Science e 3 (três) iniciativas de padrões de metadados em artigos indexados noPubMed. Na Web of Science o metadado foi encontrado no artigo “COVID-19 pandemic reveals the peril of ignoring metadata standards”. O Genomic Standards Consortium (GSC, [www.gensc.org](http://www.gensc.org) ).



## METODOLOGIA

Na PubMed foram encontradas três iniciativas o Outbreak.info no artigo “Outbreak.info Research Library: A standardized, searchable platform to discover and explore COVID-19 resources and data”



## METODOLOGIA

o PHA4GE no artigo “GA4GH: Políticas e padrões internacionais para compartilhamento de dados em pesquisa genômica e saúde” , e o GISAID no artigo “Interoperable medical data: The missing link for understanding COVID-19”.



## METODOLOGIA

Após esta etapa foi realizado um mapeamento, no Re3Data, dos repositórios de dados com conjuntos de dados de pesquisa em COVID-19.



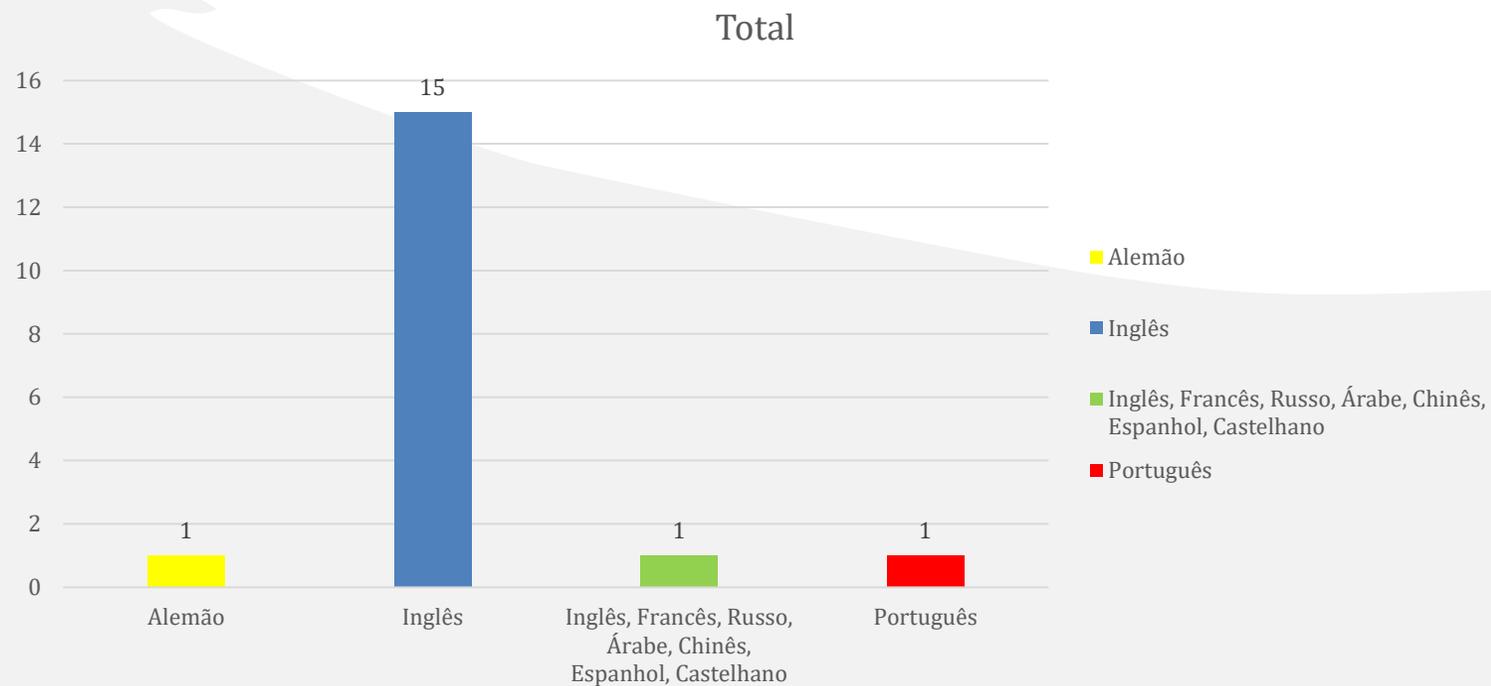


## METODOLOGIA E RESULTADOS

Utilizamos a seguinte estratégia: utilização do termo “data repository”, selecionamos o filtro Data access, no qual optamos pelo campo Open que nos deu o resultado 843. Em seguida, utilizamos o campo das palavras chave (Covid-19) e obtivemos um resultado final de 18 repositórios.

