

# Dados governamentais abertos sobre a COVID-19 no Brasil: avaliação do potencial de reuso

**Patrícia Nascimento Silva**

Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil;  
patricians@ufmg.br; ORCID <https://orcid.org/0000-0002-2405-8536>

**Resumo:** Na pandemia de COVID-19 o acesso aos dados sobre a doença tornou-se estratégico para controle das medidas de saúde pública. Diante da emergência em saúde, um grande volume de dados precisou ser minimamente organizado e disponibilizado de maneira rápida e automatizada compondo os dados governamentais abertos. Após dois anos de pandemia e com o intuito de apresentar um panorama sobre a publicação de dados abertos do governo federal do Brasil, sobre a COVID-19, este estudo buscou avaliar os dados governamentais abertos disponibilizados por meio de *Application Programming Interface*. A metodologia envolveu a identificação dos conjuntos de dados sobre a COVID-19 no Brasil, em *Application Programming Interface*, até abril de 2022, a análise da documentação e a avaliação utilizando a métrica DGABr. A avaliação considerou as cinco perspectivas da métrica que mensura elementos fundamentais sobre os dados governamentais abertos, essenciais para interoperabilidade e consequentemente reuso dos dados e pautou-se nas documentações publicadas. Como resultados, os dados governamentais abertos sobre a COVID-19, disponibilizados em *Application Programming Interface*, apresentaram uma boa pontuação na métrica, atingindo o nível 4. Esse resultado indica que o uso de APIs foi um recurso tecnológico importante e ágil para organização e disponibilização dos dados governamentais abertos, fomentando seu reuso. Contudo, é importante destacar que essa disponibilização para sociedade foi tardia, precisa de melhorias constantes, principalmente em questões técnicas como a ligação dos dados com outras fontes, e que as ações efetivas de reuso se limitaram a painéis de visualização de dados sobre a COVID-19.

**Palavras-chave:** dados governamentais abertos; dados abertos; COVID-19; API; reuso

## 1 Introdução

Após dez anos de vigência da Lei de Acesso à Informação (LAI), Lei Nº 12.527/2011, a política de dados governamentais abertos (DGA) do Brasil passou por diversas evoluções, sendo seu principal marco a formalização da

política de dados abertos, por meio do Decreto nº 8.777/2016. Com isso, a disponibilização de dados abertos do governo, inicialmente para o cumprimento de uma exigência da legislação, avançou e foi remodelada em arquivos, portais, painéis, bases de dados e sistemas.

A partir do Decreto nº 10.046/2019 (BRASIL, 2019) que dispôs sobre a governança no compartilhamento de dados no âmbito da Administração Pública Federal (APF) e instituiu o Cadastro Base do Cidadão e o Comitê Central de Governança de Dados é que alguns princípios dos dados governamentais abertos foram realmente incitados. O compartilhamento de dados no âmbito da APF estimula que dados abertos e privados possam ser compartilhados aumentando a ligação entre os dados, evitando redundâncias e promovendo a interoperabilidade.

No contexto dos DGA, o acesso aos dados deve permitir o processamento de forma automatizada. As *Application Programming Interface* (API)s permitem o processamento automatizado em meio a grandes volumes de dados e sua utilização na disponibilização dos dados abertos é crescente. Este tipo de recurso tecnológico vem sendo utilizado em vários países e a prática também foi adotada por alguns órgãos da APF do Brasil.

No Brasil, a Portaria Nº 188 de 3 de fevereiro declarou o estado de emergência em saúde pública de importância nacional (ESPIN) em decorrência da COVID-19 no Brasil. Em 11 de março a Organização Mundial de Saúde (OMS) elevou ao estado de pandemia a infecção causada pelo novo corona vírus. A partir deste momento, houve, em nível global, tensão pela abertura dos dados gerados pelos governos, não só com o objetivo de dar transparência às ações de combate à pandemia, mas também para possibilitar que setores da sociedade civil, a academia e demais entidades públicas e privadas, pudessem dar a sua contribuição (MINISTÉRIO DA SAÚDE, 2020).

Diante deste cenário, o Ministério da Saúde, por meio do DATASUS, disponibilizou a partir de outubro de 2020 dois conjuntos de dados em APIs, o primeiro conjunto relativo ao monitoramento dos casos da doença e o segundo, a partir de 2021, referente a vacinação. Frente à importância destes DGA para toda a sociedade e para o próprio governo, esta pesquisa é parte de um projeto

maior intitulado: Observatório de Dados Governamentais Abertos no Brasil: Organização e Reúso e tem o objetivo de investigar o potencial de reúso dos DGA sobre a COVID-19, disponibilizados por meio de APIs, pelo governo brasileiro, e procura responder a seguinte questão de pesquisa: Qual o potencial de reúso dos DGA disponibilizados sobre a COVID-19 no Brasil?

A pesquisa justifica-se para mapear as iniciativas existentes, após dois anos de pandemia, e publicizar novas formas de acesso aos DGA, assim como fomentar novos produtos e serviços de informação em tempos de emergência e desinformação e também para análises futuras no pós-pandemia.

## **2 Política de dados governamentais abertos no Brasil**

A política de dados governamentais abertos do governo brasileiro emerge como um subdomínio da política de informação do Brasil e engloba elementos específicos sobre os dados abertos no contexto da abertura e disponibilização. Tendo em consideração a política de informação como política de governo para documentos governamentais, o regime de acesso à informação pública no Brasil tornou-se mais conhecido a partir de 2011, com a LAI, apesar da Constituição da República já prever nos incisos XIV e XXXIII do artigo 5º, no inciso II do §3 do Artigo 37 e no §2 do Artigo 216, o acesso à informação pública (BRASIL, 1988; SILVA, 2018b).

Em 2016, o Decreto nº 8.777, de 11 de maio de 2016, instituiu a Política de Dados Abertos do Poder Executivo Federal, estabelecendo normas para a livre utilização das bases de dados e assegurando a possibilidade de uso e reúso por parte da sociedade. Este decreto também definiu um prazo para que todos os órgãos do Poder Executivo Federal publicassem seu Plano de Dados Abertos (PDA), sendo esse prazo de 60 dias a partir da publicação do Decreto no Diário Oficial da União (BRASIL, 2016a; BRASIL, 2016b; ESCOLA NACIONAL DE ADMINISTRAÇÃO PÚBLICA, 2017).

Entre 2011 e 2016, vários documentos e ações impulsionaram a política de dados governamentais abertos no país, dentre elas está a criação da Infraestrutura Nacional de Dados Abertos (INDA), em 2012. A normativa é um

conjunto de padrões, tecnologias, procedimentos e mecanismos de controle necessários para atender às condições de disseminação e compartilhamento de dados e informações públicas no modelo de Dados Abertos, em conformidade com os Padrões de Interoperabilidade de Governo Eletrônico (ePING) (INFRAESTRUTURA NACIONAL DE DADOS ABERTOS, 2022).

Em 2016, o Decreto nº 8.638/2016 que instituiu a política de governança digital no Brasil trouxe um novo olhar sobre o acesso aos dados e informações do governo, seguido do Decreto nº 9.319/2018 (BRASIL, 2018) que institui o sistema nacional para a transformação digital, estabelecendo a estrutura de governança para implantação da Estratégia Brasileira para a Transformação Digital. É a partir daí que surgem várias iniciativas da chamada transformação Digital que conforme Brasil (2020, online) tem o objetivo de “[...] tornar mais simples e ágil a relação entre o governo e o cidadão, reduzindo a burocracia e o tempo de resposta do Estado para a sociedade e aumentando a qualidade dos serviços públicos”. Muitos serviços que eram acessados somente presencialmente foram disponibilizados em sites web e aplicativos, aumentando as formas de acesso. Em 2019, o Decreto nº 10.046, estabeleceu o compartilhamento de dados no âmbito da APF. Essa iniciativa pode ser considerada um avanço na política de dados governamentais abertos, pois além de fomentar o reúso, estabeleceu padrões para a ligação dos dados (*linked data*) entre os órgãos da APF. Como consequência, os dados abertos resultantes dessas bases integradoras poderão ser utilizados por outros segmentos da sociedade, gerando novos produtos e serviços de informação. Ao longo dos anos, evoluções tecnológicas na publicação dos DGA do Brasil também foram identificadas, principalmente em relação à forma de disponibilização dos conjuntos de dados. A utilização de APIs na disponibilização dos dados abertos também é crescente nos órgãos da APF e permite o processamento automatizado (SILVA, 2022).

