Parte II – Produzindo e organizando informação para ação

14. MonitoraCovid-19: informação e disseminação de indicadores em uma pesquisa multidisciplinar

Raphael de Freitas Saldanha
Mônica de Avelar Figueiredo Mafra
Magalhães
Marcel de Moraes Pedroso

Diego Ricardo Xavier
Paulo Roberto Borges de Souza Junior
Christovam Barcellos

SciELO Books / SciELO Livros / SciELO Libros

SALDANHA, R. F., XAVIER, D. R., MAGALHÃES, M. A. F. M., SOUZA JUNIOR, P. R. B., PEDROSO, M. M., and BARCELLOS, C. MonitoraCovid-19: informação e disseminação de indicadores em uma pesquisa multidisciplinar. In: FREITAS, C. M., BARCELLOS, C., and VILLELA, D. A. M., eds. *Covid-19 no Brasil: cenários epidemiológicos e vigilância em saúde* [online]. Rio de Janeiro: Observatório Covid-19 Fiocruz; Editora Fiocruz, 2021, pp. 229-249. Informação para ação na Covid-19 series. ISBN: 978-65-5708-049-8. <https://doi.org/10.7476/9786557081211.0015>.

All the contents of this work, except where otherwise noted, is licensed under a [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença [Creative Commons Atribuição 4.0](https://creativecommons.org/licenses/by/4.0/).

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia [Creative Commons Reconocimiento 4.0](https://creativecommons.org/licenses/by/4.0/).

MonitoraCovid-19

informação e disseminação de indicadores em uma pesquisa multidisciplinar

*Raphael de Freitas Saldanha, Diego Ricardo Xavier,
Mônica de Avelar Figueiredo Mafra Magalhães,
Paulo Roberto Borges de Souza Junior,
Marcel de Moraes Pedroso e Christovam Barcellos*

O MonitoraCovid-19 foi construído no Instituto de Comunicação e Informação Científica e Tecnológica da Fundação Oswaldo Cruz (Icict/Fiocruz), que se dedica à criação e manutenção de observatórios e sistemas de informática para doenças e agravos, com recortes para informação científica e tecnológica e também para comunicação em saúde. O processo de sua construção, implementação e manutenção será aqui apresentado com ênfase nas questões institucionais e tecnológicas envolvidas.

Ainda que distantes, em outros continentes, os primeiros casos e óbitos de Covid-19 naturalmente captavam a atenção de alguns profissionais do Laboratório de Informação em Saúde (LIS/Icict), e discussões sobre a pandemia surgiam nas conversas e e-mails, principalmente sobre estimativas, taxas e outros cálculos importantes para a doença. Em seguida, começamos a nos questionar sobre a situação brasileira. Na época, as questões eram “Quando a pandemia irá chegar aqui?”, “Como o governo está se preparando?”, “O que deve ser feito?”.

Após 26 de fevereiro de 2020, com os primeiros casos autóctones de Covid-19 no Brasil, alguns cálculos começaram a ser feitos para a realidade brasileira. Acompanhamos pela imprensa os números que timidamente começaram a ser divulgados. Nos meses de fevereiro e março, entrevistas coletivas do Ministério da Saúde (MS) eram diariamente realizadas e transmitidas por praticamente todas as emissoras de televisão. Conforme os números aumentavam, a discussão entre alguns pesquisadores, tecnologistas e bolsistas ficava mais intensa.

Nessa época, o MS não disponibilizava um repositório de dados sobre casos e óbitos aberto para o público. Algumas iniciativas pontuais para registrar os números apresentados nas coletivas começaram a surgir, sendo divulgadas na internet em grupos

de conversa de jornalistas e cientistas de dados. Podem-se mencionar o *dataset* publicado na plataforma Kaggle (Fontes, 2020) e os dados publicados no repositório do GitHub (Cota, 2020) como algumas das primeiras fontes de dados estruturados e atualizados regularmente sobre a pandemia de Covid-19 no Brasil.

Com a disponibilidade de dados internacionais, como os do repositório do European Centre for Disease Prevention and Control (ECDC), e de dados nacionais por unidade da federação (UF), surgiu no Icict a primeira iniciativa formal de acompanhar os dados da pandemia de Covid-19.

Em todo o mundo, com base em alguns indicadores ou utilizando modelos estatísticos e matemáticos, foram criados diversos painéis de indicadores visando ao monitoramento e previsão de cenários em diversas escalas, da local à global (cf. cap. 12 desta coletânea). O sistema MonitoraCovid19, no entanto, adquiriu um caminho próprio, com o envolvimento de grupos de pesquisa e desenvolvedores de tecnologias de informação, caracterizando-se pela multiplicidade de indicadores e fontes de informação utilizadas e por seu uso permanente por gestores e sociedade civil. O desenvolvimento desse sistema, os desafios institucionais e tecnológicos enfrentados e os seus produtos derivados são o objeto deste capítulo.

A CONSTITUIÇÃO DO PROJETO E SEU DESENVOLVIMENTO INICIAL

Em março de 2020, com o avanço da pandemia de Covid-19 no Brasil, o trabalho presencial na Fiocruz começou a ser flexibilizado. As salas e laboratórios passaram a ficar gradativamente mais vazias, com eventos, reuniões e encontros postergados. Após algumas semanas de adequação ao novo regime de trabalho, a condução dos projetos e pesquisas em andamento foi reorganizada. Novas datas e prazos foram acordados com instituições parceiras, que invariavelmente também se encontravam na mesma situação. Reuniões virtuais periódicas começaram a ser agendadas, o fluxo de e-mails aumentou e grupos em aplicativos de conversa se multiplicaram enquanto novas diretivas e planos de contingência eram publicados e atualizados pela Fiocruz e pelo Icict.

