

## E se o *scan*COVID-19 tivesse nascido em 2022?

And if the *scan*COVID-19 had been born in 2002?

Y si el *scan*COVID-19 hubiese nacido en 2022?

Maria Cristina Soares Guimarães<sup>1,a</sup>

[crisrina.guimaraes@icict.fiocruz.br](mailto:crisrina.guimaraes@icict.fiocruz.br) | <https://orcid.org/0000-0003-2717-381X>

Janio Gustavo Barbosa<sup>1,b</sup>

[janio.barbosa@icict.fiocruz.br](mailto:janio.barbosa@icict.fiocruz.br) | <https://orcid.org/0000-0001-5829-760X>

<sup>1</sup> Fundação Oswaldo Cruz, Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Programa de Pós-graduação em Informação e Comunicação em Saúde. Rio de Janeiro, RJ, Brasil.

<sup>a</sup> Doutorado em Ciência da Informação pela Universidade Federal do Rio de Janeiro.

<sup>b</sup> Mestrado em História pela Universidade Federal do Rio Grande do Norte.

### RESUMO

Acompanhar a velocidade das mudanças, qualitativas e quantitativas, na produção de conhecimento em tempos de pandemia de covid-19 impulsionou o desenvolvimento de inúmeras iniciativas de monitoramento da informação científica. O *scan*COVID-19 foi uma delas. Essa nota de conjuntura procura situar a importância do investimento em projetos dessa natureza, que possam ter reflexos nas relações entre ciência, Estado e sociedade.

**Palavras-chave:** Monitoramento de informação; Covid-19; Ciência aberta; Ciência de dados; *scan*COVID-19.

### ABSTRACT

Keeping up with the speed of qualitative and quantitative changes in the production of knowledge in times of the covid-19 pandemic was a stimulus to the development of numerous initiatives to track scientific information. The *scan*COVID-19 was one of them. This note seeks to situate the importance of investing in projects of the same nature, which may have effects on the relations between science, state, and society.

**Keywords:** Information tracking; Covid-19; Open science; Data science, *scan*COVID-19.

## RESUMEN

Acompanhar la velocidad de los cambios, cualitativos y cuantitativos, en la producción de conocimiento en tiempos de la pandemia de covid-19 ha impulsado el desarrollo de numerosas iniciativas de seguimiento de la información científica. El *scanCOVID-19* fue uno de ellos. Esta nota de coyuntura busca situar la importancia de invertir en proyectos de esta naturaleza, que puedan tener efectos sobre las relaciones entre ciencia, Estado y sociedad.

**Palabras clave:** Seguimiento de la información; Covid-19; Ciencia abierta; Ciencia de datos; *scanCOVID-19*.

---

## INFORMAÇÕES DO ARTIGO

### Contribuição dos autores:

Concepção e desenho do estudo: ambos os autores.  
Aquisição, análise ou interpretação dos dados: ambos os autores.  
Redação do manuscrito: ambos os autores.  
Revisão crítica do conteúdo intelectual: ambos os autores.

**Declaração de conflito de interesses:** não há.

**Fontes de financiamento:** não há.

**Considerações éticas:** não há.

**Agradecimentos/Contribuições adicionais:** Agradecemos a leitura cuidadosa e contribuições de Rosane Abdala, coordenadora institucional do *scanCOVID-19*.

**Histórico do artigo:** submetido: 14 fev. 2022 | aceito: 14 fev. 2022 | publicado: 25 fev. 2022.

**Apresentação anterior:** apresentação oral, institucional, em 2021.

**Licença CC BY-NC atribuição não comercial.** Com essa licença é permitido acessar, baixar (*download*), copiar, imprimir, compartilhar, reutilizar e distribuir os artigos, desde que para uso não comercial e com a citação da fonte, conferindo os devidos créditos de autoria e menção à Reciis. Nesses casos, nenhuma permissão é necessária por parte dos autores ou dos editores.

A história da ciência nos lembra que os maiores avanços na base do conhecimento, majoritariamente indutivos, se deram em momentos de crise e, portanto, se caracterizam como disruptivos. Por mais chocante e paradoxal que possa parecer, as guerras são profícuas em demarcar recortes temporais de mudanças de paradigma no contexto científico e nas relações entre ciência, Estado e sociedade (não necessariamente para melhor, infelizmente). A ciência pode causar destruição, mas ela pode, fundamentalmente, salvar, resgatar e contribuir para a construção de novos cenários da humanidade. “Destruição criativa”, como profetizou Joseph Shumpeter, teórico da inovação (SCHUMPETER, 1950) considerado um dos principais economistas do século XX. No atual contexto pandêmico não é diferente.

