



TOOLKIT #1 | TRANSPARÊNCIA COVID-19

Publicação de microdados de casos de Covid-19

Microdados são fundamentais para a produção científica e também para a formulação de políticas públicas mais efetivas, que levem em consideração aspectos do território e dos diferentes grupos da população. As boas práticas e padrões presentes neste guia ajudam a organizar a publicação dessas bases, para que sejam mais acessíveis e reutilizáveis.

Um dos principais desafios para a transparência da Covid-19, a disponibilização de Microdados é o instrumento com maior potencial para qualificar as informações que o Estado pode oferecer a pesquisadores, jornalistas, voluntários e demais protagonistas e interessados no enfrentamento à pandemia.

Aliada às devidas técnicas de anonimização e aos padrões abertos, a publicação de uma base de Microdados detalhada atende aos padrões internacionais de abertura, e permite o desenvolvimento de análises, comparações e projeções mais precisas, complexas e informativas. Além disso, a disponibilização da base pode fazer o ente pontuar em diversos critérios no Índice de Transparência da Covid-19 (ITC-19), como os das dimensões Conteúdo e Formato, impulsionando o desempenho no ranking geral.

Nesta publicação, você conhecerá alguns passos para abrir um conjunto de Microdados. O Toolkit de Publicação de Microdados será atualizado com novas orientações conforme identificarmos os avanços dos estados nas avaliações. Não deixe de conferir também [nosso Template](#), que pode tanto te ajudar a organizar os dados que já possui, quanto vislumbrar aqueles que ainda precisa coletar.¹

ABRINDO E PROTEGENDO DADOS

O grau de detalhamento (granularidade) das informações sobre casos suspeitos e confirmados de Covid-19 mostrou-se uma das maiores barreiras para a ampliação da transparência da pandemia no Brasil. Como revelado na [última avaliação](#) do ITC-19 em 2020, realizada em dezembro, a publicação de microdados foi o critério menos cumprido pelos entes. Isso quer dizer que não é possível acessar dados individualizados sobre os casos, contendo detalhamentos como sexo, faixa etária, raça/cor e localização para cada um deles, na maior parte dos estados e, principalmente, das capitais brasileiras -- mesmo após meses de pandemia e de registro diário desses casos.

A publicação de microdados sempre suscita dúvidas em gestores e gestoras devido ao fato de a publicação detalhada dos dados apresentar potencial para ferir a privacidade dos cidadãos a que os dados se referem. No entanto, também é fato que quanto mais detalhes tivermos sobre os casos, mais análises de qualidade conseguimos produzir. Como resolver este aparente conflito?

¹ A primeira versão deste toolkit foi publicada em 06 de maio de 2020 e pode ser acessada [aqui](#).

Dados de saúde são considerados dados sensíveis pela Lei Geral de Proteção de Dados Pessoais, e, por isso, é necessário ter o máximo de cautela ao coletar e administrar esses dados. No caso da disponibilização de dados sobre a Covid-19, a solução mais simples para mitigar os riscos à privacidade se encontra nos diferentes **níveis de detalhamento e agregação** que podem ser empregados aos dados coletados. Quando a publicação vincula diretamente atributos de idade e sexo dos pacientes a seus respectivos locais de internação, por exemplo, isso pode facilitar a identificação de um determinado paciente, sobretudo em municípios pequenos. **Esse tipo de publicação não é incentivada pelo ITC-19.**

No entanto, uma opção para não expor os pacientes e nem deixar de publicar esses dados -- que são todos igualmente importantes -- é divulgar o quantitativo de casos por hospital, e, separadamente, publicar os Microdados de casos com as informações de idade e sexo. Afinal, não é tão importante saber especificamente se um homem de 35 anos está internado no Hospital das Flores, mas é urgente saber quantas pessoas estão sendo atendidas no Hospital das Flores, e é fundamental saber que o vírus está acometendo homens de 35 anos.

Lógica similar se aplica à disponibilização de dados de casos por bairro/distrito em municípios pequenos e médios, que ainda possuem poucas confirmações de Covid-19. Para essas situações, é necessário redobrar o cuidado com a exposição de pacientes, e muitas vezes analisar caso a caso, dado que é comum que muitos munícipes se conheçam. Por este motivo, o ideal é publicar informações (qualitativas e quantitativas) de casos por bairro/distrito somente para capitais e grandes cidades.