A missão do Icict se traduz em ações integradas de pesquisa e ensino; comunicação e informação; e gestão e desenvolvimento institucional, alinhadas a quatro eixos temáticos: Desafios do SUS; Ciência e Tecnologia, Saúde e Sociedade; Inovação na Gestão; e Saúde, Ambiente e Sustentabilidade. O objetivo comum desses eixos é fortalecer o SUS e promover melhores condições de vida e saúde da população. O instituto atua no campo da informação e comunicação em saúde, área interdis-

ciplinar do conhecimento que por sua ampla abrangência gera, ao mesmo tempo, desafios e oportunidades (Icict/Fiocruz, 2020).

O LIS do Icict desenvolve e mantém diversos observatórios e sistemas dedicados a monitorar a saúde da população brasileira e de outros países conveniados, acumulando larga experiência na área. Em colaboração com o MS e com o Instituto Brasileiro de Geografia e Estatística (IBGE), o LIS é responsável pelo aprimoramento do Sistema de Informação sobre Mortalidade, Sistema de Monitoramento da Aids (MonitorAids) e pelo planejamento e desenvolvimento da Pesquisa Nacional de Saúde do IBGE (PNS). O LIS desenvolve sistemas de indicadores que foram criados e são ofertados publicamente com foco no acesso e uso de serviços de saúde pela população brasileira (Proadess), população de idosos (Sisap), qualidade da água e saúde (Água Brasil), monitoramento de clima e saúde (Observatório de Clima e Saúde) e acompanhamento de indicadores de mortalidade infantil (Monitorimi). Esses sistemas dispõem de informações que servem como subsídios para gestores e tomadores de decisão, e também buscam apresentar as informações em linguagem acessível à comunidade acadêmica e ao cidadão usuário do Sistema Único de Saúde (SUS). O LIS abriga o Núcleo de Geoprocessamento, voltado para a produção, adequação e atualização de dados e análises espaciais que relacionam dados socioeconômicos, de saúde e ambiente. Seus pesquisadores, tecnologistas e bolsistas estão rotineiramente envolvidos na construção e/ou manutenção de algum desses sistemas, o que permitiu a acumulação de conhecimentos e técnicas que cobrem diversos aspectos da implementação de observatórios. Nesse contexto surgiu o projeto MonitoraCovid-19.

DESENVOLVIMENTO TECNOLÓGICO

Em meados de março de 2020 foi criada uma planilha no *software* Excel. Com dados do ECDC, foram produzidos alguns gráficos de casos e óbitos e cálculos de fatores de crescimento comparativos entre países.

O projeto nasceu em uma planilha feita para atender à demanda por antevisão de cenários sobre a evolução da pandemia nos países e sua chegada no Brasil. Contudo, a atualização da planilha com dados novos e seu incremento com mais funções se mostrou insustentável. A cada dia entravam novos países na lista, as fontes de dados mudavam seus padrões repentinamente, os dados precisavam ser baixados manualmente, convertidos para outro formato e inseridos na planilha manualmente, e as fórmulas precisavam ser atualizadas para a entrada de novas datas. Em função dessa dinâmica, o editor de planilhas estava em seu limite.

Figura 1 – Planilha de cálculos

Nesse momento, percebemos que a quantidade de dados aumentava de forma acelerada e que esta seria a tônica da pandemia. O que era só uma curiosidade sobre o processo epidêmico foi se transformando em necessidade de contribuir no enfrentamento do problema de saúde pública. Percebemos nessa época que a pandemia produziria uma grave crise de saúde pública, mas nem nossas piores previsões apontavam para o cenário que se desenharia.

Surgiu a ideia de reproduzir todo esse processo manual de maneira mais automatizada. Um programa que conseguisse baixar os dados da internet, criar uma base única de casos e óbitos por Covid-19 para países e UFs, produzir gráficos e fazer cálculos de forma automática, conforme novos dados fossem disponibilizados. Nesse momento surge o conceito fundamental do MonitoraCovid-19. Passamos, então, a procurar uma ferramenta mais flexível e escalável para a enxurrada de dados e ideias sobre o que fazer com esses dados. O pacote estatístico R nos possibilitou realizar a importação de diversos tipos de dados com mais facilidade e criar um aplicativo on-line.

Uma das primeiras versões do sistema, ainda sem o nome, apresentava apenas seis abas, conforme a Figura 2, de 23 de março de 2020.

Figura 2 – Imagem de uma das primeiras versões do MonitoraCovid-19

De março ao final de novembro de 2020, as atualizações das bases de dados eram feitas manualmente. Durante 253 dias, os dados foram baixados de suas fontes de forma manual, organizados e padronizados diariamente, e passaram a alimentar a aplicação web.

Não fazíamos ideia do quão pesada seria essa rotina de atualização. Ela surgiu ingênua, da nossa intenção de ver os dados novos nos gráficos que estávamos criando. Conforme mais pessoas se interessavam por nossos gráficos, ainda dentro do LIS, crescia em nós o senso de dever: manter o projeto atualizado e funcionando, todos os dias.

Apenas em dezembro de 2020 uma rotina completamente automática de atualização foi colocada a serviço. Essa etapa será melhor descrita adiante.

DATA DRIVEN ANALYSIS

Ao escrever sobre o projeto, é tentador afirmar que ele foi o fruto esperado de um extenso e detalhado planejamento, no qual tudo foi pensado e antevisto, e que cada aba e conteúdo foi planejado para ser daquele jeito, desde o começo. Mas seria desonesto. O projeto nasceu sem ser um projeto, ele era um teste, um tubo de ensaio, e hoje, mesmo sendo uma plataforma estável e confiável, continua sendo um laboratório para explorar possibilidades de novas análises e dados.

