

Infraestruturas de dados da COVID-19: condições para depósito, acesso e usos

Vanessa de Arruda Jorge

Fundação Oswaldo Cruz, Brasil

Sarita Albagli

Instituto Brasileiro de Informação em Ciência e Tecnologia, Brasil

Infraestrutura de dados COVID-19 Acesso Aberto Emergência em Saúde Pública Compartilhamento de Dados	Diante do cenário de emergência em saúde pública de importância internacional da COVID-19 e da importância do compartilhamento e abertura de dados para pesquisa neste contexto, o trabalho tem por objetivo identificar infraestruturas de dados que declaram o termo COVID-19, analisando aspectos e condições de depósito, acesso e uso desses dados, incluindo termos e/ou licenças para sua utilização. As buscas pelas infraestruturas de dados foram realizadas no diretório re3data.org e os resultados foram apresentados a partir de categorias pré-definidas. Verificou-se que há 65 infraestruturas de dados que têm o termo COVID-19 indexado e a maioria delas tem <i>upload</i> restrito, acesso aberto aos dados e bancos de dados, algumas vezes associados a cadastros e questionários, mas sem licenças que favoreçam o livre reuso dos dados. Considera-se que há avanços consideráveis na criação de infraestruturas de dados quando comparado a outras ESPII, porém há necessidade de discussão qualificada para responder às emergências.
--	--

COVID-19 data infrastructures: conditions for deposit, access and uses

Data infrastructure COVID-19 Open Access Public Health Emergencies Data Sharing	In view of Covid-19's public health emergency of international concern and the importance of data sharing and open data for research in this context, the work aims to identify data infrastructure that declared the term COVID-19, analyzing aspects and conditions of deposit, access and use of this data, including terms and/or licenses for its use. Searches for data infrastructures were carried out in the re3data.org directory and the results were presented from pre-defined categories. There are 65 data infrastructures that have the term Covid-19 indexed and most of them have restricted upload, open access to data and databases, sometimes associated with registers and questionnaires, but without licenses that favor free reuse of the data. It is considered that there are considerable advances in the creation of data infrastructures when compared to other ESPIIs, but there is a need for qualified discussion to respond to emergencies.
---	--

INTRODUÇÃO

Em 30 de janeiro de 2020, a Organização Mundial da Saúde (OMS) declarou Emergência em Saúde Pública de Importância Internacional (ESPII) nos termos do Regulamento Sanitário Internacional, aprovado em 2005 (Brasil, 2009), pela circulação do vírus SARS-COV-2, e que popularmente ficou conhecido por novo coronavírus. Em 11 de março de 2020 a OMS declarou que a ESPII tem características de uma pandemia. A doença causada pelo SARS-COV-2 chama-se COVID-19 e tem números de

contaminação alarmantes e crescentes, assim como o número de óbitos notificados.

Neste cenário, houve uma corrida científica global para conhecer o vírus, sua história, suas características, seus impactos, cenário epidemiológico, medidas de controle, cuidado e tratamento, grupos de risco e possíveis medicamentos e vacinas para a doença causada pelo vírus. As expectativas e apelos sociais em prol de respostas científicas à pandemia aumentam e os atores do campo científico,

principalmente da saúde, mobilizam-se na vigilância e na realização de inúmeras pesquisas, a partir de grandes quantidades de dados e informações que estão sendo coletadas diariamente.

A velocidade da geração, circulação e publicação de informações científicas aumenta nestas situações de emergências e pode potencializar práticas e movimentos de abertura da ciência, fortalecendo os argumentos sobre a importância do acesso à informações e dados para pesquisa científica. Considera-se que um dos elementos constituintes de um ecossistema de compartilhamento e abertura de dados para pesquisas, principalmente no contexto da ESPII do novo coronavírus, são as infraestruturas para depósito e publicação dos dados. Essas infraestruturas, sejam elas repositórios, plataformas ou outros tipos, permitem que o pesquisador tenha um local para disponibilizar, acessar e coletar dados, possibilitando o uso de dados já publicados, de acordo com as autorizações pré-estabelecidas.

O objetivo deste trabalho é identificar infraestruturas de dados sobre a temática da COVID-19, analisando aspectos e condições de depósito, acesso e uso desses dados, incluindo termos e/ou licenças para sua utilização. Para tanto, realizou-se uma busca no diretório internacional RE3DATA.org, que é um registro global de infraestruturas de dados de pesquisa que permitem o armazenamento permanente e o acesso a conjuntos de dados para pesquisadores, órgãos de financiamento, editores e instituições acadêmicas. O re3data.org foi lançado em 2012, com financiamento da Fundação Alemã de Pesquisa (DFG) (RE3DATA.ORG, [s.d.]).