A busca de soluções para enfrentamento da atual emergência sanitária global provocada pela covid-19, que eclodiu nos primeiros meses do ano de 2020, imprimiu uma dinâmica diferenciada no regime de produção de conhecimento, quantitativa e qualitativamente. Os principais marcadores dessa dinâmica são, por um lado, o volume brutal de informação científica produzida, em diferentes tipologias e de natureza variada que, diariamente, vem a público em diferentes fontes, mais, ou menos, credenciadas. Essa produção acelerada pressionou, como uma avalanche, os meios tradicionais de difusão do conhecimento. Por outro lado, qualitativamente, as colaborações científicas se tornaram quase que regra, as pesquisas inter e multidisciplinares se impuseram e, talvez mais importante, a autonomia dos pesquisadores/autores foi, de certa forma, redimensionada. Para alguns, a confiança na ciência, nos cientistas e entre os cientistas vem sendo, aos poucos, restaurada, senão redimensionada.

Mas mudanças disruptivas não prescindem de um ‘chão de fábrica’, de um trampolim, de onde se pode dar o salto qualitativo. Para nós, a ciência aberta e a ciência de dados são os dois pilares que sustentam e fomentam o processo criativo de soluções e alternativas inovadoras que vêm promovendo a ampliação do acesso ao conhecimento científico, em suas mais diversas dimensões, conteúdos, tipologias e formatos. Assim foi com o [scanCOVID-19](#).

Enquanto projeto de pesquisa, o *scanCOVID-19* nasceu no âmbito do Programa de Pós-graduação em Informação e Comunicação em Saúde (PPGICS/Icict/Fiocruz), em meados de 2020, imerso na incerteza que orientava a própria dinâmica da ciência. Como auxiliar pesquisadores, gestores e a sociedade a acompanhar o que estava acontecendo nas diferentes frentes de pesquisa? Onde coletar os dados? Como coletá-los? Como organizá-los? A resposta era simples: usar as tecnologias a nosso favor. Resposta simples para uma pesquisadora em ciência aberta, mas longe do trivial para um cientista de dados. O compromisso: usar fontes abertas. A primeira e mais radical aposta foi nos repositórios de *preprints* os quais, na época, ainda nutriam uma certa desconfiança na própria comunidade científica. E, para nossa sorte, eles se transformaram na vedete inusitada da ciência pandêmica!

Chegamos em 2022 com uma proliferação dos repositórios de *preprints*, virtualmente todos com curadoria, e com estratégias variadas de *rapid review*, de retirada e retratação de artigos, e incorporação de mídias sociais. Algumas grandes casas editoriais, por outro lado, quando não passaram a investir elas mesmas nesta estratégia de abertura, passaram a aceitar (com algumas restrições) a publicação dos *preprints* como artigos originais o que, atualmente, é facilmente rastreável por meio do uso de tecnologias disponíveis (a organização dos repositórios por meio de metadados em muito facilita esse rastreamento). Há um movimento inequívoco de fortalecimento dos repositórios de *preprints* por meio de estratégias de seus próprios gestores o que, como anteriormente apontado, agora são conduzidos por gestão profissional. Ou seja, de certa forma, os *preprints* vão se incorporando a lógica da economia da informação: é um *business* valioso demais para ser deixado na mão dos pesquisadores. Alguns resistem, como o [arXiv](#), mas é uma exceção.

Todo esse movimento joga uma outra luz, pelo menos por ora, no conceito de autoria científica e no que se configura como uma ‘inovação’, determinante para um artigo ser publicado como ‘original’ por um

periódico científico. De fato, a ciência aberta se insinua e tensiona diferentes dimensões do fazer científico, especialmente nas entranhas ortodoxas das diversas comunidades científicas. Cada área do conhecimento, por certo, vai responder de forma diferente. Mas provavelmente nenhuma vai passar incólume pela pandemia.

São, na atualidade, inúmeros os serviços de monitoramento/*tracking* do avanço científico em tempos pandêmicos (CHEN *et al.*, 2021). Exemplos podem ser citados: o [iSearch COVID-19](#) do National Institutes of Health (NIH), uma base curada que inclui artigos da PubMed e de vários repositórios de *preprints* sobre SARS-CoV-2 e covid-19. Por seu lado, o [LitCovid](#) tem seu foco unicamente nos artigos científicos publicados na PubMed, curados e categorizados por temáticas (mecanismo, transmissão, diagnóstico, tratamento, prevenção, case report e forecasting). O Worldwide Non-pharmaceutical Interventions Tracker for COVID-19 (WNTRAC) monitora, coleta e disponibiliza as intervenções não farmacêuticas para enfrentamento da covid-19 implementadas em cerca de 160 países, desde o início da pandemia. Acompanhar o financiamento de projetos de pesquisa sobre covid-19, em diferentes áreas do conhecimento, é possível pelo acesso ao [COVID-19 Research Project Tracker](#).