O método cartesiano do planejamento, com objetivos, métodos e resultados, tem, obviamente, seu valor. Mas neste caso, invertemos de alguma maneira essa sequência. Não ter uma ideia formalizada do que queríamos nos possibilitou maior liberdade e criatividade para explorar diversas possibilidades de análise *data driven* (Provost & Fawcett, 2013). Gastamos obscenas quantidades de tempo em gráficos que não publicamos, em reuniões longas sobre temas que não abordamos.

Na medida em que o projeto avançava, ainda que de forma discreta, precisávamos de um nome para registrar um domínio na internet e podermos nos referir a ele de forma mais precisa. Após algumas sugestões, decidimos por MonitoraCovid-19.

DIVERSIDADE DE FONTES DE DADOS

O projeto apoia-se atualmente em mais de dez fontes de dados diferentes, que contemplam desde o número de casos e óbitos de Covid-19 até variáveis ligadas direta ou indiretamente à doença, como por exemplo mobilidade urbana, população em risco e medidas legislativas.

Atualmente, utilizamos os dados disponibilizados diariamente pelo John Hopkins Coronavirus Resource Center sobre casos e óbitos para os países; dados do MS para casos e óbitos no Brasil, UFs e municípios; InfoGripe para dados de síndrome respiratória aguda grave (Srag); dados do Facebook e da Universidade de Maryland para pesquisa de sintomas; dados de congestionamentos de trânsito da plataforma Waze disponibilizados pelo IDB Coronavirus Impact Dashboard; dados de mobilidade urbana disponibilizados pelo Google; dados de utilização de transporte público disponibilizados pela plataforma Waze (2020); e metadados de legislações municipais, estaduais e federais sobre Covid-19 disponibilizados pelo site Leis Municipais.

Algumas intercorrências aconteceram, principalmente devido a dificuldades de acesso a dados de casos e óbitos de Covid-19 para as UFs e municípios brasileiros. No início da pandemia (fevereiro e março de 2020), o MS informava a população sobre o número de casos e óbitos exclusivamente por meio de coletivas de imprensa. O MS disponibilizou os dados de casos e óbitos em pelo menos sete sistemas diferentes até o momento, apresentados na Figura 3.

Figura 3 – Telas de diferentes sistemas de informação de casos e óbitos de Covid-19 do Ministério da Saúde

Esses sistemas foram apresentados seguidamente pelo MS, muitas vezes mantendo dois ou três deles em funcionamento e, algumas vezes, com dados divergentes. Embora alguns sistemas ou versões permitissem o *download* de dados, em vários momentos era oficialmente impossível baixar os dados estruturados oficiais de casos e óbitos por Covid-19 a partir do MS.

Essas transições entre sistemas e constantes dificuldades em realizar o *download* diário de dados dificultavam imensamente o processo de atualização da plataforma. Basicamente, a cada dia podia-se esperar algo diferente, fosse na oferta de dados, fosse

até mesmo variações no formato, extensão, horário de disponibilização dos dados e variáveis constantes nos arquivos. Isso exigia grande esforço da equipe para manter os dados diários atualizados.

Como não havia um repositório oficial de dados históricos da epidemia no Brasil, dependíamos de projetos e iniciativas que os coletavam diariamente, durante as coletivas, e disponibilizavam os dados históricos estruturados de forma *on-line*.

Em 3 de junho de 2020, data em que o painel Coronavírus Brasil mantido pelo MS foi parcialmente descontinuado, ocorreu o chamado “apagão de dados”. Nessa data, regredimos quatro meses em termos de transparência de dados oficiais e passamos a depender diretamente de projetos e iniciativas privadas para manter MonitoraCovid com dados atualizados. Passamos, então, a utilizar diariamente os dados disponibilizados pelo projeto Brasil.IO.

Nos dias seguintes ao apagão, chegamos em um único dia (7 de junho de 2020) a até 11.565 usuários que procuravam dados atualizados e confiáveis sobre Covid-19 no Brasil.

O projeto Brasil.IO é uma iniciativa de código aberto liderada pelo cidadão Álvaro Justen, reconhecido desenvolvedor nas comunidades de *software* livre e jornalismo de dados brasileiros. Contando com o trabalho voluntário de dezenas de pessoas, o projeto diariamente acessa os painéis, planilhas, postagens em redes sociais e outras mídias disponibilizadas pelas secretarias estaduais de Saúde para reunir os dados de casos e óbitos de Covid-19 no Brasil, para UFs e municípios.

Como contrapartida institucional e reconhecimento do serviço essencial prestado, o MonitoraCovid-19 procurou direcionar parte de seu financiamento e recursos para o projeto Brasil.IO por meio de doações.

De 3 de junho até 1 de dezembro de 2020, utilizamos diariamente os dados sobre Covid-19 disponibilizados pelo Brasil.IO. A partir de dezembro, notada maior estabilidade e qualidade dos dados disponibilizados pelo MS, voltamos a utilizá-los como fonte primária de dados de casos e óbitos por Covid-19 no Brasil. Essa estabilidade foi interrompida pontualmente no dia 4 de novembro de 2020 por um ataque *hacker* aos servidores do departamento de informática do Sistema Único de Saúde, o Datasus. A disponibilização de dados de Covid-19 pelo MS foi parcialmente interrompida nessa data, mas plenamente retomada apenas alguns dias depois.

GERENCIAMENTO DO PROJETO

Com o início do trabalho remoto, os pesquisadores, tecnologistas e bolsistas do LIS testaram algumas ferramentas de interação: além do e-mail, passamos a usar plataformas de reuniões on-line e, principalmente, grupos de conversa no aplicativo WhatsApp. Além de um grande grupo para interação de todos no laboratório e outro, mais específico para o Núcleo de Geoprocessamento do LIS, criamos alguns grupos para o gerenciamento específico do MonitoraCovid-19. A participação nesses grupos seguiu uma premissa básica: será incluído no grupo quem se dispuser a contribuir, direta e ativamente, na evolução do projeto. Naturalmente, diversas pessoas apoiam o projeto e suas iniciativas, mas a inclusão de todos seria contraprodutiva. Seguimos com a lógica de *need to know basis*.