Uma API, ou uma interface de programação de aplicação, são padrões de programação para acesso a aplicações na web. As APIs podem ser consideradas uma forma de comunicação com o software que procura abstrair totalmente os detalhes da implementação subjacente, ou pelo menos o que for possível. O propósito de uma API é tornar o uso do sistema fácil e conveniente para que

desenvolvedores não familiarizados com ele possam criar código por meio da API rapidamente. Ao disponibilizar uma API para o público expande-se o público-alvo do sistema ou aplicação e esta passa a contemplar também desenvolvedores que queiram integrar suas funcionalidades em seus próprios sistemas. Uma boa API deve ser o mais auto-explicativa possível, ter uma boa documentação e ser retrocompatível. Esta última característica pode ser implementada através de versionamento, evitando que regras já estabelecidas sejam alteradas à medida que a API evolui (SAUDATE, 2021).

Em 2020, fenômenos como *big data* (grande volume de dados) já eram visualizados em diversas áreas e contextos digitais, mas com a pandemia de COVID-19, a transformação digital foi acelerada, e no contexto dos DGA não seria diferente. Além disso, a emergência em saúde fez com que a sociedade repensasse a importância e urgência da publicação de DGA sobre a COVID-19, em meio a painéis fora do ar, ataques cibernéticos e eleições. Em outubro de 2020 os dados do sistema e-SUS Notifica foram disponibilizados no *Portal Open DataSUS* por meio de uma API (OPEN KNOWLEDGE BRASIL, 2021). Estes eram os primeiros dados sobre a COVID-19, em APIs, no Brasil.

Em 2022, após dez anos da LAI, seis anos da instituição da Política de Dados Abertos e dois anos de pandemia de COVID-19, espera-se que além do acesso e processamento por máquina, a interoperabilidade encontre-se cada vez mais presente nos DGA, para que o seu maior objetivo, o reúso, seja plausível.

### **3 Procedimentos metodológicos**

Este estudo caracteriza-se como exploratório com abordagem qualitativa e quantitativa. A pesquisa é caracterizada como pesquisa ação e possui três etapas: (i) seleção dos conjuntos de dados sobre COVID-19 disponibilizados em APIs, (ii) análise da documentação das APIs e (iii) avaliação dos conjuntos de dados com a métrica DGABr, conforme detalhamento nas seções 3.1 a 3.3.

### 3.1 Seleção dos conjuntos de dados sobre COVID-19 disponibilizados em APIs

A seleção da amostra inicialmente foi feita através do Portal Brasileiro de Dados Abertos (PBDA) (<https://dados.gov.br/>), instrumento da política de DGA do Brasil, onde todos os conjuntos de DGA devem estar catalogados. Em seguida, também foi realizada uma busca ativa no *Portal Open DataSUS* (<https://opendatasus.saude.gov.br/>), que é um portal do governo que disponibiliza informações para subsidiar análises objetivas da situação sanitária, tomadas de decisão baseadas em evidências e elaboração de programas de ações de saúde (OPEN DataSUS, 2022).

Os filtros existentes no PBDA foram analisados, assim como as informações existentes sobre a publicação de conjuntos de dados sobre a COVID-19. Foi identificado somente o filtro “Formato” com o tipo “API” e um campo de pesquisa textual. As buscas no PBDA foram realizadas em abril de 2022 e retornaram três órgãos que disponibilizam DGA em APIs: Banco Central do Brasil (528 ocorrências), Ministério da Economia (oito ocorrências) e Ministério da Saúde (três ocorrências). Ao realizar uma pesquisa com o termo “covid”, nas ocorrências retornadas para o formato API, foram identificados somente dois conjuntos de dados do Ministério da Saúde: “(1) Notificações de Síndrome Gripal” (NSG) e “(2) Campanha Nacional de Vacinação contra Covid-19” (CNVC). No *Portal Open DataSUS* também foi utilizado o mesmo critério de busca e foram encontrados os mesmos conjuntos de dados identificados no PBDA.

### 3.2 Análise da documentação das APIs

A análise da documentação envolveu uma busca ativa em todos os documentos disponibilizados no PBDA e no *Open DataSUS* para os conjuntos de dados selecionados, a fim de identificar: a documentação da API, métodos de acesso, dicionário de dados e padrões utilizados na organização e publicação dos dados e o PDA do Ministério da Saúde, órgão responsável pela publicação dos conjuntos de dados sobre a COVID-19 em APIs. O Quadro 1 apresenta os

respectivos conjuntos de dados e o endereço da documentação no PBDA e no *Open DataSUS*.

**Quadro 1** - Relação de documentos sobre os conjuntos de dados selecionados

| Conjunto de dados                                         | Endereço no Portal de Brasileiro de Dados Abertos                                                             | Endereço no Open DataSUS                                                                                                                                                                            |
|-----------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (1) Notificações de Síndrome Gripal (NSG)                 | <a href="https://dados.gov.br/dataset/covid-19-vacinacao">https://dados.gov.br/dataset/covid-19-vacinacao</a> | <a href="https://opendatasus.saude.gov.br/dataset/covid-19-vacinacao">https://opendatasus.saude.gov.br/dataset/covid-19-vacinacao</a>                                                               |
| (2) Campanha Nacional de Vacinação contra Covid-19 (CNVC) | <a href="https://dados.gov.br/dataset/casos-nacionais">https://dados.gov.br/dataset/casos-nacionais</a>       | <a href="https://opendatasus.saude.gov.br/dataset/notificacoes-de-sindrome-gripal-api-elasticsearch">https://opendatasus.saude.gov.br/dataset/notificacoes-de-sindrome-gripal-api-elasticsearch</a> |

Fonte: Elaborado pela autora.

O PDA publicado e disponibilizado no site do Ministério da Saúde (<https://portalsage.saude.gov.br/planoDadosAberto>) para o ano de 2020 e 2021 foi analisado. Apesar de não haver indicação sobre o ano de 2022 no PDA publicado no site do Ministério da Saúde, no Painel de Monitoramento de Dados Abertos, instrumento de monitoramento da política de dados governamentais abertos, a vigência indicada para o PDA é 22/10/20 a 22/10/22 (CONTROLADORIA-GERAL DA UNIÃO, 2022), conforme apresentado na Figura 1. Assim, analisando o conteúdo do PDA foi identificado que ambos os endereços apontam para o mesmo documento e que há a abertura de alguns conjuntos de dados em 2022. Assim, considerou-se que o documento disponível está dentro da vigência e que, provavelmente, o PDA foi atualizado para contemplar o ano de 2022. Contudo, essa informação não está clara para o usuário.



**Figura 1** - Monitoramento de Dados Abertos do Ministério da Saúde



Fonte: Controladoria Geral da União (2022)

Após realizar uma consulta nos endereços apresentados no Quadro 1, um resumo das documentações e recursos selecionados, para análise de cada conjunto de dados, é apresentado no Quadro 2:

**Quadro 2** - Documentos analisados para cada conjunto de dados<sup>1</sup>

| Conjunto de dados                     | (1) Notificações de Síndrome Gripal (NSG)                                                                                                                                                                                                                                                                                                                                                                                                                         | (2) Campanha Nacional de Vacinação contra Covid-19 (CNVC)                                                                                                                                                                                                                                                                 |
|---------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Documentação identificada nos Portais | <ul style="list-style-type: none"> <li>- Dicionário de dados</li> <li>- Ficha de Notificação COVID</li> <li>- Instrutivo de preenchimento da ficha de notificação de caso de Síndrome Gripal suspeito de doença pelo Coronavírus 2019 COVID-19 (B34.2)</li> <li>- Elasticsearch API – Notificações</li> </ul> <p>Documentos disponíveis em:<br/><a href="https://dados.gov.br/dataset/covid-19-vacinacao">https://dados.gov.br/dataset/covid-19-vacinacao</a></p> | <ul style="list-style-type: none"> <li>- Dicionário de dados</li> <li>- Manual para Consumo da API</li> <li>- Elasticsearch API - Registros de Vacinação COVID19</li> </ul> <p>Documentos disponíveis em:<br/><a href="https://dados.gov.br/dataset/casos-nacionais">https://dados.gov.br/dataset/casos-nacionais</a></p> |
| PDA                                   | <a href="https://portalsage.saude.gov.br/planoDadosAberto">https://portalsage.saude.gov.br/planoDadosAberto</a>                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                           |
| Endpoint da API                       | <a href="https://elasticsearch-saps.saude.gov.br/desc-esus-notificacao-estado-*/_search">https://elasticsearch-saps.saude.gov.br/desc-esus-notificacao-estado-*/_search</a>                                                                                                                                                                                                                                                                                       | <a href="https://imunizacao-es.saude.gov.br/desc-imunizacao/_search">https://imunizacao-es.saude.gov.br/desc-imunizacao/_search</a>                                                                                                                                                                                       |

Fonte: Elaborado pela autora.