Todos esses serviços nasceram da recente ideia de *'tracker'* de fontes e dados, só possível com a ciência de dados e sua capacidade de tratar grandes agregados de dados. Esse tipo de ação (de monitoramento) tem sido cada vez mais importante. Recentemente, os grandes centros de tratamento de informação estratégica e as grandes corporações privadas têm investido esforços no *'eye tracking'* (segunda geração do *tracker*) que coloca o foco no usuário e na sua interação e comportamento de navegação na web.

Tecnologias de acompanhamento (*tracking*) são possíveis pelo conceito de ecologia da informação (BAKER; BOWKER, 2007) por meio da qual usuários, contexto e conteúdo compõem um todo de diferentes complexidades no uso dos dados e no tratamento da informação. Essa dinâmica tem aumentado, exponencialmente, nos últimos anos, globalmente. O número crescente de usuários; o aumento substantivo da velocidade das conexões com advento da tecnologia 5G, e o massivo desenvolvimento de ferramentas de captação de dados vêm tornando a internet uma dimensão com tamanha informação, que a qualificação de fontes e o escaneamento da produção, *in loco*, têm se mostrado uma estratégia necessária e eficiente em face das atuais demandas estratégicas e competitivas que emergiram dentro próprio campo da ciência.

Todas essas mudanças não passam despercebidas das grandes casas editoriais e dos gestores dos repositórios de *preprints*, os quais têm lançado mão dessas tecnologias, além da inteligência artificial (AI) não só para identificar os documentos de maior audiência, como também para identificar e remover artigos que contêm plágios, e/ou trabalhos experimentais que não atendam aos padrões exigidos de ética, sobretudo, nos relatos que envolvam experimentos com animais e seres humanos. Todos esses arranjos são filhos do tempo pandêmico, e da urgência do mundo contemporâneo.

O *scanCOVID-19* utiliza as duas principais técnicas que envolvem o *tracking* de dados: a primeira, de leitura e coleta de dados; a segunda, o uso de protocolos de interação entre bases de dados, por meio de linguagem de programação, por meio de requisições e respostas computacionais ao que se denomina API (*Application Programming Interface*). Tecnologias que realizam *crawler* (coleta) e *webscrapping* (raspagem de dados) apostam que a captura em larga escala de dados para disponibilização simplificada e centralizada democratiza o conhecimento e promovem maior acessibilidade às abordagens metodológicas atualmente utilizadas no fazer da pesquisa acadêmica. Dessa forma, a dinamicidade e a velocidade da divulgação das produções passam a não ser mais um problema, a questão é como articular as várias proposições diante da emergência de respostas sociais de que necessitamos.

Tal como todos os dispositivos de *tracking*, o *scanCOVID-19* comporta-se como um robô que, ensinando sobre como navegar numa página, reconhece quais as características necessárias para uma busca, de modo a coletar os dados de maneira precisa. Essa coleta nada mais é que o *'copiar e colar'* que se faz

costumeiramente, com a diferença de ser feito em milissegundos, com recorrência automática. O grande diferencial do *scanCOVID-19* é a diversidade de fontes monitoradas, respeitando a especificidade de cada uma delas. *Preprints*, artigos de periódicos, ensaios clínicos, protocolos Conep (Comissão Nacional de Ética em Pesquisa), em um mesmo ambiente *web*. Para nós, mais do que o escaneamento das fontes, nos parecia importante acompanhar como os ‘problemas’ a serem enfrentados se pulverizavam/descentralizam entre diferentes estratégias de ação/tipologias de publicações.

Os desafios do *scanCOVID-19*, no ano de 2020, não foram poucos, ou menores. O primeiro deles diz respeito às fontes monitoradas que, a despeito de serem ditas abertas, estão em constante mudança, o que pedia/pede ajustes quase diários no robô. Além disso, outro desafio foi que os grandes portais de periódicos constantemente rastreavam e bloqueavam as ações desse tipo de robô. Um terceiro se refere ao fato de se tratar de uma questão ética, de comprometimento, observação e vigilância de coleta apenas de dados abertos e públicos, ou seja, artigos que não afetassem os direitos autorais e/ou morais de terceiros. O quarto desafio tem a ver com a continuidade e a reprodutibilidade dessa tecnologia para ser empregada em novas fontes, adições de novos parceiros e novos temas de pesquisas. A tecnologia precisava estar disponível para ser atualizada e alterada a qualquer momento para outros fins e, por esse motivo, deveria prever seu uso misto e apresentar natureza híbrida.