Por intermédio desses grupos, conseguimos discutir os assuntos e tomar decisões de forma efetiva, diretamente com os envolvidos no projeto. Essa estrutura colaborativa de gestão e atuação no projeto é interinstitucional, contando com a colaboração de vários técnicos e pesquisadores, dentro do Ict e de outros institutos da Fiocruz.

Observamos que, dentro do Ict, com algumas exceções, poucos pesquisadores puderam colaborar diretamente no desenvolvimento contínuo do projeto. Apesar de enviarem sugestões e pedidos de dados em alguns momentos, outras demandas parecem ter limitado a participação de mais profissionais no projeto. Fora do Ict, por sua vez, conseguimos uma colaboração interessante com um pesquisador da Escola Politécnica de Saúde Joaquim Venâncio, que resultou em uma nota técnica publicada e em artigos submetidos a periódicos científicos. Em situações pontuais, no entanto, quando solicitados os pesquisadores com expertise em determinados temas se prontificaram a colaborar. Um exemplo de sucesso foi a parceria entre os pesquisadores para a elaboração da nota técnica sobre a volta às aulas, na qual foi estimada a população com fatores de risco e idosa que convive na mesma habitação com pessoas em idade escolar.

CONSTRUINDO PONTES INTERINSTITUCIONAIS

Ao longo do desenvolvimento do projeto, procuramos agregar novas fontes de dados e análises e, assim, conseguimos estabelecer algumas parcerias, dentro da própria Fiocruz e com outras instituições e empresas.

Na Fiocruz, estreitamos laços com a equipe do Projeto InfoGripe, responsável pela modelagem e monitoramento de dados de Srag no Brasil. Já conhecíamos parte da equipe de outros projetos e eventos institucionais, mas essa parceria se estabeleceu através de canais bem informais, como mensagens trocadas por meio do Twitter e do WhatsApp.

Com a rotina de atualização de dados, passamos a conversar mais intensamente, dirimir dúvidas sobre os dados de Srag que divulgamos, indicar dados e eventos e propor colaborações entre os projetos.

Nessa ocasião, vários grupos que estudavam a epidemiologia e padrões da Covid-19 se reuniram para discutir os dados e possibilidades de análise. Capitanado pela equipe do InfoGripe, o webinar “O Panorama da Covid-19 no Rio de Janeiro e Brasil: onde estamos e para onde vamos?” possibilitou o intercâmbio de informações e alternativas de análise e estudos compartilhados entre diferentes iniciativas.

Figura 4 – Peça de divulgação do evento “O Panorama da Covid-19 no Rio de Janeiro e Brasil: onde estamos e para onde vamos?”

The poster is for a webinar titled "O Panorama da Covid-19 no Rio de Janeiro e Brasil: Onde estamos e para onde vamos?". It features a dark blue background with a teal gradient on the right side. On the left and right edges, there are stylized white virus particles. The text is centered and uses various colors (white, orange, teal) for emphasis. At the bottom, there is a row of logos for participating organizations.

WEBINAR

O Panorama da Covid-19 no Rio de Janeiro e Brasil
Onde estamos e para onde vamos?

Quinta-feira, 10 de setembro, das 19h às 20h30

Participação
 Covid19Analytics, MAVE, Observatório CovidBR
 COVID19: Observatório Fluminense, Monitora Covid

Mediação
 Amanda Rossi (Colaboradora da Revista Piau)

Inscrições pelo link
<https://eesp.fgv.br/evento/webinar-o-panorama-da-covid-19-no-rio-de-janeiro-e-no-brasil>

Logos at the bottom: ICICT, OBR, COVID-19 Observatório Fluminense, PROCC, FIOCRUZ, PUC-RIO, COVID-19 ANALYTICS, FGV EMAP, FGV EESP.

O MonitoraCovid-19 também foi apresentado em alguns outros eventos. Podem-se destacar o de 10 de junho, realizado pelo Programa de Pós-Graduação em Informação e Comunicação em Saúde/Icict, intitulado “Acesso Aberto a Dados de Saúde na Perspectiva da Pandemia”, o de 19 de junho de 2020 organizado pelo Centro de Estudos do Icict, intitulado “Informação em Saúde: importância e desafios no enfrentamento da pandemia”, e o de 4 de agosto de 2020, organizado pelo Fórum de Reportagem sobre a Crise Global em Saúde e intitulado “Como usar o MonitoraCovid-19, Sistema da Fiocruz que Agrupa Dados da Pandemia”.

Externamente, conseguimos parceria direta com empresas para o fornecimento de dados, como a Moovit, que fornece dados sobre circulação de transporte coletivo nas capitais e regiões metropolitanas. A aplicação avalia tendências globais de transporte público e combina pesquisas de opinião com dados remotos para construir um retrato de como as pessoas transitam por suas cidades. No período epidêmico da Covid-19, a empresa criou o indicador de porcentagem de redução no uso de transporte público em comparação à média anterior ao período da pandemia.

Usamos, também, dados de acesso público do Google que apontam indicadores de mobilidade da população com dados agregados e anônimos usados em produtos como o Google Maps. Os indicadores apontam tendências de deslocamento ao longo do tempo por região e em diferentes categorias de locais, como varejo e lazer, mercados e farmácias, parques, estações de transporte público, locais de trabalho e áreas residenciais.