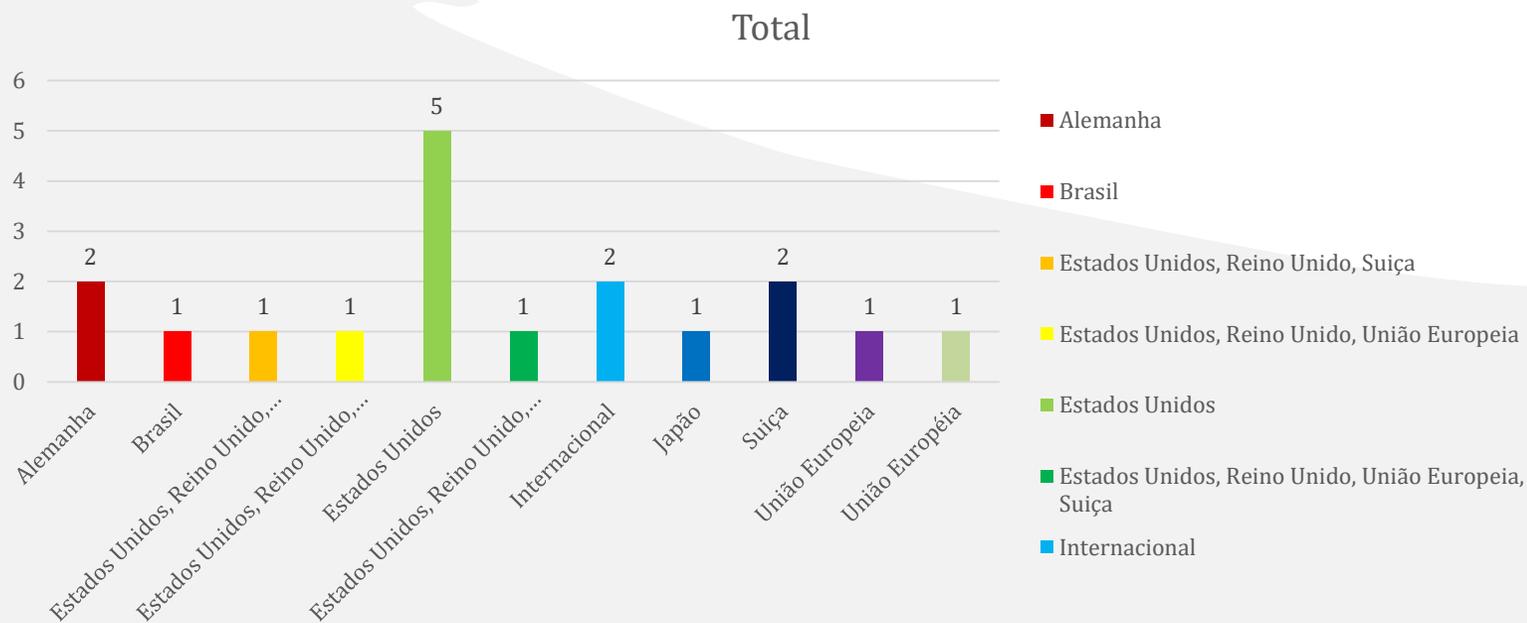
# RESULTADOS

## IDIOMAS DOS REPOSITÓRIOS



# RESULTADOS

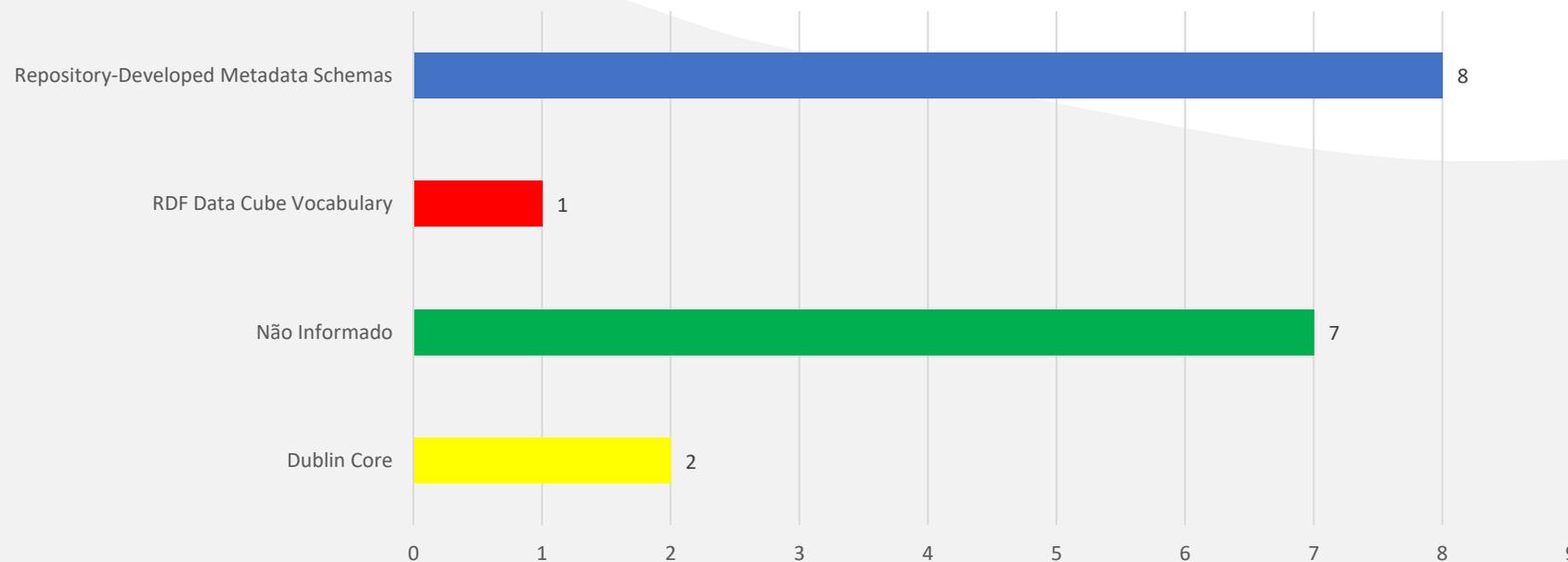
## PAÍSES DOS REPOSITÓRIOS



# RESULTADOS

## PADRÃO DOS METADADOS

Contagem de NOME por NOME PADRÃO METADADOS





## RESULTADOS

### PADRÃO DOS METADADOS

O RDF Data Cube Vocabulary fornece um meio de publicar dados multidimensionais, como estatísticas, na web de forma que possam ser vinculados a conjuntos de dados e conceitos relacionados usando o padrão W3C RDF (Resource Description Framework).



## RESULTADOS

### PADRÃO DOS METADADOS

O Dublin Core é um esquema de metadados que visa descrever objetos digitais, tais como, vídeos, sons, imagens, textos e sites na web. Aplicações de Dublin Core utilizam XML e o RDF (Resource Description Framework).



## RESULTADOS

### PADRÃO DOS METADADOS

A maioria dos repositórios analisados identificou a necessidade de desenvolver metadados próprios para representar os conjuntos de dados em seus repositórios. Isto aponta a necessidade da criação de padrões de metadados, analisados e validados pela comunidade de domínio e da comunidade científica geral, para garantir a interoperabilidade no campo científico.



## RESULTADOS

### PADRÃO DOS METADADOS

A maioria dos repositórios analisados identificou a necessidade de desenvolver metadados próprios para representar os conjuntos de dados em seus repositórios.