ENCONTRANDO E ORGANIZANDO DADOS

Sistema e-SUS VE/Notifica

Em todos os níveis de governo, diversos sistemas e tecnologias têm sido empregados para coletar dados de casos sobre o novo coronavírus. No entanto, o [e-SUS Notifica](#) (anteriormente conhecido como e-SUS VE), sistema criado pelo Governo Federal para registro obrigatório dos casos de novo Covid-19, lança luz sobre quais tipos de dados a administração deve coletar, e, portanto, pode disponibilizar, após aplicar as devidas medidas de anonimização.

Conforme a versão mais recente do [manual de navegação](#) do e-SUS Notifica publicado pelo Ministério da Saúde, gestores devem cadastrar os seguintes dados referentes às notificações de Covid-19:

- Tem CPF? Caso não haja, o CNS (Cartão Nacional de Saúde) e o nome completo da mãe serão solicitados.
- Estrangeiro? Se sim, país de origem e passaporte serão solicitados.
- É profissional de saúde?
- Profissional de segurança
- CBO (Código/ocupação, caso seja profissional da saúde)
- CNS
- CPF
- Nome completo
- Data de nascimento
- Sexo
- Raça/cor
- Etnia (em caso de indígena)
- CEP
- Passaporte (em caso de estrangeiros)
- Logradouro
- Número
- Complemento
- Bairro
- Estado de residência
- Município de residência
- Telefone
- Data da notificação
- Sintomas (Febre, Tosse, Outros, Dor de garganta, Dispneia, Dor de cabeça, Distúrbios gustativos, Distúrbios olfativos, Coriza, Assintomático,)
- Data do início dos sintomas
- Condições (Doenças respiratórias crônicas descompensadas; Doenças renais crônicas em estágio avançado; Portador de doenças cromossômicas ou estado de fragilidade; Diabetes, Imunossupressão; Doenças cardíacas crônicas; Gestante; Puérpera - até 45 dias do parto; Obesidade)
- Estado do teste (Solicitado; Coletado; Concluído; Exame não solicitado)

- Tipo de teste (Teste rápido - Anticorpo; Teste rápido - Antígeno; RT-PCR; Testes sorológico)
- Data do teste (PCR/Rápidos)
- Data do testes (sorológico)
- Resultado (PCR/Rápidos) (Negativo; Positivo; Inconclusivo ou Indeterminado)
- Teste Sorológico (IgA; IgG; IgM; Anticorpos Totais)
- Resultado IgA (Reagente; Não reagente; Inconclusivo ou indeterminado)
- Resultado IgG (Reagente; Não reagente; Inconclusivo ou indeterminado)
- Resultado IgM (Reagente; Não reagente; Inconclusivo ou indeterminado)
- Resultados Totais (Reagente; Não reagente; Inconclusivo ou indeterminado)

Como, segundo o manual do Ministério da Saúde, o e-SUS Notifica permite a **exportação da planilha CSV** com o compilado de dados inseridos no sistema pela administração, facilita-se a disponibilização de uma base com Microdados. Além disso, o documento também destaca que é **possível baixar a base do e-SUS Notifica filtrando por Estado, Município, Evolução do Caso, Estado do Teste, Classificação Final, Resultado do Teste, CNES, Data de Notificação Início, Data da Notificação Fim, Data de Nascimento Início, Data de Nascimento Fim, Dados da coleta do testes, Dados da Coleta Teste Fim, Data de Encerramento Início, Data de Encerramento Fim**. Assim, é possível extrair segmentos de informações que podem ser disponibilizadas de forma agregada ou ainda gerar análises para subsidiar a tomada de decisão. Por exemplo, filtrando pelo **CNES**, é possível contabilizar a quantidade de casos por unidade de saúde, enquanto usando os filtros de **Data de Notificação Início** e **Data de Notificação Fim** é possível saber quantos casos foram registrados em determinado período.

Dentre os dados requeridos para a notificação no sistema do Governo Federal, é possível observar que ao suprimir somente as colunas contendo **CPF, Passaporte (no caso de estrangeiros), Nome completo, Telefone, CNS, CEP, Logradouro, Número, Complemento e Bairro**, os gestores já podem avançar significativamente na publicação periódica de um conjunto de informações bastante completo. É preciso, entretanto, realizar um processo de pseudoanonimização (substituindo o CPF por outro código único que não remeta a um indivíduo) de forma a garantir que não haja registros duplicados. Veja nas referências ao final deste guia

Recomendamos que sejam disponibilizados dados de **todos os casos notificados**, não somente os de casos confirmados. Isso porque, com esses números, os pesquisadores conseguem, por exemplo, acompanhar a capacidade de testagem dos entes ou mesmo estimar com mais precisão a quantidade de casos que podem estar fora das estatísticas oficiais de confirmados.