Também incluímos dados do Waze com informações sobre congestionamento de trânsito que podem ser utilizadas para acompanhamento da mobilidade da população (Waze, 2020). Os dados de variação percentual de quilômetros compartilhados são agregados, anonimizados e foram gerados com base em critérios de privacidade dos usuários. Esses relatórios mostram o aumento ou a redução dos quilômetros rodados como uma variação percentual comparada a dados da linha de base, calculada pela média do dia da semana correspondente ao período de duas semanas antes da decretação de emergência (11 a 25 de fevereiro de 2020).

Também incluímos no sistema dados do Facebook, que disponibiliza indicadores de Covid-19 derivados de pesquisas globais de sintomas em sua plataforma. Os dados são mantidos pela Universidade de Maryland, que os compartilha com outros pesquisadores em saúde. Embora seja uma pesquisa que considera informações individuais, esses dados são disponibilizados de forma agregada, o que garante a privacidade do participante. Com base nessas informações, é estimada a porcentagem de pessoas que em determinada região geográfica e em determinado tempo relataram sintomas de Covid-19.

Ganhamos apoios de empresas como a RStudio e a Digital Ocean, com cupons de gratuidade para utilização de seus serviços por um tempo limitado. Essa oportunidade foi muito importante para manter em funcionamento o projeto, enquanto seguimos amadurecendo a nossa infraestrutura de hospedagem do site, tópico que será abordado a seguir.

ORGANIZAÇÃO INSTITUCIONAL

A resposta da Fiocruz à epidemia se iniciou de forma orgânica, originada em grupos preexistentes ou que se formaram espontaneamente congregando profissionais e estudantes. À medida que aumentava o número de casos e óbitos, nos meses de fevereiro e março de 2020, algumas iniciativas foram surgindo nos diferentes institutos e laboratórios da fundação.

Visando a organizar e institucionalizar essas iniciativas, foi criado na Fiocruz o Observatório Covid-19: informação para ação. Após o reconhecimento de cada iniciativa por seu laboratório e instituto, os projetos foram cadastrados no Observatório Covid-19. Em sua página, hospedada no portal Fiocruz, a produção de cada iniciativa é divulgada por meio de notícias, *press releases*, notas técnicas, links para seminários e outros materiais. Essa foi uma forma acertada de apoiar e organizar as iniciativas existentes na instituição, evitando hierarquizações desnecessárias e sobreposições de trabalhos.

No âmbito do MonitoraCovid-19, a iniciativa foi rapidamente apoiada pelo Núcleo de Geoprocessamento e pela Plataforma de Ciência de Dados Aplicada à Saúde, ambos vinculados ao Laboratório de Informação em Saúde do Ictict. A estratégia de desenvolvimento colaborativo do MonitoraCovid-19, sempre aberto à participação de novas pessoas, de dentro e fora da Fiocruz, inclusive voluntários da sociedade civil, levou a discussões sobre seus objetivos e produtos, problematização sempre necessária ao fazer científico. No início do seu desenvolvimento, a interação do projeto com outros grupos de pesquisa dentro da Fiocruz foi agitada por sugestões vagas, ou mesmo perturbada por declarações de inutilidade do projeto. Com gradativo amadurecimento da iniciativa, a plataforma MonitoraCovid-19 passou a ser reconhecida com orgulho pela instituição e usada nas apresentações oficiais da Presidência da Fiocruz.

Para a continuidade sustentável do projeto, procuramos alguns editais de financiamento. Além de recursos financeiros pontuais obtidos com a Presidência da Fiocruz, o projeto foi contemplado pelo Edital Inova Fiocruz Ideias e Produtos Inovadores - Covid-19 - Encomendas Estratégicas, de 2020. Os editais Inova da Fiocruz visam a financiar projetos internos da instituição cujo objetivo geral seja incentivar ambientes favoráveis à pesquisa, desenvolvimento tecnológico e inovação em saúde em todas as áreas de atuação da instituição.

Concorrendo com mais de cem projetos, o MonitoraCovid-19 foi contemplado e recebeu integralmente o valor planejado na proposta. Esse recurso tem sido destinado principalmente ao pagamento de bolsas e à compra de equipamentos para a manutenção do projeto.

Além do Edital Inova, o projeto foi agraciado com uma premiação da Escola Nacional de Administração Pública (Enap), na categoria “Desafios Covid-19”, em reconhecimento ao seu mérito e qualidade. O recurso tem sido destinado ao pagamento de bolsas e também ao auxílio financeiro de projetos de parceiros essenciais para a continuidade do MonitoraCovid-19, bem como à melhoria de infraestrutura na unidade.

A RELAÇÃO COM A IMPRENSA

A publicação de notas técnicas e postagens em redes sociais sobre novos gráficos e análises logo atraiu a atenção de algumas pessoas, e pedidos de entrevistas e solicitações de esclarecimentos sobre a doença e evolução da pandemia começaram a surgir. As demandas da imprensa eram direcionadas à Assessoria de Comunicação (Ascom) do Ictict, que nos auxiliava na triagem e atendimento desses pedidos. Em outros momentos, éramos diretamente acionados por órgãos de imprensa para comentar determinado tema em pauta ou produzir indicadores sobre um problema de saúde, sempre usando a plataforma MonitoraCovid-19.

Sempre tivemos o entendimento de que o papel prioritário do MonitorCovid-19 era manter a população brasileira informada adequadamente sobre os dados de Covid-19, e nisso o relacionamento com a imprensa era algo essencial.

Em alguns momentos tivemos dificuldade em atender a todos os pedidos da imprensa, que demandava também dados específicos para certas regiões ou municípios. O atendimento a esse tipo de demanda era complexo e dividia nossa atenção entre manter o projeto sempre atualizado e atender às solicitações. Procuramos seguir considerando que o conhecimento gerado pelo projeto deveria ser divulgado e propor pautas à imprensa. Até 6 de janeiro de 2021, o portal de notícias do Google apresenta 215 reportagens em que o projeto é citado, em diversos veículos de imprensa, nacionais e internacionais.

