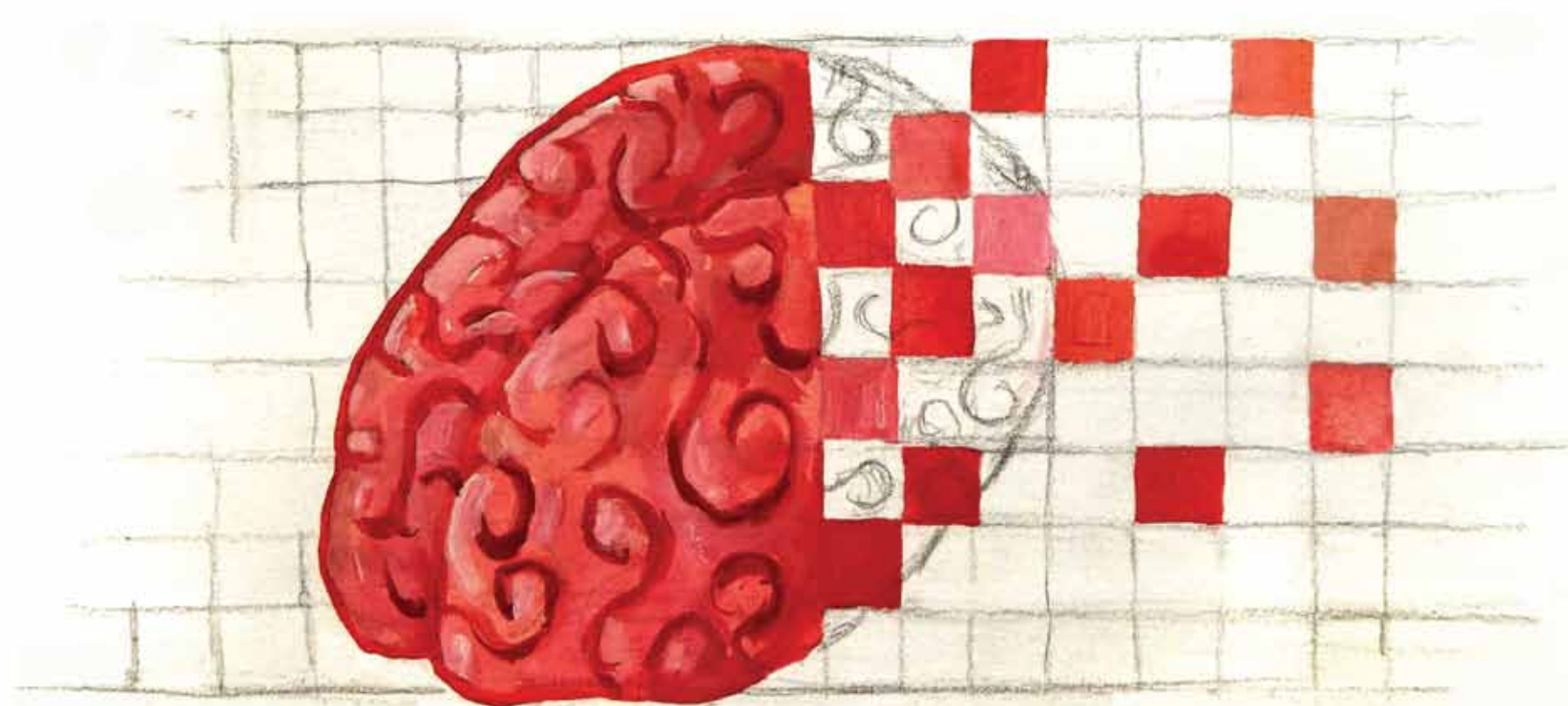


CIÊNCIA ABERTA

# Em defesa do compartilhamento público de dados científicos

O movimento da ciência aberta preconiza que as ferramentas e os dados utilizados pelos cientistas sejam disponibilizados publicamente para ampliar seus benefícios a toda a sociedade. Embora haja inegáveis avanços nos últimos anos, ainda há forças que resistem a essa ideia