Vale atentar ainda para o fato de o e-SUS Notifica permitir atualização caso a caso. Gestores que focarem a disponibilização de Microdados somente neste sistema deverão extrair e publicar as bases com frequência constante para permitir o acompanhamento da evolução do número de recuperados e óbitos. É de suma importância que o ente seja capaz de **informar com clareza a data de atualização dos dados publicados, o que pode ser feito adicionando a coluna de "Data de Atualização" na base de dados**. Recomendamos que a administração faça backup diário das bases de dados, uma vez que o sistema sobrescreve as informações.

BAIXE O MODELO DE PADRONIZAÇÃO DOS MICRODADOS

A OKBR produziu um template com sugestão de atributos e formato para padronização da base de microdados [Faça o download aqui.](#)

Código da Variável	Nome da Variável	Formato	Descrição
data_publicacao	Data de publicação da base	Data (DD-MM-AAAA)	Data em que o arquivo CSV foi atualizado. Alterar a cada publicação.
data_sintomas	Data de início dos sintomas	Data (DD-MM-AAAA)	Data em que se registrou a existência dos primeiros sintomas de Covid-19 no paciente.
data_notificacao	Data da Notificação	Data (DD-MM-AAAA)	Data em que o paciente deu entrada no sistema de saúde e recebeu as primeiras informações sob
classificacao_final	Classificação Final	Texto	Descrição do caso, após investigação.
critério_confirmacao	Critério de confirmação	Texto	Tipo de exame realizado para investigação do caso.
data_encerramento	Data de encerramento	Data (DD-MM-AAAA)	Data em que o caso foi encerrado, após alta ou óbito do paciente.
situacao_teste	Situação do teste	Texto	Descrição do status de testagem do paciente.
data_coleta	Data da coleta do teste	Data (DD-MM-AAAA)	Data em que o teste foi coletado.
tipo_teste	Tipo de teste aplicado	Texto	Padrão do teste utilizado para confirmação dos casos.
resultado_teste	Resultado do teste	Texto	Descrição do resultado do teste aplicado.
evolucao	Evolução	Texto	Evolução clínica do caso acompanhado.
municipio	Município (Nome)	Texto	Nome do Município onde o caso foi registrado.
cod_ibge	Município (Código)	Texto	Código IBGE do Município.
estado	Estado	Texto	Nome do Estado onde o caso está sendo tratado.
municipio_residencia	Município de Residência (Nome)	Texto	Nome do Município onde o paciente reside (se houver). Caso o registro seja de pessoa em situação
cod_ibge_res	Município de Residência (Código)	Texto	Código IBGE do Município onde o paciente reside (se houver)
estado_residencia	Estado de Residência	Texto	Nome do Estado onde o paciente reside (se houver). Caso o registro seja de pessoa em situação de
pais_origem	País de origem	Texto	País de origem do paciente, em caso de estrangeiro

Se tiver comentários para aprimoramento ou dúvidas sobre como implementar o modelo, escreva para transparenciacovid19@ok.org.br

OUTRAS REFERÊNCIAS

“Publicadores de Dados: da gestão estratégica à abertura” — para uma visão mais ampla sobre o processo de publicação de dados no contexto de uma organização, o e-book produzido pela Open Knowledge Brasil por meio de sua Escola de Dados é uma boa referência.

Faça o download gratuito em vários tipos de formato [aqui](#).

“Pseudoanonimização de dados com editores de planilhas”. Este tutorial da Escola de Dados explica, de maneira bastante simples, o processo de pseudoanonimização para uma base de pequeno porte. Ainda que não seja aplicável ao processo de uma base com mais de um milhão de linhas, como é o caso de muitas cidades, o conteúdo é interessante para compreender o conceito de substituição de nomes por códigos únicos (a ideia de “pseudônimos”), que não permitam a identificação de indivíduos. Acesse [aqui](#). A Escola de Dados possui, ainda, diversos outros tutoriais básicos, intermediários e avançados para o trabalho com dados.

Disponível em: <https://escoladedados.org/tutoriais/>.

SOBRE A OKBR

A OKBR, também conhecida como Rede pelo Conhecimento Livre, é uma organização da sociedade civil sem fins lucrativos e apartidária que atua no país desde 2013. Desenvolvemos e incentivamos o uso de tecnologias cívicas e de dados abertos, realizamos análises de políticas públicas e promovemos o conhecimento livre para tornar a relação entre governo e sociedade mais transparente e participativa.

Saiba mais no site: <http://ok.org.br>.

Contato: transparenciacovid19@ok.org.br