Considerando esses desafios, nos questionamos: e se fizéssemos o *scanCOVID-19* em 2022? Certamente o seu desenvolvimento encontraria maiores facilidades pelo ciclo dinâmico dos processos de inovação proporcionados pela abertura da ciência e do uso das abordagens da ciência de dados. Por exemplo, devido ao crescente comportamento de rastreio de páginas e monitoramento de sites, as contramedidas que bloqueavam robôs, realizadas pelos repositórios e páginas governamentais oficiais, não se mostraram mais efetivas. Os periódicos e organizações governamentais adotaram o mantra “se não pode com o inimigo, alia-se a ele”. Passaram então a disponibilizar arquivos .xml (arquivos de textos com algumas *tags*), com os resumos das produções diárias e metadados qualificados a partir da seleção que o usuário, único ou automatizado, fizer. Entretanto, não significa que seja possível acessar todo o conteúdo a custo zero. Algumas taxas para acesso já estão sendo precificadas.

A covid-19 trouxe também para a prática científica uma aproximação inequívoca com as redes sociais. Os periódicos e os repositórios de *preprints* passaram a ter interseções em muito ampliadas com mídias sociais abertas, o que os permite ‘calibrar’ o reconhecimento de autores e centros de pesquisa envolvidos com o desenvolvimento de soluções para o enfrentamento da pandemia. Vídeos, aulas *online*, participação em webnários e eventos correlatos, entrevistas em várias mídias, dentre outras, foram algumas das inúmeras estratégias que, orientadas ou não, colocaram foco na ciência pandêmica. Ainda é cedo para antecipar se essa forma de democratização vai trazer alguma consequência para os processos avaliativos das atividades científicas, locais e globais, mas, certamente, são dimensões que devem passar a compor, no futuro próximo, os estudos altimétricos. Precisamos igualmente de respostas sobre como cobrir todo o universo web pandêmico, qualificando as fontes, separando o joio do trigo.

Tantas inovações mostram o quanto esse acontecimento protagonizado pela covid-19 vem ressignificando a relação entre ciência e sociedade. Os métodos de difusão da informação passaram a ser híbridos e os instrumentos de produção da informação científica mais flexíveis, uma vez que a dinamicidade da ciência começou a trabalhar com um elemento essencial para o estabelecimento do saber: o tempo. A urgência da informação científica é, notadamente e temporalmente, diversa do que ocorria outrora. Vacinas produzidas em tempo recorde, ensaios clínicos estabelecidos num menor espaço de tempo, agências sanitárias com novos e encurtados protocolos de autorização do uso de medicamento e vacinas, audiências públicas sobre os temas que envolvem a covid-19 com grande audiência midiática são eventos que, sem dúvida, vêm

redimensionando a relação entre ciência e sociedade. Talvez, nunca como antes, a sociedade, em face da incerteza em que estamos imersos, esteja procurando amparo na ciência. Parte dela, pelo menos.

Muito ainda pode ser feito, e deve ser feito, no casamento virtuoso entre ciência aberta e ciência de dados. Muito do que aprendemos até aqui é que, com investimento e comprometimento de atores-chave, podemos mais e melhor. As demandas e os desafios sociais estão postos. Cabe não desistir, e insistir.

## REFERÊNCIAS

BAKER, Karen S.; BOWKER, Geoffrey C. Information ecology: open system environment for data, memories, and knowing. **Journal of Intelligent Information Systems**, [s. l.], v. 29, n. 1, p.127-144, 2007. Disponível em: <https://link.springer.com/article/10.1007/s10844-006-0035-7>. Acesso em: 17 fev. 2022.

CHEN, Qingyu *et al.* Artificial intelligence in action: addressing the covid-19 pandemic with natural language processing. **Annual Review of Biomedical Data Science**, [s. l.], v.4, p. 313-339, 2021. DOI: <https://doi.org/10.1146/annurev-biodatasci-021821-061045>. Disponível em: <https://www.annualreviews.org/doi/abs/10.1146/annurev-biodatasci-021821-061045>. Acesso em: 17 fev. 2022.

SCHUMPETER, Joseph A. **Capitalism, socialism, and democracy**. 3. ed. Nova York: Harper and Brothers, 1950.