As fontes documentais apresentadas no Quadro 2 correspondem ao corpus da pesquisa. Todos os documentos foram acessados e analisados juntamente com o acesso às APIs, por meio do *endpoint*, também indicado no Quadro 2. Destaca-se que o acesso à API foi utilizado somente para analisar se os dados



retornados estavam em conformidade com a documentação disponibilizada para o conjunto de dados. Os acessos à API foram realizados de forma automatizada e exclusivos para coleta de dados, por meio de código implementado na linguagem *Python*. A avaliação dos DGA foi realizada manualmente e detalhada na terceira etapa metodológica.

### 3.3 Avaliação dos conjuntos de dados com a Métrica DGABr

Para avaliar os conjuntos de dados foi selecionada a métrica DGABr. A DGABr é a primeira métrica brasileira criada para avaliar o potencial de reúso dos DGA disponibilizados no Brasil. A métrica permite avaliar conjuntos de dados individuais e é composta por dimensões qualitativas e quantitativas que avaliam elementos do ciclo de vida dos DGA para mensurar o potencial de reúso (SILVA, 2018b).

A avaliação considerou as informações sobre os conjuntos de dados no PDA, o acesso direto a cada uma das APIs, implementado na linguagem *Python* pela autora, e toda a documentação selecionada no PBDA e no *Open DataSUS*. Toda a avaliação foi feita manualmente, por meio do *template* da métrica, publicado no site: <https://www.dgabr.com.br/>, para avaliar cada uma das 28 dimensões divididas em cinco perspectivas de análise. Para cada dimensão foram avaliadas as características da publicação dos dados, conforme o respectivo peso, atribuindo um nível de 0 a 5, onde zero é o menor e cinco o maior valor, conforme critérios da métrica DGABr, detalhados no Quadro 3 e Quadro 4.

**Quadro 3 - Pesos da métrica DGABr**

| Pesos da métrica DGABr |                                                                                     |
|------------------------|-------------------------------------------------------------------------------------|
| Peso                   | Descrição                                                                           |
| 1                      | Peso de menor valor, corresponde a uma perspectiva básica e essencial.              |
| 2                      | Peso atribuído a uma perspectiva com elementos necessários aos DGA.                 |
| 3                      | Peso de maior valor, corresponde aos elementos com maior relevância para a métrica. |

Fonte: Adaptado de Silva (2018a).

**Quadro 4 - Níveis da métrica DGABr**

| <b>Níveis da métrica DGABr</b>            |                                                                                                                                                                    |
|-------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Nível</b>                              | <b>Descrição</b>                                                                                                                                                   |
| Nível 0 ( <b>Inexistente</b> )            | Inexistência ou não cumprimento da dimensão.                                                                                                                       |
| Nível 1 ( <b>Em construção</b> )          | Inexistência ou não cumprimento da dimensão, porém há um estudo ou uma padronização em construção.                                                                 |
| Nível 2 ( <b>Não executado</b> )          | Não cumprimento da dimensão, apesar de existir uma recomendação ou padronização regulamentada pelo órgão.                                                          |
| Nível 3 ( <b>Executado parcialmente</b> ) | Observou-se o cumprimento da dimensão, que atingiu 50%, ou seja, foi considerado que 50% representa um valor mediano em relação ao cumprimento da dimensão.        |
| Nível 4 ( <b>Resultados Existentes</b> )  | Observou-se o cumprimento da dimensão, que atingiu 80%, ou seja, foi considerado que 80% representa um valor acima da média em relação ao cumprimento da dimensão. |
| Nível 5 ( <b>Resultados Avançados</b> )   | A dimensão foi cumprida integralmente, ou seja, 100%.                                                                                                              |

Fonte: Adaptado de Silva (2018a).

A avaliação considerou os documentos e conjuntos de dados publicados até abril de 2022, que inclui dados até março de 2022, compreendendo assim dois anos da pandemia de COVID-19. Destaca-se que os resultados encontrados representam um espelho do momento em que a pesquisa foi realizada e estão fundamentados na documentação fornecida pelo órgão. Os acessos às APIs não foram realizados de forma exaustiva e, neste estudo, foram utilizados como apoio para interpretação dos documentos analisados.

## 4 Resultados e discussões

A avaliação dos conjuntos de dados (1) NSG e (2) CNVC considerou todas as perspectivas e dimensões da métrica DGABr, com o intuito de avaliar o potencial reúso dos DGA de maneira completa. Nas seções 4.1. a 4.5 são detalhadas as perspectivas, respectivas dimensões de análise e a pontuação alcançada em cada dimensão.

### 4.1 Perspectiva Dados Abertos

A perspectiva Dados Abertos compreende as dimensões D1 a D10 e avaliou princípios e recomendações técnicas sobre a publicação de dados abertos. A avaliação das dimensões D1 a D10 são detalhadas a seguir:

- a) D1 - Dados completos: na documentação é informado que a cobertura

temporal para os dados de (1) NSG é a partir de fevereiro de 2020 e que foram utilizados dados de sistemas legados, porém na API não é possível acessar os dados por período, somente por Estado. Ao analisar as datas de notificação é possível identificar casos desde janeiro de 2020, mas há notificações sem data e com valores incorretos, como exemplo uma notificação em 26/05/1993. Assim, foi alcançado o nível 4 que considera que pelo menos 80% dos DGA estão completos. Para o conjunto de dados (2) CNVC são identificados registros a partir de 2021, início da campanha de vacinação no Brasil, porém também não há um parâmetro na API para recuperar os dados por período, mas como grande parte dos dados são apresentados na consulta geral, incluindo cada uma das doses da vacina, também foi alcançando o nível 4;

- b) D2 - Dados Primários: na documentação de ambos os conjuntos de dados são informados os dados coletados na fonte e as modificações realizadas são para não identificação pessoal, alcançando o nível 5;
- c) D3 - Dados Atualizados: a documentação indica que os dados são atualizados na frequência diária e conforme verificado na API a frequência é cumprida, alcançando o nível 5;
- d) D4 - Dados Acessíveis: Os dados são disponibilizados via API, via acesso web e sem nenhuma forma de identificação, alcançando o nível 5;
- e) D5 - Dados Processáveis por máquina: na API todos os conjuntos de dados estão estruturados, são disponibilizados em formatos abertos e processáveis por máquina, alcançando o nível 5;
- f) D6 - Acesso Não Discriminatório: os dados são disponibilizados para qualquer pessoa mediante algum tipo de identificação. Porém, a identificação corresponde a um login e uma senha informados no PBDA e no manual da API, ou seja, todas as pessoas que utilizam a API possuem o mesmo login e senha, assim considerou-se que não há nenhum tipo de exigência de identificação, alcançando o Nível 5;
- g) D7 - Formato de Dados não Proprietário: os dados são publicados em formato não proprietário na API (JSON), alcançando o nível 5;
- h) D8 - Dados Livres de Licença: Os DGA publicados são livres de licença,

restrições ou privilégios de acesso, alcançando o nível 5;

- i) D9 - URIs desenhadas: os DGA publicados utilizam URIs que os identificam na API, alcançando o nível 5;
- j) D10 - Dados Ligados: os DGA publicados não são ligados com outras fontes, porém há recomendações e padrões de publicação do órgão que incentivam e permitem essa ligação, alcançando o nível 2. São utilizados códigos do IBGE para identificar os municípios e códigos do Cadastro Nacional de Estabelecimentos de Saúde.