OUTRAS FORMAS DE INTERAÇÃO COM A POPULAÇÃO

Já no início do projeto, quando a primeira versão foi disponibilizada ao público, uma conta de e-mail institucional foi criada para receber dúvidas, sugestões, críticas e pedidos de esclarecimentos dos usuários. Esse canal de comunicação direta com o projeto se mostrou bastante interessante, ainda que exaustivo em alguns momentos. Por ele recebemos demandas de diversos tipos e origens, de cidadãos comunicando imprecisões e pedindo esclarecimentos sobre seus municípios até entes públicos requisitando dados extraoficialmente. Procuramos responder às mensagens enviadas

com rapidez e precisão, mas sem prejudicar o andamento do projeto como um todo. Em alguns momentos, essas demandas por dados e análises específicas foram enviadas pelo Ministério Público.

O acesso ao projeto foi disponibilizado na internet no dia 23 de março de 2020. Dessa data até o dia 15 de dezembro de 2020, o site do MonitoraCovid-19 teve 211.778 usuários, que abriram as abas do projeto 501.727 vezes (sessões). Atendemos usuários em mais de cem países de todos os continentes, conforme os dados informados pela plataforma Google Analytics (Figura 5b).

Figura 5a – Dados de acesso ao site do projeto

Figura 5b – Dados de acesso ao site do projeto

No Brasil, atendemos usuários em todas as 27 UFs, o que se pode creditar, em parte, à divulgação do projeto por meio perfis da Fiocruz nas redes sociais.

Twitter

Conforme o MonitoraCovid-19 avançava com novas funções, bases de dados e gráficos, procurávamos uma forma de divulgar estas novidades. Um dos principais canais que utilizamos é o Twitter. Apesar de não termos uma conta exclusiva para o projeto, procuramos divulgar as novidades do projeto em nossas contas pessoais, já conectadas com diversos usuários e grupos. Em geral, essas postagens alcançam certa repercussão, com curtidas e reportagens por centenas de usuários dessa rede social. O Twitter possibilitou também a formação de redes de análise e a troca de informações entre equipes de especialistas de diversas áreas de conhecimento.

Um efeito interessante ocorreu quando o pesquisador e divulgador científico Atila Iamarino fez uma postagem em 3 de julho de 2020 sobre o crescimento de casos nas UFs, usando gráficos produzidos no MonitoraCovid-19, incluindo um link para o site do projeto. Essa postagem gerou grande número de usuários acessando o site do projeto em curto tempo, o que levou à instabilidade do sistema. Conforme a postagem repercutia e alcançava mais pessoas, fomos acumulando um número crescente de usuários. No dia seguinte, atingimos um pico de 3.487 usuários em um único dia.

O projeto também fornece seus dados para outros projetos e instituições. O IBGE tem um link direto de acesso aos nossos dados atualizados para alimentar seu portal sobre a Covid-19. Outra iniciativa de cessão de dados é destinada ao monitoramento da epidemia na Região Metropolitana de São Paulo.

INFRAESTRUTURA DE COMPUTAÇÃO E HOSPEDAGEM

O crescimento do projeto em termos do número de dados trabalhados, gráficos disponíveis e número de usuários nos deu melhor dimensão da responsabilidade ativa do projeto em informar a população, diariamente, sem atrasos ou interrupções.

Para isso, foi necessário rever a infraestrutura utilizada para a hospedagem do projeto. Desde o início, a Plataforma de Ciência de Dados aplicada à Saúde (PCDaS) tem apoiado o MonitoraCovid-19 com sua infraestrutura computacional e revisão técnica do sistema.

Inicialmente, o projeto era hospedado em um servidor da PCDaS localizado fisicamente no Laboratório Nacional de Computação Científica (LNCC), em Petrópolis, RJ. Apesar da alta qualidade ofertada pelo LNCC para computação científica, percebemos que a rede disponível não atendia a pré-requisitos que permitissem atender a grande número de acessos ao site do MonitoraCovid-19.

Com o crescimento do número de usuários, optamos, então, por hospedar o site do projeto dentro da sala cofre da Fiocruz, contando com a ajuda da equipe da PCDaS e do Centro de Tecnologia da Informação e Comunicação em Saúde (CTIC/Icict) para a migração e manutenção do sistema. Essa sala cofre é uma infraestrutura computacional criada na Fiocruz para prover a instituição de um ambiente computacional seguro, física e virtualmente, oferecendo serviços computacionais como hospedagem de sites, com alta confiabilidade e disponibilidade.

PRODUTOS CIENTÍFICOS DO PROJETO

Além das abas e gráficos apresentados disponíveis no site, o projeto produziu uma série de notas técnicas cujo teor detalhamos a seguir.

A dinâmica da doença demandou maior celeridade na divulgação das informações para que estas pudessem se tornar política pública. Se antes o caminho habitual passava pela construção de um artigo científico e sua publicação para que houvesse alguma visibilidade, durante a pandemia de Covid-19 esse caminho moroso não traria muito ganho no sentido de subsidiar decisões e intervenções. Nem mesmo a velocidade dos *preprints* possibilitava ganho de oportunidade. A decisão foi: precisamos analisar

os dados do sistema e emitir notas técnicas sobre os cenários da pandemia de forma rápida e independente.

A primeira nota técnica foi divulgada em 2 de abril de 2020 e alertava sobre o processo de interiorização da doença. A doença havia chegado pelas grandes cidades e não havia razão nenhuma para acreditar que não avançasse para as cidades do interior. Foi o que aconteceu. A parceria com o IBGE possibilitou acesso aos dados da pesquisa Regiões de Influência das Cidades (Regic, edição 2018), em especial do componente Saúde, que foi liberado para uso antes mesmo da sua publicação em julho de 2020. Com base nesses dados, estruturamos outras notas e apontamos o caminho esperado da difusão espacial do vírus.