POR CLAUDIA DOMINGUES VARGAS E FABIO KON\*



© Odyr

**A** ideia de que o conhecimento deveria estar ao alcance de todos aqueles que quisessem apreciá-lo é uma questão recorrente na história da humanidade. Ela está presente desde os filósofos gregos da Antiguidade até os cientistas da Renascença, dos trovadores medievais aos grandes compositores eruditos do século XIX. Ao mesmo tempo, mecanismos de controle do conhecimento, proteção da informação e até criptografia existem há algumas centenas de anos. No século XX, houve um movimento na direção de restringir o acesso ao conhecimento como forma de gerar receita financeira ou vantagem comercial. Desse modo, apenas quem pagasse pelo direito de executar uma peça musical receberia autorização para sentar-se ao piano em um concerto público e somente quem pagasse pelo acesso a um artigo científico teria direito de lê-lo. Uma parte significativa dos avanços científicos do século foi motivada por fins militares, cuja prática de esconder as descobertas

do inimigo é compreensível. No entanto, o que podemos observar a longo prazo é que, de maneira geral, quando há um nível maior de compartilhamento de ideias e abertura do conhecimento, o avanço da ciência é mais rápido e as sociedades tornam-se mais avançadas, ricas e democráticas.

Nos últimos anos, boa parte da comunidade científica internacional, com apoio de órgãos de fomento governamentais tais como a Fundação da Ciência Nacional (NSF) dos Estados Unidos, a Comissão Europeia e a Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp), no caso do Brasil, tem defendido o que se conhece como “ciência aberta”. Esse novo modelo de compartilhamento da informação científica está baseado em três fundamentos. Primeiro, os resultados científicos devem ser divulgados em veículos de “acesso aberto”, para que qualquer cientista ou cidadão tenha fácil acesso a essas descobertas, independentemente de sua origem ou situação financeira. Segun-

do, as ferramentas utilizadas no processo científico devem também ser compartilhadas abertamente; como boa parte da ciência hoje em dia depende de ferramentas computacionais, isso indica que elas devem ser disponibilizadas como “software livre”. Finalmente, os dados utilizados nas pesquisas devem ser compartilhados como “dados abertos”: não só os dados brutos e processados devem ser disponibilizados abertamente, mas também descrições do formato e significado desses dados (chamados metadados) devem ser distribuídas publicamente. Quando essas informações são coletadas em seres humanos, especial atenção é necessária para que a distribuição seja feita de forma criteriosa, respeitando a privacidade e o anonimato dos envolvidos. Um dos pilares da ciência experimental é sua reprodutibilidade. E a ciência só se torna reprodutível se os dados e ferramentas utilizados nos experimentos, simulações e análises forem também disponibilizados de forma aberta e livre.

A ideia de ciência aberta tem avançado em diferentes velocidades em distintos ramos. Em áreas como Ciência da Computação, Genética e Química, por exemplo, esses conceitos têm sido muito bem recebidos. Na Neurociência, o compartilhamento não é a praxe. Tanto a coleta como o armazenamento de dados ainda se fazem, de modo geral, de maneira artesanal. Há grande variabilidade nos tipos de dado que se coletam. Bancos de dados nessa área do conhecimento deverão contemplar desde informações sobre a forma e o comportamento de neurônios individuais, passando por medidas de funcionamento cerebral, até medidas comportamentais. Essa grande quantidade e variedade de informações requer um tipo de banco de dados desenhado especialmente para tal. Além disso, há, ainda hoje, grande desinformação na comunidade sobre mecanismos públicos de compartilhamento de dados.

## O CASO DA NEUROCIÊNCIA

A construção, manutenção e curadoria de bancos de dados públicos são consideradas fundamentais por muitos membros da comunidade neurocientífica para que se possa avançar mais efetivamente na compreensão do funcionamento e no tratamento de patologias do cérebro. O novo paradigma de compartilhamento de dados surgiu de maneira mais sistemática na literatura neurocientífica a partir da década de 1990. Data também dessa época a primeira grande iniciativa de compartilhamento de dados coletados a partir de medidas de ressonância magnética funcional, o Consórcio Internacional de Mapeamento do Cérebro. Esse trabalho foi financiado pelas agências de fomento governamentais norte-americanas NSF e Institutos Nacionais de Saúde (NIH). O novo modelo acompanhava o aumento substancial da capacidade de geração de dados experimentais nas Neurociências e as novas possibilidades computacionais e de compartilhamento público de informações, decorrentes do grande avanço ocorrido nas últimas décadas em tecnologia da informação.