Na perspectiva de Dados Abertos, a completude dos DGA é um dos elementos mais críticos em relação à publicação dos DGA sobre a COVID-19. No início da pandemia, houve uma grande resistência por parte do governo em divulgar dados sobre a COVID-19 e este cenário gerou incertezas em relação aos dados publicados. Com isso, foi criado em 8 de junho de 2020 o Consórcio de Veículos de Imprensa (CVI) que é uma parceria estabelecida entre os veículos de imprensa brasileiros: *O Estado de São Paulo*, *G1*, *O Globo*, *Extra*, *Folha de São Paulo* e *UOL* para informar dados da pandemia de COVID-19 no Brasil, recebidos das secretarias estaduais de saúde. Sua fundação ocorreu devido à restrição que o Ministério da Saúde promoveu na época à divulgação dos dados de números de casos e óbitos decorrentes de COVID-19 (G1 *et al.*, 2020). Posto este cenário, e avaliando somente a documentação dos dados publicados, é importante ressaltar que a pontuação obtida nas dimensões avaliadas D1 a D10 é alta. Assim, abstraindo a questão da insegurança dos dados e também possíveis subnotificações, os dados publicados permitem o seu reúso de forma rápida e facilitada por meio da API.

Com relação à organização da informação, a versão disponibilizada da API aceita somente parâmetros (filtros) por Estado. Desta forma, as informações retornadas são volumosas e exigem que o usuário organize a informação em menor granularidade<sup>2</sup> para análise. Para Ferneda (2012) o processo de recuperação de informação compreende identificar em um acervo documental os documentos que venham a atender a uma determinada necessidade de informação do usuário. Nesta mesma reflexão Baeza-Yates e Ribeiro-Neto (2013) destacam que para ser efetivo, um sistema de recuperação da informação

precisa satisfazer a necessidade de informação de seu usuário e que usuários de sistemas modernos têm necessidades de informação de diferentes níveis de complexidade. A disponibilização de dados e informações em API é um recurso contemporâneo que precisa ser projetado para atender as necessidades dos usuários e permitir buscas avançadas em consonância com a evolução tecnológica e conceitos da web semântica.

A inexistência de ligação com outros conjuntos de dados é algo que, apesar de iniciado de maneira muito tímida, não foi implementado. Contudo, visto que a API foi construída em um momento de emergência, essa questão poderá ser tratada como uma melhoria, a partir da maturidade alcançada com a manipulação destes dados. A ligação dos dados abertos ou *Linked Open Data* (LOD) é de extrema importância e vai de encontro com as iniciativas do Governo Federal, através do Decreto nº 10.046/2019 que dispôs sobre a governança no compartilhamento de dados no âmbito da APF. A associação dos conjuntos de dados sobre a COVID-19 com outros conjuntos de dados poderão propiciar uma maior interoperabilidade e consequentemente uma maior rastreabilidade das informações, subsidiando análises e estudos sobre pandemia, em diversos contextos e diferentes épocas, já que estes dados possuem uma rica relevância histórica para o país e quiçá para a humanidade. Conforme apresentado por Macedo e Lemos (2021) a prática de LOD no âmbito governamental atinge uma nova dimensão, em especial pela grande capacidade dos governos em centralizar e coletar uma ampla gama de dados e informações sensíveis a toda uma sociedade.

## 4.2 Perspectiva Legal

A Perspectiva Legal compreende as dimensões D11 a D13 e avalia os elementos específicos da legislação brasileira, como o tipo de informação a ser publicada, a classificação da informação e o atendimento ao (PDA). A avaliação das dimensões D11 a D13 são detalhadas a seguir:

- a) D11 - Tipo de Informação: os dados publicados estão previstos na legislação, no PDA e são DGA, alcançando o nível 5. Para os dados da

vacinação ainda são consideradas as seguintes legislações:

Medida Provisória nº 1.026, de 6 de janeiro de 2021, dispõe sobre as medidas excepcionais relativas à aquisição de vacinas, insumos, bens e serviços de logística, tecnologia da informação e comunicação, comunicação social e publicitária e treinamentos destinados à vacinação contra a covid-19 e sobre o Plano Nacional de Operacionalização da Vacinação contra a Covid-19.

Portaria GM/MS nº 69, de 14 de janeiro de 2021 – institui a obrigatoriedade de registro de aplicação de vacinas contra Covid-19 nos sistemas de informação do Ministério da Saúde.

Nota Informativa nº 1/2021 – CGPNI/DEIDT/SVS/MS, de 10 de janeiro de 2021, dispõe sobre as orientações para o registro de vacinas no sistema de informação e sobre acesso às informações referentes à vacinação contra Covid-19 (PORTAL BRASILEIRO DE DADOS ABERTOS, 2022a, online).

- b) D12 - Gestão de dados sensíveis: os DGA publicados têm os dados dessensibilizados, conforme as normas e padrões definidos pelo órgão e pela Lei Geral de Proteção de Dados (LGPD) que entrou em vigor em 2020. Quanto aos prazos de sigilo, não há como avaliar ainda, mas por se tratar de dados pessoais não deverão ser divulgados. Desta forma, considerando que o sigilo será mantido, foi atribuído o nível 5;
- c) D13 - Plano de Dados Abertos: Em abril de 2022, somente um conjunto de dados do PDA do Ministério da saúde não tinha sido disponibilizado, desta forma, pelo menos 80% dos compromissos e metas do PDA foram cumpridos pelo órgão, alcançando o nível 4.

Na perspectiva legal os dados estão em conformidade com a legislação, principalmente com o PDA que foi elaborado no ano de 2020 e incluiu a abertura dos dados sobre a COVID-19 e outras atualizações até 2022. Os dados pessoais foram retirados dos conjuntos de dados, conforme já previsto na publicação dos DGA e ratificado pela LGPD. Legislações específicas com relação à vacinação também foram essenciais para a gestão e o controle dos estoques de vacina e também para evitar fraudes, que não foram descartadas, mas podem ter sido minimizadas. A disponibilização dos dados de vacinação permitiu que a sociedade acompanhasse de alguma maneira a evolução da campanha, prioridades e o quantitativo de vacinados.

### 4.3 Perspectiva Técnica

A Perspectiva Técnica avalia a utilização de padrões técnicos existentes no Brasil para a publicação de DGA. A avaliação das dimensões D14 a D20 são detalhadas a seguir:

- a) D14 - Fonte da informação: as fontes dos DGA são disponibilizadas no PDA e nos documentos da API, alcançando o nível 5. Conforme indicado na documentação, grande parte dos dados são provenientes de base de dados do sistema e-SUS;
- b) D15 - Formato dos dados: os DGA publicados utilizam os formatos abertos recomendados pelos padrões existentes nos serviços de API, alcançando nível 5;
- c) D16 - Qualidade: Com base na média dos níveis alcançados nas dez dimensões da perspectiva Dados Abertos, para ambos conjuntos foi obtido o nível 4 (média igual a 4,6);
- d) D17 - Metadados: o Ministério da Saúde disponibiliza os metadados mínimos sugeridos para os conjuntos de dados nos Portais, porém no PDA não há um detalhamento que considere toda a especificidade dos atributos disponibilizados sobre a COVID-19. Analisando as documentações dos Portais foram identificados dicionários de dados, onde há um detalhamento completo dos atributos do conjunto (1) NSG, que alcança o nível 5, o que não ocorre para o conjunto de dados (2) CNVC, cujo o dicionário de dados não detalha a maioria dos campos e categorias, atingindo assim o nível 3;
- e) D18 - Vocabulário: os DGA publicados não utilizam um vocabulário controlado, obtendo o nível 0. Não há nenhuma indicação sobre o uso de algum tipo de controle de vocabulário na documentação de ambos conjuntos de dados;
- f) D19 - Dimensão Geográfica: os DGA publicados estão conforme a cobertura geográfica definida no PDA, cuja granularidade é o município, alcançando assim o nível 5;
- g) D20 - Catalogação de dados: o Ministério da Saúde cataloga todos os DGA disponibilizados em APIs no PBDA, alcançando o nível 5.