Em maio de 2020, buscamos avaliar a velocidade de disseminação da doença; como já imaginávamos, a chegada da doença no interior do país e a historicamente reduzida disponibilidade de recursos para atendimento de alta complexidade (UTI) trariam um quadro de desassistência à saúde. Levantamos esses dados e estimamos a distância de deslocamento em busca de atendimento e a disponibilidade de recursos dentro das Regiões de Saúde e o fluxo de pacientes que já estava ocorrendo, de um município em busca de outro município que pudesse atendê-los. Os dados permitiram antever uma situação trágica que infelizmente se confirmou. Ainda em maio, em outra nota se avaliavam os dados de Srag do Sistema de Informação da Vigilância Epidemiológica da Gripe (Sivep Gripe) e como esse sistema permitiria avaliar aspectos como sexo, raça, idade, atendimento, entre outros, e se criava um espectro mais amplo para a avaliação da doença.

Em junho de 2020 começou a ocorrer um movimento, originado em setores da economia e da política, em favor da flexibilização das medidas de isolamento e distanciamento social. Esse movimento, junto com outros processos de desgaste de estratégias de proteção da saúde, resultou na demissão do ministro Mandetta em abril e do ministro Teich em maio. Nesse período nos dedicamos a trabalhar com dados de mobilidade urbana apontando os índices de mobilidade disponíveis no sistema. Também buscamos explicar que existiam tempos epidêmicos diferentes em função da chegada da doença. À medida que a doença chegava, os chamados epicentros da doença iam se alterando e uma flexibilização geral defendida pelo governo federal se mostrava uma decisão equivocada.

No mês de julho de 2020, passamos a avaliar os impactos indiretos, sobretudo os óbitos por outras causas e os indiretamente ocorridos pela Covid-19. Outros sistemas estavam disponibilizando os dados de cartórios do registro civil, informações que

chegamos a implementar no sistema, mas, em razão de inconsistências que observamos à época, bem como de distorções e *fake news* baseadas em tais informações, foi preciso retirar. Avaliamos então os padrões de mortalidade no município do Rio de Janeiro, que possui seu próprio site de disponibilização de dados, mostrando grande volume de óbitos em domicílios, vias públicas e outras unidades de saúde e sem assistência médica, o que poderia estar se repetindo em outras cidades.

Ainda em julho de 2020, a questão da volta às aulas ganhou força no processo de relaxamento das medidas de isolamento social. Nesse período tal medida se mostrava preocupante diante da alta da taxa de contágio e, principalmente, da exposição da população idosa e da população com fatores de risco que conviviam com pessoas em idade escolar. Fizemos um levantamento com a equipe que detém conhecimento sobre os dados da Pesquisa Nacional de Saúde e apontamos que quase 10 milhões de pessoas idosas ou com fatores de risco conviviam com crianças em idade escolar. Buscamos ser taxativos, mostrando que a retomada das aulas sem testagem, rastreio, medidas de distanciamento físico, de higiene e o uso de máscara seria o fim do isolamento social para essa população.

Os boletins epidemiológicos das secretarias estaduais de Saúde (SES) apresentaram, em alguns momentos, represamento de casos para além do esperado no comportamento cíclico das informações, em função dos fins de semana. Decidimos buscar os sistemas de referência (e-SUS-VE e Sivep Gripe) para comparar a data do evento (óbitos ou data dos primeiros sintomas nos casos) e a data de divulgação dos dados pelas SES. A diferença entre os cenários que a população acompanhava na mídia e aqueles baseados na data em que de fato ocorriam os eventos chegava, para alguns estados, a quase dois meses. Essa nota foi lançada em agosto de 2020. Em setembro, em um evento com outros grupos de análise sobre Covid-19 do Brasil, elaboramos uma carta aberta que buscava alertar a mídia e a sociedade sobre essa situação. Concordamos que não era um processo simples e que, como já havia o hábito da divulgação dos dados pelos boletins, isso não deveria ser alterado, mas complementado. A divulgação continuou sendo pautada pela mídia com os dados dos boletins das SES e continuamos sujeitos a esse represamento, que se acentuou próximo ao período eleitoral e nos recessos de fim de ano.

Em dezembro de 2020 retomamos a análise sobre o excesso de óbitos no município do Rio de Janeiro, e os números revelaram um cenário trágico de desassistência na cidade, que com certeza se repetiu em muitas outras cidades. O aumento de casos no fim do ano e maior mobilidade das pessoas motivaram a nota técnica que apontava o fim do processo de interiorização e a sincronização da pandemia no espaço. O estudo mostra duas fases na dinâmica espacotemporal da Covid-19: se em abril e maio havia

um processo de difusão da doença, das capitais para cidades de menor porte, e aumento de casos à medida que a doença avançava no território, em novembro a doença ocupava todo o território e a movimentação das pessoas entre cidades definiria o aumento de casos de forma sincronizada em vários locais.

MODELOS DE PREVISÃO

Conforme os casos avançavam no Brasil, em meados de março de 2020, e os primeiros óbitos começavam a ser reportados, diversas propostas nacionais e internacionais surgiram para tentar prever quando ocorreria o auge do número de casos, o chamado pico da curva, e qual seria o número de casos nessa data. Essas informações seriam de grande importância para preparar o sistema de saúde, com a preocupação dominante à época de achatar a curva, para melhor atender a população.

Nessas propostas se utilizaram, em geral, métodos de modelagem epidemiológico compartimentais, como os modelos SIR (Suscetível - Infectado - Recuperado) e algumas variantes.