Apesar dos grandes avanços na concepção e introdução de bancos de dados públicos, seu compartilhamento não era então um consenso entre os neurocientistas. Em 2000, a *Revista de Neurociência Cognitiva*, publicada nos Estados Unidos, determinou que os artigos aceitos para publicação deveriam compartilhar seus dados brutos em bancos de dados públicos. Essa política de compartilhamento estimulou postura semelhante em outras publicações de grande circulação. Entretanto, por pressão da comunidade de neurocientistas, essa proposta de compartilhamento público foi revogada. Felizmente, esse primeiro conjunto de iniciativas prenunciava uma nova era e, desde então, várias iniciativas de compartilhamento de dados têm sido colocadas em prática, seja no formato de consórcios, como no caso da Rede de Pesquisa do Cérebro e Neurociência Integrada (BrainNet), seja em projetos públicos, como Análise de Código, Repositório e Modelagem para a Neurociência Eletrônica (Carmen) e Ontologias Eletromagnéticas Neurais (Nemo). Um exemplo interessante de compartilhamento público de dados clínicos é o banco de dados de pacientes com Parkinson, coordenado pela Fundação Michael Fox sobre a Doença de Parkinson. Essa bela iniciativa ilustra o fato de que cada vez mais se reconhece a necessidade de bancos de dados públicos para que se possa avançar na identificação de marcadores precoces de patologias do cérebro.

Em todos os exemplos citados, é necessário realizar um cadastro no site que abriga o banco de dados e assi-

nar um termo de responsabilidade quanto à privacidade dos indivíduos cujos dados se encontram disponibilizados. A quebra do compromisso firmado no ato do cadastro pode ter consequências legais. Solicita-se também frequentemente que a origem dos dados e os artigos nos quais eles foram publicados sejam citados na nova publicação. Em alguns consórcios, o artigo deve ser submetido ao comitê científico que gerencia o banco de dados. Em certos casos, o pesquisador tem a opção de depositar os dados no banco sem disponibilizá-los publicamente, decidindo o melhor momento para o compartilhamento.

Stephen Koslow está entre os defensores mais ferrenhos do compartilhamento público de dados. Então diretor da Divisão de Neurociências do Instituto Nacional de Saúde Mental (NIMH) e um dos fundadores do consórcio BrainNet, Koslow publicou em 2000 um manifesto na revista *Nature Neuroscience*, de grande circulação e alto impacto na comunidade de neurocientistas, defendendo a necessidade de uma mentalidade voltada ao compartilhamento público de dados e ferramentas.

## O novo paradigma de compartilhamento de dados surgiu de maneira mais sistemática na literatura neurocientífica a partir da década de 1990

Entre as reações negativas mais comuns à ideia dos bancos de dados públicos, Koslow destacou os argumentos de que os dados brutos são muito complexos para serem compreendidos por outros neurocientistas e que a análise dos dados realizada por outra pessoa poderia levar a resultados diferentes dos originais. Outros argumentos contra o compartilhamento são, por exemplo, a resistência em tornar públicos dados muitas vezes duramente coletados, ou ainda a ausência de mecanismos legais de proteção no caso de fraude ou uso indevido das informações. Além disso, há críticas quanto a alguns dos modelos vigentes de compartilhamento de dados, nos quais, talvez por falta de uma curadoria, tanto a origem como a qualidade do dado disponibilizado são questionáveis.

Contrapondo-se a essas premissas, Koslow argumentou que é desejável e necessário que os dados sejam corretamente comentados e etiquetados, para que sejam compreendidos e utilizados por outros pesquisadores. Além disso,

ele argumenta que a publicação dos resultados na forma de artigos científicos pressupõe que os dados já estejam prontos para ser compartilhados e que perspectivas complementares produzidas por novas análises dos dados poderão ajudar a comunidade a compreender melhor o fenômeno em questão.

Assim, Koslow concluiu que os benefícios científicos do compartilhamento dos dados superavam os argumentos contra o compartilhamento e levantou algumas estratégias para que essa prática possa ser adotada mais amplamente na comunidade científica. Por exemplo, a já citada política de compartilhamento de dados adotada no passado por algumas revistas de grande circulação poderia estimular os pesquisadores a compartilhar seus dados.