A Perspectiva Técnica é a parte da métrica que avalia os maiores impactos para a organização da informação. O formato dos dados foram bem pontuados, uma vez que a API permite a visualização dos DGA em vários formatos, inclusive no formato JSON que é semi-estruturado e permite o reúso em diversas aplicações, facilitando o armazenamento em bancos de dados.

A organização da informação é um processo de arranjo de acervos tradicionais ou eletrônicos realizado por meio da descrição física e de conteúdo (assunto) de seus objetos informacionais (CAFÉ; SALES, 2010, p. 118). Os metadados são elementos de descrição utilizados na representação bibliográfica desde o surgimento dos primeiros instrumentos de controle bibliográfico, como as bibliografias e os catálogos (BRANDT *et al.*, 2019, p. 2) e atualmente continuam sendo utilizados em recursos digitais, como APIs, para descrever os dados. Os metadados foram detalhados no conjunto (1) NSG, porém no conjunto (2) CNVC há muitos atributos não detalhados, o que pode gerar dúvidas e interpretações incorretas pelos usuários no reúso dos DGA. Desta forma, a pontuação obtida considerou estas limitações.

A inexistência do uso de um vocabulário controlado pode impactar diretamente na recuperação da informação que utiliza linguagem natural e termos técnicos da área de saúde. Conforme destacado por Cruz, Ferneda e Fujita (2022) em uma pesquisa sobre a disponibilização de vocabulário controlado aos usuários para a recuperação da informação:

Na Ciência da Informação os vocabulários controlados geralmente são utilizados pelos indexadores com preocupação na padronização e consistência dos conceitos descritivos do conteúdo informacional dos documentos. Em um sistema de recuperação de informação, o vocabulário controlado pode também ser disponibilizado aos usuários a fim de auxiliá-los na elaboração de suas estratégias de busca (CRUZ; FERNEDA; FUJITA, 2022, p. 269).

Com isso, a definição de conceitos específicos da área e a utilização de um vocabulário controlado impactarão positivamente nas estratégias de recuperação da informação, atendendo as necessidades de busca dos usuários. Ações que estimulem o enriquecimento semântico dos termos também pode ser uma solução para melhoria da qualidade dos dados publicados que obtiveram a

pontuação zero na dimensão D18, caracterizando a inexistência do item avaliado.

A interoperabilidade semântica está ligada a capacidade de trocar e reusar conteúdos contextualizados semanticamente e reside na estruturação da informação (atribuição de metadados) por meio do controle de vocabulário, a partir de linguagens estruturadas que considerem relações sintáticas e semânticas (MARTINS, 2013). Desta forma, o tratamento semântico demanda da inclusão de instrumentos lógicos durante todo o ciclo de vida dos dados. Para Santarém Segundo (2015) a não utilização de esquemas lógicos como as ontologias e os principais vocabulários para descrever dados a serem publicados certamente é um caminho desacertado para publicação de dados em formato semântico, principalmente do ponto de vista da recuperação da informação.

#### **4.4 Perspectiva Gerencial**

A Perspectiva Gerencial avalia a gestão dos DGA pela organização e pode ser mensurada pela organização ou pelo cidadão, por meio das informações divulgadas pelo próprio órgão. A avaliação das dimensões D21 a D23 são detalhadas a seguir:

- a) D21 - Priorização e Estratégias de Abertura: conforme informado no site do Ministério da Saúde, após pressões da sociedade para divulgação sobre os dados da COVID-19, o órgão priorizou esta demanda e publicou alguns DGA. Desta forma, considerou-se que os DGA são publicados conforme o critério de priorização de abertura do órgão, alcançando o nível 5. Contudo, carências foram listadas para o conjunto de dados (1) NSG:

Os dados disponíveis não apresentam informações de estados e municípios que utilizam sistemas próprios de notificação de casos suspeitos de Covid-19 e, portanto, os dados para esses locais podem apresentar informações distintas. Estes dados são preliminares, sujeitos a avaliação. As bases estão sendo avaliadas e harmonizadas, com o objetivo de ser verificada sua consistência, principalmente em relação a atualização dos dados (PORTAL BRASILEIRO DE DADOS ABERTOS, 2022b, online).

- b) D22 - Estratégias de Investimento e cronograma: Inicialmente os dados sobre COVID-19 foram disponibilizados em arquivos, com formatos abertos. Assim, o desenvolvimento da API pode ser considerado uma melhoria na disponibilização dos dados. Porém, por ainda haver conjuntos de dados pendentes, dados que não estão completos ou totalmente corretos em ambos os conjuntos de dados, foi alcançado o nível 4, onde considera-se que pelo menos 80% dos DGA foram publicados conforme o cronograma de publicação, atualização e melhoria no PDA;
- c) D23 - Manutenção e correção de problemas: o Ministério da saúde informa ter ciência dos problemas do catálogo, conforme divulgado no PBDA e disponibiliza um e-mail para contato. Porém, as informações sobre manutenção não são divulgadas, há somente a indicação da última atualização dos conjuntos de dados no PBDA, sem detalhes das alterações realizadas. Desta forma, como ainda há pendências de informações, principalmente sobre os metadados para o conjunto (2) CNVC e dados incorretos em ambos os conjuntos, alcançou-se o nível 3, para ambos conjuntos de dados, pois não há informações claras sobre correções e manutenções.

Após diversas pressões da sociedade, os dados sobre COVID-19 foram incluídos no PDA do Ministério da Saúde. Embora o resultado sobre o potencial de reúso de dados seja alto, há de se destacar o atraso em relação aos outros países na disponibilização dos DGA pelo governo, o que consequentemente impossibilitou o uso dos dados no início da pandemia. As dimensões avaliadas na perspectiva gerencial apresentam resultados medianos e requerem melhorias.

#### **4.5 Perspectiva Reúso**

A Perspectiva Reúso avalia dimensões relacionadas ao reúso dos dados, mas que ainda não possuem uma legislação específica no país. A avaliação das dimensões D24 a D28 são detalhadas a seguir:

- a) D24 - Quantidade de conjuntos de dados Publicados: conforme indicado no Painel de Monitoramento de Dados Abertos, o órgão publica mais de 80% dos conjuntos de DGA propostos. Nesta dimensão considerou-se o PDA de maneira geral e não somente o recorte deste estudo, pois a avaliação contempla o órgão;
- b) Para as dimensões D25 - Quantidade de aplicativos desenvolvidos e D26- Quantidade de download de aplicativos, foram identificados no site do Ministério da Saúde dois aplicativos: o Conecte SUS e o Coronavírus - SUS;
- o Conecte SUS é um programa de governo que materializa a Estratégia de Saúde Digital para o Brasil, potencializando o compartilhamento de informações entre todos os pontos da Rede de Atenção em Saúde pública e privada (BRASIL, 2022a). No aplicativo é possível visualizar dados individualizados sobre a vacinação contra COVID-19, contudo o acesso é mediante login e senha do usuário. Desta forma, o aplicativo não foi contabilizado, por utilizar dados pessoais individualizados e não DGA;
  - o Coronavírus - SUS é o aplicativo oficial do governo para a COVID-19 e tem o objetivo de conscientizar a população sobre a COVID-19. No aplicativo são disponibilizados informativos sobre os sintomas e medidas de prevenção, informações sobre a unidade básica de saúde mais próxima, notificações sobre contato com pessoas que tiveram a doença e notícias oficiais do Ministério da Saúde. A última atualização foi realizada em abril de 2021 e há mais de 10 milhões de downloads (GOOGLE, 2021). No Aplicativo não é possível visualizar estatísticas ou informações sobre os casos, conforme disponibilizado no conjunto (1) NSG, assim o aplicativo também não foi contabilizado;
  - em uma busca ativa na loja de aplicativos do Google (<https://play.google.com/store/>) não foram identificados outros aplicativos do Ministério da Saúde. Com isso, para ambos os conjuntos de dados não foram identificados aplicativos oficiais e/ou

reconhecidos pelo Ministério da Saúde que reusam os DGA, obtendo o nível 0;