Isto aponta a necessidade da criação de padrões de metadados, analisados e validados pela comunidade de domínio e da comunidade científica geral, para garantir a interoperabilidade no campo científico.



## CONSIDERAÇÕES

Os Estados Unidos, foi identificado como o país com o maior número de repositórios que disponibilizam dados de pesquisa em covid-19.



## CONSIDERAÇÕES

Quanto ao padrão de metadados verificamos que a maioria destes repositórios possui um esquema de metadados próprio, o que pode significar que os esquemas atuais não estão atendendo as demandas de representação descritiva e temática dos dados de pesquisa.



## CONSIDERAÇÕES

Também constatou-se que os repositórios analisados não adotaram os padrões de metadados específicos do campo da saúde, o que pode por um lado comprometer o alinhamento destes dados aos princípios FAIR, e por outro talvez favorecer a interoperabilidade com outras disciplinas.



## CONSIDERAÇÕES

Por fim, constatamos que é necessário um estudo mais aprofundado para verificar a contribuição dos novos esquemas de metadados criados pelos repositórios para a descrição dos conjuntos de dados em Covid-19, e o grau de FAIR destes esquemas.



## REFERÊNCIA

DataCite Metadata Working Group. (2017). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.1. DataCite e.V. 10.5438/0014.  
[https://schema.datacite.org/meta/kernel-4.1/doc/DataCite-MetadataKernel\\_v4.1.pdf](https://schema.datacite.org/meta/kernel-4.1/doc/DataCite-MetadataKernel_v4.1.pdf)