No âmbito do MonitoraCovid-19, inicialmente procuramos desenvolver modelos de previsão e chegamos a discutir alguns modelos de previsão a curto prazo considerando o crescimento exponencial da doença. Contudo, no decorrer do processo epidêmico, observamos que a epidemia era muito mais complexa do que se esperava e vários dos modelos propostos inicialmente começaram a ser descartados, pois a incerteza, nos modelos, para efeitos de médio e longo prazos traziam enorme insegurança, devido à falta de conhecimento sobre a doença.

Além disso, enfrentamos – como todos os outros proponentes de modelos – grandes dificuldades com a qualidade e as limitações dos dados. Os dados disponíveis sobre a epidemia versavam basicamente sobre o número de casos e óbitos por data de notificação, o que acaba por violar alguns dos pressupostos do modelo. Seria necessário obter a data dos primeiros sintomas dos casos, conhecer precisamente o número de testes aplicados nas populações, detalhes sobre as medidas de prevenção e combate à epidemia e outros fatores que não estavam (e não estão até o momento) disponíveis como dados. Por esses motivos, decidimos não apresentar um modelo publicamente. Priorizamos o trabalho em análises de curto prazo, sobretudo sobre o processo de disseminação espacial da doença e as implicações relacionadas ao atendimento.

De fato, observamos que diversos dos modelos apresentados publicamente foram discretamente retirados, conforme a epidemia avançava. Em geral, eles eram reajustados diariamente com os novos dados, o que acabava por sempre postergar o pico de casos.

Os valores resultantes desses modelos não pareciam ser compatíveis com a realidade do momento, ou necessitavam de dezenas de parâmetros adicionais, fornecidos pelo usuário para modelar a curva epidêmica, o que virtualmente impedia qualquer previsão. Com efeito, nenhum modelo de previsão alertou para a possibilidade futura de um segundo pico de casos, realidade que começou a se definir em novembro de 2020.

A ESTRUTURA DE CÓDIGO

A estrutura de código de programação do projeto se divide em duas partes: Extraction, Transform and Load (ETL) e o aplicativo web.

ETL

A estrutura de ETL é responsável pela atualização rotineira dos dados do projeto. Atualmente ela é realizada de forma automática em um servidor e seus passos são notificados via mensagens no Telegram para a equipe técnica do projeto. A execução do *script* de ETL leva, atualmente, em torno de uma hora para ser executada.

Esta estrutura é composta de uma série de *scripts* em linguagem R. O primeiro bloco de *scripts* realiza o *download* dos dados utilizados no projeto. Outros *scripts* efetuam transformações e cálculos necessários, enquanto o último bloco de *scripts* realiza exportações de dados para que projetos parceiros possam consultar nossos dados.

Aplicativo

O site do projeto é um aplicativo também desenvolvido com a linguagem R, utilizando-se o pacote Shiny para construção de aplicações on-line. Trata-se de um arquivo único contendo em torno de 4.000 linhas, onde são definidas a estrutura visual do site e a execução de códigos dinâmicos para a construção das tabelas, gráficos e mapas apresentados no site.

AVANÇOS E DESAFIOS FUTUROS

No projeto MonitoraCovid-19 pretende-se continuar cumprindo a missão institucional do Ict, disponibilizando informação de forma aberta e criando interfaces para consultas interativas e aquisição dos dados de forma simples e acessível para a população, os gestores e a sociedade civil.

A curto prazo, o sistema passará por uma remodelagem de *layout* e usabilidade com a qual se pretende tornar a plataforma ainda mais simples e intuitiva e, portanto, acessível a

um público mais amplo. Para tanto, esperamos incluir dados sobre Covid-19 dos demais sistemas de informação de saúde (SIM, SIH, SIA etc.) que devem disponibilizar esses dados seguindo uma defasagem inerente a cada sistema de informação, mas trazem outras possibilidades de análise e entendimento da doença e do processo epidêmico.

As bases de dados do Sivep Gripe e do sistema e-SUS VE serão trabalhadas para criar interfaces de consultas amigáveis e acessíveis a qualquer usuário, sem necessidade de conhecimento especializado sobre modelagem de bancos de dados. Além disso, esperamos incluir os dados sobre vacinação disponíveis tanto no Brasil quanto no mundo.

Esperamos expandir as parcerias e fornecer informação para estudos aprofundados sobre a Covid-19 que nos ajudem a entender a epidemia, para enfrentá-la e criar conhecimento de base para eventuais novos processos endêmicos que venham a ocorrer. Nesse sentido, espera-se maior aproximação das instituições de ensino e das secretarias de Saúde para colaboração na melhoria da informação e na capacitação de pessoal para análise.

REFERÊNCIAS

COTA, W. Repositório do GitHub-Covid19Br, 2020. Disponível em: <<https://github.com/wcota/covid19br>>. Acesso em: dez. 2020.

FONTES, R. Coronavirus – Brazil; Brazil Dataset, 2020. Disponível em: <www.kaggle.com/unanimad/corona-virus-brazil>. Acesso em: dez. 2020.

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE/FUNDAÇÃO OSWALDO CRUZ (ICICT/FIOCRUZ). Site. Disponível em: <www.iciet.fiocruz.br/content/sobre-o-iciet>. Acesso em: 20 dez. 2020. Acesso em: dez. 2020.

PROVOST, F. & FAWCETT, T. Data Science and its relationship to big data and data-driven decision making. *Big Data*, 1(1): 51-59, 2013. Disponível em: <<http://doi.org/10.1089/big.2013.1508>>. Acesso em: dez. 2020.

WAZE. Mantendo o Waze atualizado durante a pandemia Covid-19, 2020. Disponível em: <www.waze.com/pt-BR/covid19>. Acesso em: dez. 2020.