O cadastro de infraestruturas de dados no re3data.org envolve o atendimento a um conjunto de requisitos, registrados por meio de preenchimento de formulário, cuja análise é feita pelos administradores do diretório. Há uma preocupação especial com o campo acesso do formulário, que é dividido em três categorias, sendo elas: acesso aos dados, acesso ao banco de dados e acesso ao *upload* de dados, cada qual com três possíveis níveis de acesso – aberto,

restrito e/ou fechado. Aberto significa que não há barreiras de acesso. Restrito significa que usuários externos precisam superar barreiras de acesso, como cadastro com aprovação prévia do perfil do solicitante, resposta a questionários disponibilizados pelos administradores e outras. Fechado significa que os usuários externos não podem superar as barreiras de acesso (RE3DATA.ORG, [s.d.]).

No trabalho considerou-se então que o acesso abrange não somente a permissão aos usuários para obter ou visualizar dados, mas também a permissão, aos autores dos dados, para admissão e depósito na infraestrutura. Esta é uma questão importante porque o acesso às infraestruturas pode limitar a circulação dos dados em uma ESPII. Isso envolve também os termos e/ou licenças adotadas sobre os dados e bancos de dados que indicam e autorizam os possíveis tipos de usos.

Na pesquisa no re3data.org, utilizou-se como termo de busca a palavra «COVID-19» no re3data.org, recuperando os resultados obtidos a partir de sete filtros: 1) acesso aos dados, 2) restrições de acesso a dados, 3) acesso ao banco de dados, 4) restrições de acesso ao banco de dados, 5) licenças de banco de dados, 6) licenças de dados, 7) *upload* de dados e 8) restrições de *upload* de dados. Os resultados foram apresentados na Tabela 1.

Como principais resultados do trabalho foram localizadas 65 infraestruturas de dados que têm o termo Covid-19 indexado pela busca no re3data.org. Trata-se de plataformas e repositórios de dados de diversas regiões do mundo, principalmente Estados Unidos e Europa, que abrigam diferentes tipos de conteúdos, como dados audiovisuais, bancos de dados, imagens, gráficos estruturados, dados brutos, documentos e dados científicos e estatísticos. Comparando com os resultados do número de infraestruturas de dados para duas últimas ESPII declaradas pela OMS no mesmo diretório, onde há 1 resultado para busca pelo termo «zika» e 3 para o termo «ébola», verifica-se que há um considerável aumento do número de infraestruturas de dados para a atual ESPII, o

que representaria um avanço visível para o movimento dos dados abertos. Nesse levantamento, avaliaram-se os resultados a partir das condições para acesso às

infraestruturas e uso de dados sobre a COVID-19, conforme Tabela 1.

Tabela 1: Condições para acesso às infraestruturas e uso de dados sobre a COVID-19

CATEGORIA DE ANÁLISE	FILTRO DO DIRETÓRIO	PRINCIPAIS RESULTADOS	PERCENTUAL
Admissão dos dados e banco de dados	Upload de dados	Fechado	23%
		Aberto	2%
		Restrito	75%
	Restrições de <i>upload</i> de dados	Taxa requerida	7%
		Filiação institucional	4%
		Outro	35%
		Cadastro	55%
	Acesso aos dados e banco de dados	Acesso aos bancos de dados	Aberto
Restrito			3%
Restrição de acesso aos bancos de dados		Outro	33%
		Cadastro	67%
Acesso aos dados		Fechado	5%
		Embargado	9%
		Aberto	61%
		Restrito	25%
Restrição de acesso aos dados		Taxa requerida	17%
		Filiação institucional	3%
		Outro	43%
		Cadastro	37%
Condições para uso dos dados e banco de dados	Licença dos bancos de dados	Apache License 2.0	14%
		BSD	5%
		CC	19%
		CC0	10%
		<i>Copyrights</i>	24%
		ODC	5%
		Outro	24%
		Licença dos dados	Apache License 2.0
	BSD		1%
	CC		28%
	CC0		3%
	<i>Copyrights</i>		29%
	ODC		3%
	OGI		1%
	Public Domain		7%
	Outro	26%	