O apoio financeiro adicional para a construção de bancos de dados em projetos financiados com recursos públicos e a valorização acadêmica do investimento em tempo e recursos alocados na construção dos bancos de dados também são mecanismos de mudança cultural propostos pelo autor. Essas estratégias poderiam levar a um ambiente em que os dados seriam organizados de modo a serem compartilhados durante o próprio processo de aquisição, não apenas no final desse processo. Críticos poderiam objetar que isso levaria a um custo operacional. Para enfrentar essa dificuldade, faz-se necessário o desenvolvimento de tecnologias de baixo custo para compartilhamento, manutenção e curadoria dos dados.

### O BANCO DE DADOS NEUROMAT

Participamos atualmente no Brasil do desenvolvimento de um banco de dados que permitirá o acesso público a dados de Neurociências (medidas fisiológicas e avaliações funcionais). Trata-se de um trabalho pioneiro desenvolvido pelo Centro de Pesquisa, Inovação e Disseminação em Neurociência (Cepid-NeuroMat), coordenado por Antonio Galves e financiado pela Fapesp. O projeto, que envolve principalmente pesquisadores da Universidade Federal do Rio de Janeiro (UFRJ) e da Universidade de São Paulo (USP), pretende construir um repositório público que permita progressos na compreensão do funcionamento cerebral, assim como no tratamento de doenças neurológicas.

Entre as linhas de pesquisa cujos dados serão hospedados no banco de dados NeuroMat está o projeto de reorganização cortical após lesão e reconstrução do plexo braquial, o conjunto de nervos que conecta o braço ao cérebro, em andamento hoje no Instituto de Neurologia Deolindo Couto (INDC) da UFRJ. A fim de albergar esse e outros projetos com perfil básico-clínico no

âmbito do banco de dados do NeuroMat, a equipe desenhou um protótipo que permitirá registrar e armazenar a história clínica progressiva dos pacientes, documentar as lesões e registrar a evolução clínica destes por meio de avaliações fisioterapêuticas e neurofisiológicas longitudinais. Esse trabalho detalhado de construção e digitalização das avaliações vem sendo realizado por uma equipe multidisciplinar composta por médicos, fisioterapeutas e neurocientistas, além da equipe de cientistas da computação do Instituto de Matemática e Estatística (IME) da USP. O resultado dessa iniciativa do Cepid-NeuroMat será a criação de uma base comum para o diagnóstico, a avaliação clínica e o prognóstico funcional de pacientes com lesão de plexo braquial. O modelo de banco de dados adotado pelo NeuroMat permitirá ainda colocar em um “terreno comum” as avaliações clínicas e todos os dados eletrofisiológicos coletados nos pacientes, possibilitando uma grande flexibilidade na consulta e análise dos dados. Estamos agora trabalhando no desenvolvimento de um protótipo eletrônico para a armazenagem, a manipulação e o compartilhamento dos dados, e esperamos em breve disponibilizá-lo para uso público. Quando tornada pública, essa base de dados poderá servir como um modelo para avaliação de outros pacientes com lesões semelhantes em nível mundial.

A criação do banco de dados NeuroMat abre uma oportunidade para que cientistas tenham acesso não somente a um universo de dados bem documentados e etiquetados, mas também ao processo que gerou essa ferramenta de trabalho compartilhada. Além disso, a disponibilização pública dos programas de análise que geraram os resultados cria um círculo virtuoso na medida em que permite a verificação pública de sua qualidade e veracidade. Dessa maneira abre-se uma janela de oportunidade para o avanço rápido do conhecimento nessa área. Esperamos contribuir para que haja mais compartilhamento de dados abertos na comunidade brasileira e internacional de Neurociência para que todos tenhamos melhores condições de trabalhar em conjunto e a Neurociência avance mais rápido, beneficiando diretamente a população. ①

\***Claudia Domingues Vargas** é professora associada do Programa de Neurobiologia do Instituto de Biofísica Carlos Chagas Filho e chefe do Laboratório de Neurociências e Reabilitação (LabNeR) do Instituto de Neurologia Deolindo Couto, ambos sediados na UFRJ; e **Fabio Kon** é professor titular em Ciência da Computação e vice-diretor do Centro de Competência em Software Livre (CCSL) do IME-USP. Ambos são pesquisadores do NeuroMat.