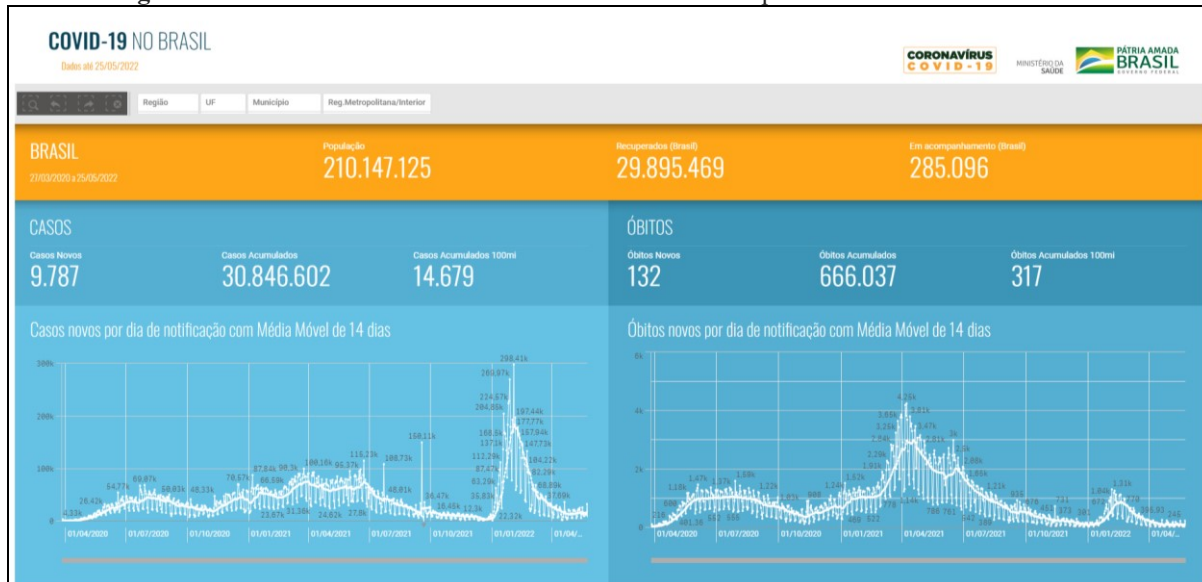
- c) D27 - Engajamento da sociedade civil: no PDA foi planejada a publicação de matérias com informações sobre os conjuntos de dados disponibilizados, eventos sobre dados abertos e um *Hackathon*, que corresponde a uma maratona de programação para explorar DGA. Em uma busca ativa no site do Ministério da Saúde, utilizando o recurso de pesquisa com a expressão “covid dados”, foram encontradas matérias sobre a divulgação dos dados e painéis sobre os dados da COVID-19. A página oficial é: <https://www.gov.br/saude/pt-br/coronavirus> e os painéis estão disponíveis em: <https://covid.saude.gov.br/>. Não foram identificadas notícias sobre o *Hackaton* até abril de 2022. Desta forma, considerou-se que pelo menos 80% das ações de engajamento planejadas foram executadas, alcançando o Nível 4;
- d) D28 - Informações sobre o reúso de DGA: o Ministério da Saúde criou painéis de visualização que consomem os dados publicados sobre a COVID-19. Não há informações sobre outros produtos e serviços disponibilizados pelo Ministério da Saúde. Com isso, o serviço desenvolvido pelo órgão tem informações sobre as fontes e atualização dos DGA utilizados, alcançando o nível 5.

Na perspectiva reúso os níveis obtidos não foram satisfatórios. O reúso ainda é pequeno, face relevância dos conjuntos de dados publicados sobre a COVID-19 no Brasil. Foram identificados painéis, construídos pelo Ministério da Saúde, para visualização dos dados sobre a COVID-19. Este tipo de implementação é rápida de ser concluída e seu acesso é facilitado por estar disponibilizado em páginas web, conforme apresentado na figura 2.

É importante destacar que para que ocorra reúso, a informação precisa ser recuperada. Com isso, além dos instrumentos lógicos utilizados nos novos sistemas, o comportamento informacional também é um elemento fundamental. Estudos sobre os comportamentos informacionais dos usuários, especialmente os de busca e recuperação da informação, os quais incluam as variáveis que interferem nestes comportamentos, podem contribuir para o desenvolvimento de

novos sistemas e ferramentas para a organização e representação da informação, assim como para o uso da informação (GARCIA, 2007).

**Figura 2** - Painel de Interativo sobre a COVID-19 no Brasil para o Estado de Minas Gerais



Fonte: Brasil (2022b).

O fato de não existirem aplicativos oficiais reutilizando os conjuntos de dados não significa que os dados não são utilizados em outros contextos. Porém esperava-se que as próprias ações de engajamento do Ministério da Saúde estimulassem o reúso dos DGA e o reconhecimento de outras aplicações pelo órgão, por meio do diálogo com os usuários e consumidores da API. Contudo, diante de tantos desafios e dificuldades institucionais para publicação dos dados sobre a COVID-19, as ações realizadas podem ser consideradas uma conquista da sociedade, em tempos de desinformação e negacionismo, o que não impede a cobrança por melhorias e evoluções.

Após realizar o somatório da pontuação, para cada conjunto de dados, foi identificado que o conjunto de dados (1) NSG obteve o indicador 4 (211 pontos) e o (2) CNVC também obteve o indicador 4 (205 pontos) conforme detalhado na Tabela 1.

Tabela 1 - Resumo da avaliação com a métrica DGABr

| Métrica DGABr    |     |                                             |      | (1) Notificações de Síndrome Gripal |                               | (2) Campanha Nacional de Vacinação contra Covid-19 |                               |
|------------------|-----|---------------------------------------------|------|-------------------------------------|-------------------------------|----------------------------------------------------|-------------------------------|
| Perspectiv<br>a  | ID  | Dimensão                                    | Peso | Nível<br>adquirido                  | Total<br>(peso<br>x<br>nível) | Nível<br>adquirido                                 | Total<br>(peso<br>x<br>nível) |
| Dados<br>abertos | D1  | Dados completos                             | 1    | 4                                   | 4                             | 4                                                  | 4                             |
|                  | D2  | Dados primários                             | 1    | 5                                   | 5                             | 5                                                  | 5                             |
|                  | D3  | Dados atualizados / DGA atualizado          | 1    | 5                                   | 5                             | 5                                                  | 5                             |
|                  | D4  | Dados acessíveis                            | 1    | 5                                   | 5                             | 5                                                  | 5                             |
|                  | D5  | Dados processáveis por máquina              | 1    | 5                                   | 5                             | 5                                                  | 5                             |
|                  | D6  | Acesso não discriminatório                  | 1    | 5                                   | 5                             | 5                                                  | 5                             |
|                  | D7  | Dados não proprietários                     | 1    | 5                                   | 5                             | 5                                                  | 5                             |
|                  | D8  | Dados livres de licença                     | 1    | 5                                   | 5                             | 5                                                  | 5                             |
|                  | D9  | URLs desenhadas                             | 1    | 5                                   | 5                             | 5                                                  | 5                             |
|                  | D10 | Dados ligados                               | 1    | 2                                   | 2                             | 2                                                  | 2                             |
| Legal            | D11 | Tipo de informação                          | 3    | 5                                   | 15                            | 5                                                  | 15                            |
|                  | D12 | Gestão de dados sensíveis                   | 3    | 5                                   | 15                            | 5                                                  | 15                            |
|                  | D13 | Plano de Dados Abertos                      | 3    | 4                                   | 12                            | 4                                                  | 12                            |
| Técnica          | D14 | Fonte da informação                         | 2    | 5                                   | 10                            | 5                                                  | 10                            |
|                  | D15 | Formato dos dados                           | 2    | 5                                   | 10                            | 5                                                  | 10                            |
|                  | D16 | Qualidade dos dados                         | 2    | 4                                   | 8                             | 4                                                  | 8                             |
|                  | D17 | Metadados                                   | 2    | 5                                   | 10                            | 3                                                  | 6                             |
|                  | D18 | Vocabulário                                 | 2    | 0                                   | 0                             | 0                                                  | 0                             |
|                  | D19 | Dimensão geográfica                         | 2    | 5                                   | 10                            | 5                                                  | 10                            |
|                  | D20 | Catalogação dos dados                       | 2    | 5                                   | 10                            | 5                                                  | 10                            |
| Gerencial        | D21 | Priorização e Estratégias de abertura       | 2    | 5                                   | 10                            | 5                                                  | 10                            |
|                  | D22 | Estratégias de investimento e cronograma    | 2    | 4                                   | 8                             | 4                                                  | 8                             |
|                  | D23 | Manutenção e correção de problemas          | 2    | 3                                   | 6                             | 3                                                  | 8                             |
| Reúso            | D24 | Quantidade de conjuntos de dados publicados | 3    | 4                                   | 12                            | 4                                                  | 12                            |
|                  | D25 | Quantidade de aplicativos desenvolvidos     | 3    | 0                                   | 0                             | 0                                                  | 0                             |
|                  | D26 | Quantidade de download dos aplicativos      | 3    | 0                                   | 0                             | 0                                                  | 0                             |
|                  | D27 | Engajamento da sociedade civil              | 3    | 4                                   | 12                            | 4                                                  | 12                            |
|                  | D28 | Informações sobre o reúso de DGA            | 3    | 5                                   | 15                            | 5                                                  | 15                            |
| Total            |     |                                             |      | 211                                 |                               | 205                                                |                               |