Por último, realizou-se uma avaliação sobre as condições estabelecidas para uso dos dados e banco de dados sobre a Covid-19, nas infraestruturas analisadas no trabalho. Como uma das etapas do processo de depósito, no ato da admissão de dados nas infraestruturas, o autor deve explicitar as condições de uso do material que está disponibilizando, seja de acesso aberto, restrito ou fechado, mas também sobre que tipos de uso são permitidos, como citação obrigatória, usos comerciais ou não, somente para fins de pesquisa científica entre outros. Desta forma, verificou-se que, em relação aos bancos de dados, as condições mais utilizados são o *copyright* (24%) e termos e condições de uso específicos, na tabela 1 discriminados como outro (24%). O *copyright* é restritivo e, quando aplicado em dados e bancos de dados, o uso somente poderá ocorrer a partir dos termos de uso estabelecidos pelos autores. Percebe-se que as licenças de direitos autorais são utilizadas para promover, encorajar e deixar claras as condições de compartilhamento e de reuso. Apenas 19% dos autores dos bancos de dados analisados adotaram alguma dentre as diferentes licenças da Creative Commons (CC), que são consideradas licenças livres e fomentam a cultura da abertura e liberdade do uso. Desta forma, sem licenças livres, o reuso dos dados fica comprometido nos termos indicados pelas infraestruturas, pois o limitam e condicionam a uma autorização para uso. Além da família de licenças da CC, verificou-se que outras licenças foram utilizadas pelas

infraestruturas para o depósito e o acesso aos dados da Covid-19, ainda que, em menor frequência de uso, tais como as licenças Apache License 2.0, BSD, ODC, OGL e Public Domain.

Ao abordar a temática do acesso a dados e respectivas infraestruturas, verifica-se a necessidade de permissões claras sobre as condições de depósito para compartilhamento, acesso e uso de dados. Estas informações têm impacto direto nos processos de produção do conhecimento, pois podem restringir o compartilhamento, o acesso e o reuso dos dados. Apesar do avanço no número de infraestruturas de dados em COVID-19, percebeu-se que o acesso para *upload* de dados nas plataformas avaliadas ainda não é tão aberto para muitos pesquisadores, sendo dificultado o depósito em infraestruturas temáticas. Também há condições restritivas de acesso em muitas infraestruturas, assim como condições restritivas de uso, não estando os dados e nem os bancos de dados tão livremente disponíveis para o público.

Conclui-se que há a necessidade de avançar nas discussões e reflexões sobre infraestruturas para compartilhamento e abertura de dados em ESPII, sobretudo para qualificar as questões e condições de acesso e usos, inclusive envolvendo questões de acessibilidade, e seus impactos na produção e circulação dos dados para pesquisa, para que possam beneficiar não somente as comunidades científicas, mas também outros setores da sociedade.

REFERÊNCIAS

BRASIL. AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA (2009). *Regulamento Sanitário Internacional*. Brasília: Ministério da Saúde, p. 11.

PORTAL BRASILEIRO DE DADOS ABERTOS. *O que são dados abertos?* [Em linha]. [Consult. 8 ago. 2018]

Disponível na Internet: <URL:
<http://dados.gov.br/pagina/dados-abertos>>.

RE3DATA.ORG. *FAQ - re3data.org*. [Em linha] [Consult. 10 abr. 2021] Disponível na Internet: <URL:
<https://www.re3data.org/faq>>.

Jorge, V. A.; Albagli, S. (2021). «Infraestruturas de dados da COVID-19: condições para depósito, acesso e usos». *Cadernos BAD*, n. 1-2. <https://doi.org/10.48798/cadernosbad.2753>

Acesso e Licença

Artigo em acesso aberto distribuído nos termos da licença Creative Commons Atribuição 4.0 Internacional (CC-by 4.0).

Revisão por Pares

Esta revista usa um sistema de revisão por pares, duplamente cega, assegurada pelo Conselho Científico da *Cadernos BAD*.

Conflitos de Interesse

Os autores declaram a inexistência de conflitos de interesse na realização do presente trabalho.

Confidencialidade dos Dados

Os autores declaram ter seguido os protocolos do RGPD.

Financiamento, Apoio e Patrocínios

Não existiram fontes externas de financiamento para a realização deste trabalho.

Recebido

07/10/2021

Aceite

13/10/2021

Publicado

26/09/2022