Fonte: Elaborado pela autora.

O indicador alcançado na métrica DGABr por ambos os conjuntos de dados é um bom resultado. A disponibilização dos DGA por meio de API é um dos fatores que contribuíram para aumentar a pontuação, principalmente nas dimensões que consideram os princípios dos dados abertos e nas dimensões



técnicas avaliadas, já que a arquitetura de uma API já garante a existência de padrões, mínimos que sejam, para organização dos dados dentro da aplicação.

Porém, conforme apontado por Buckland (1999), os mapeamentos entre vocabulários são uma questão de interesse central para a Ciência da Informação. Essa questão ainda permanece atual, visto que a interoperabilidade entre sistemas, requisito para o efetivo reúso dos DGA, deve incluir também padrões para tratar o controle terminológico. Desta forma, o conceito de dados abertos e de web semântica deve estabelecer um modelo de estrutura de dados que favoreça o processamento por máquina e a ligação entre dados e informações de bases diferentes, através de relacionamentos semânticos, proporcionando o aumento do conhecimento para tomadas de decisão e facilitando o acesso aos dados pela sociedade civil (SANTARÉM SEGUNDO, 2015).

A princípio, o número de conjuntos de dados sobre a COVID-19 disponibilizados em APIs é pequeno, mas analisando a quantidade de atributos disponibilizados por cada conjunto de dados, o volume é extenso. A política de DGA do Brasil tem evoluído em alguns sentidos e retrocedido em outros (SILVA, 2022), porém para os DGA sobre COVID-19, a utilização de API foi fundamental para que os dados fossem processados por máquina e o seu acesso fosse ampliado, inclusive compartilhado com outros países.

Passados dois anos de pandemia, espera-se que as incertezas iniciais sobre o compartilhamento de DGA sobre a COVID-19 estejam superadas e que o Ministério da Saúde garanta a manutenção e correção de eventuais problemas identificados nos DGA. A flexibilização de regras e novas formas de parametrização na API será importante para que a recuperação de dados seja mais assertiva e rápida para o usuário que não tenha experiência com desenvolvimento de sistemas e a manipulação de grandes conjuntos de dados.

## **5 Considerações finais**

Através deste estudo foi possível apresentar um breve mapeamento dos DGA sobre a COVID-19, disponibilizados por meio de API, no Brasil. Os dados sobre a notificação de casos de COVID-19 e a campanha de vacinação contra a

COVID-19 são dados estratégicos e relevantes para monitoramento da pandemia no Brasil e estão publicados. A disponibilização destes conjuntos de dados, através de APIs, permitiu um compartilhamento mais rápido e efetivo, aumentando seu potencial de reúso. Na avaliação realizada com a métrica DGABr, foi alcançado o indicador 4 para ambos conjuntos de dados, que mantiveram uma grande similaridade em todas as dimensões de análise. O resultado mostra que os DGA foram publicados de forma minimamente ordenada pelo órgão, acusando um maior grau de maturidade em relação a publicação de DGA através de arquivos para *download*.

Apesar de terem sido identificados vários pontos de melhoria nos conjuntos de DGA e até questões políticas que podem ter influenciado na veracidade e completude dos dados, espera-se que a manutenção e correção dos dados sejam constantes para que essa valiosa fonte de informação possa contribuir com análises sobre a pandemia de COVID-19 nos próximos anos. A utilização de métodos e técnicas que contribuam para a ligação e interoperabilidade desses DGA com outras bases de dados também são melhorias fundamentais para garantia da qualidade e genuinidade dos dados. Por fim, evoluções na política de DGA, que incluam a definições de recursos e padrões, e não somente recomendações, serão essenciais para melhoria dos DGA e seu reúso pela sociedade.

A avaliação realizada apresentou resultados do começo de uma pesquisa, pertencente a um projeto maior, sobre a temática no Brasil, e considerou, neste momento, as documentações publicadas sobre os conjuntos de dados. O consumo de DGA direto à API foi utilizado como ferramenta de apoio. Como trabalho futuro, e continuidade desta pesquisa, sugerem-se estudos avaliando o efetivo consumo dos DGA e a comparação e validação com outras fontes de informação, sobretudo de menor granularidade, como Estados e Municípios, sobre a COVID-19 no Brasil.

## Referências

- BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de Informação:** conceitos e tecnologia das máquinas de busca. Porto Alegre: Bookman, 2013.
- BRANDT, M. B *et al.* Catalogação de metadados: descrição de metadados de negócio a partir dos princípios e objetivos bibliográficos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 24, n. 3, p. 3-18, 2019. Disponível em: <https://doi.org/10.1590/1981-5344/2930>. Acesso em: 18 jan. 2022.
- BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Senado Federal, 1988.
- BRASIL. Ministério do Planejamento, Orçamento e Gestão. **Plano de dados abertos para o Ministério do Planejamento**, 2016a. Disponível em: [http://www.planejamento.gov.br/tema/governo-aberto/plano-de-dados-abertos-pda/plano\\_dados\\_abertos.pdf](http://www.planejamento.gov.br/tema/governo-aberto/plano-de-dados-abertos-pda/plano_dados_abertos.pdf). Acesso em: 19 abril 2022.
- BRASIL. Decreto nº 8.638, de 15 de janeiro de 2016. Institui a Política de Governança Digital no âmbito dos órgãos e das entidades da administração pública federal direta, autárquica e fundacional. **Diário Oficial da União**: Seção 1, Brasília, DF, p. 2-3, 18 jan. 2016b.
- BRASIL. Decreto nº 9.319/2018, de 21 de março de 2018. Institui o Sistema Nacional para a Transformação Digital e estabelece a estrutura de governança para a implantação da Estratégia Brasileira para a Transformação Digital. **Diário Oficial da União**: Seção 1, Brasília, DF, p.2, 21 mar. 2018.
- BRASIL. Decreto nº 10.046/2019, de 9 de outubro de 2019. Dispõe sobre a governança no compartilhamento de dados no âmbito da administração pública federal e institui o Cadastro Base do Cidadão e o Comitê Central de Governança de Dados. **Diário Oficial da União**: Seção 1, Brasília, DF, p. 2, 9 out. 2019.
- BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. Governo Digital: transformação digital - o que é? **Gov**, Brasília, 2020. Disponível em: <https://www.gov.br/governodigital/pt-br/transformacao-digital/o-que-e>. Acesso em: 19 abril 2022.
- BRASIL. Ministério da Saúde. Conecte SUS cidadão. **Gov**, Brasília, 2022a. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/conecte-sus>. Acesso em: 19 de abril 2022.
- BRASIL. Ministério da Saúde. **COVID-19 NO BRASIL**. 2022b. Disponível em: [https://infoms.saude.gov.br/extensions/covid-19\\_html/covid-19\\_html.html](https://infoms.saude.gov.br/extensions/covid-19_html/covid-19_html.html). Acesso em: 19 abril 2022.

BUCKLAND, M. Vocabulary as a central concept in library and information science. *In*: THIRD INTERNATIONAL CONFERENCE ON CONCEPTIONS OF LIBRARY AND INFORMATION SCIENCE (CoLIS3), 3., 1999, Dubrovnik. **Proceedings** [...]. Lokve: Benja Publishing, 1999. p. 3-12.

CAFÉ, L.; SALES, R. Organização da informação: conceitos básicos e breve fundamentação teórica. *In*: ROBREDO, J.; BRÄSCHER, M. (orgs.). **Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento**. Brasília DF: IBICT, 2010. p. 335.

CONTROLADORIA-GERAL DA UNIÃO (CGU). **Painel Monitoramento Dados Abertos**. Brasília, 2022. Disponível em: <http://paineis.cgu.gov.br/dadosabertos/index.htm>. Acesso em: 19 abril 2022.

CRUZ, M. C. A. E.; FERNEDA, E.; FUJITA, M. S. L. A disponibilização de vocabulário controlado aos usuários para a recuperação da informação. **Revista Ibero-Americana de Ciência da Informação**, Brasília, v. 15, p. 266-282, 2022.

ESCOLA NACIONAL DE ADMINISTRAÇÃO PÚBLICA (ENAP). **Elaboração de Plano de Dados Abertos: módulo 1 - conceitos de dados abertos**. Brasília: ENAP, 2017.

FERNEDA, Edberto. **Introdução aos modelos computacionais de recuperação de informação**. Rio de Janeiro: Ciência Moderna, 2012.

G1 *et al.* Veículos de comunicação formam parceria para dar transparência a dados de Covid-19. **G1**, São Paulo, 8 jun. 2020. Disponível em: <https://g1.globo.com/politica/noticia/2020/06/08/veiculos-de-comunicacao-formam-parceria-para-dar-transparencia-a-dados-de-covid-19.ghtml>. Acesso em: 19 abril 2022.

GARCIA, R. M. **Modelos de comportamento de busca de informação: contribuições para a organização da informação**. 2007. Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências de Marília, 2007.

GOOGLE. **Google Play Store**. 2021. Disponível em: <https://play.google.com/store>. Acesso em: 19 abril 2022.

INFRAESTRUTURA NACIONAL DE DADOS ABERTOS (INDA). **Política de Dados Abertos**. [2022?]. Disponível em: <https://wiki.dados.gov.br>. Acesso em: 19 abril 2022.

MACEDO, D. F.; LEMOS, D. L. S. Dados abertos governamentais: iniciativas e desafios na abertura de dados no Brasil e outras esferas internacionais. **Atoz: Novas Práticas em Informação e Conhecimento**, Curitiba, v. 10, n. 2, p. 14-26, 2021. Disponível em: <http://dx.doi.org/10.5380/atoz.v10i2.77737>. Acesso em: 18 jan. 2022.

MARTINS, M. C. B. **Indexação e controlo da terminologia em bibliotecas do ensino superior politécnico em Portugal: o sistema no instituto politécnico de Portalegre**. Salamanca: Ediciones Universidad de Salamanca, 2013.

MINISTÉRIO DA SAÚDE. Secretaria-Executiva. **Plano de Dados Abertos para o Ministério da Saúde 2020-2021**. Brasília: Ministério da Saúde, 2020. Disponível em: [https://sage.saude.gov.br/sistemas/apresentacoes/Plano-de-Dados-Abertos\\_MS\\_2020-2021.pdf](https://sage.saude.gov.br/sistemas/apresentacoes/Plano-de-Dados-Abertos_MS_2020-2021.pdf). Acesso em: 19 abril 2022.

OPEN DataSUS. **Portal**. 2022. Disponível em: <https://opendatasus.saude.gov.br/>. Acesso em: 19 abril 2022.

OPEN KNOWLEDGE BRASIL (OKB). **Emergência dos dados: Como o Índice de Transparência da Covid-19 impulsionou a abertura de dados da pandemia no Brasil**. São Paulo: Open Knowledge Brasil, 2021.

PORTAL BRASILEIRO DE DADOS ABERTOS (PBDA). **Campanha Nacional de Vacinação contra Covid-19**. 2022a. Disponível em: <https://dados.gov.br/dataset/covid-19-vacinacao>. Acesso em: 19 abril 2022.

PORTAL BRASILEIRO DE DADOS ABERTOS (PBDA). **Notificações de Síndrome Gripal**. 2022b. Disponível em: <https://dados.gov.br/dataset/covid-19-vacinacao>. Acesso em: 19 abril 2022.

SANTARÉM SEGUNDO, J. E. Web semântica, dados ligados e dados abertos: uma visão dos desafios do Brasil frente às iniciativas internacionais. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 16., 2015, João Pessoa. **Informação, memória e patrimônio: do documento às redes**. João Pessoa: UFPB, 2015.

SAUDATE, A. **APIs REST: Seus serviços prontos para o mundo real**. São Paulo: Casa do Código, 2021.

SILVA, P. N. **DGABr: Métrica Brasileira**. Belo Horizonte, 2018. Disponível em: <https://www.dgabr.com.br/>. Acesso em: 19 abril 2022a.

SILVA, P. N. **Dados governamentais abertos: métricas e indicadores de reúso**. 2018. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2018b.

SILVA, P. N. Acesso à informação no Brasil: política de dados governamentais abertos. *In*: GERALDES, E. *et al.* (orgs.). **Dez anos da lei de acesso à informação: limites, perspectivas e desafios**. São Paulo: INTERCOM, 2022. p. 748.

## **Open Government Data on COVID-19 in Brazil: assessing the Reuse Potential**

**Abstract:** In the COVID-19 pandemic, access to data on the disease has become strategic for controlling public health measures. Faced with the health emergency, a large volume of data needed to be minimally organized and made available in a quick and automated way, composing the open government data. After two years of a pandemic and in order to present an overview of the publication of open data by the federal government of Brazil, on COVID-19, this study sought to evaluate the open government data made available through the Application Programming Interface (API). The methodology involved the identification of datasets on COVID-19 in Brazil, in Application Programming Interface, until April 2022, the analysis of the documentation and the evaluation using the DGABr metric. The evaluation considered the five perspectives of the metric that measures fundamental elements about the open government data, essential for interoperability and consequently reuse of the data and was based on the published documentation. As results, the open government data on COVID-19, made available in Application Programming Interface, presented a good score in the metric, reaching level 4. This result indicates that the use of APIs was an important and agile technological resource for the organization and availability of open government data, promoting its reuse. However, it is important to highlight that this availability to society was late, it needs constant improvements, mainly in technical issues such as the connection of data with other sources, and that the effective reuse actions were limited to data visualization panels on COVID-19.

**Keywords:** open government data; open data; COVID-19; API; reuse

Recebido: 27/05/2022

Aceito: 21/12/2022

### **Declaração de autoria**

**Concepção e elaboração do estudo:** Patrícia Nascimento Silva.

**Coleta de dados:** Patrícia Nascimento Silva.

**Análise e interpretação de dados:** Patrícia Nascimento Silva.

**Redação:** Patrícia Nascimento Silva.

**Revisão crítica do manuscrito:** Patrícia Nascimento Silva.

**Como citar:**

NASCIMENTO SILVA, Patrícia. Dados governamentais abertos sobre a COVID-19 no Brasil: avaliação do potencial de reúso. **Em Questão**, Porto Alegre, v. 29, e-124820, 2023. <https://doi.org/10.19132/1808-5245.29.124820>



---

<sup>1</sup>Todos os documentos do Quadro 2 estão disponíveis em repositório público, no endereço:  
[https://github.com/pprof2022/DGA\\_covid](https://github.com/pprof2022/DGA_covid)

<sup>2</sup> O termo granularidade é utilizado para indicar partes menores de um sistema.