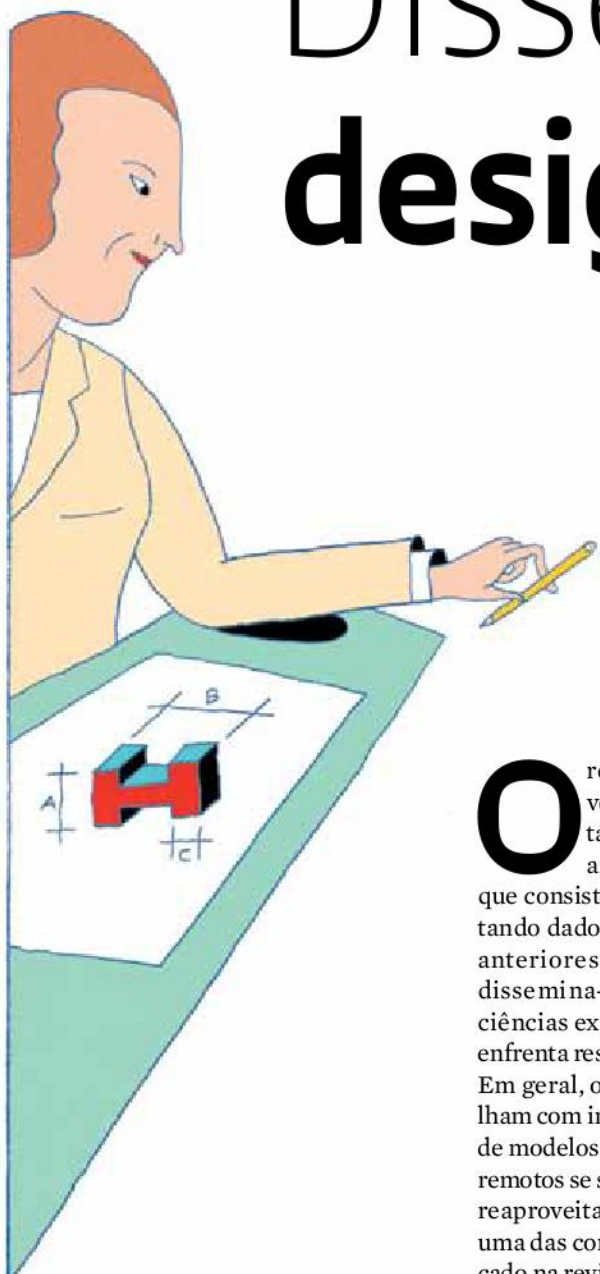


Disseminação desigual



Reúso de informações científicas ainda é baixo e varia de acordo com a área do conhecimento

Rodrigo de Oliveira Andrade

O reúso de dados de pesquisa vem crescendo, mas ainda está longe de se consolidar no ambiente científico. A prática, que consiste em fazer estudos aproveitando dados gerados em experimentos anteriores de outros pesquisadores, dissemina-se mais efetivamente nas ciências exatas e biológicas, enquanto enfrenta resistência nas ciências sociais. Em geral, os pesquisadores que trabalham com informações obtidas por meio de modelos computacionais ou sensores remotos se sentem mais confortáveis em reaproveitar dados de terceiros. Essa é uma das conclusões de um artigo publicado na revista *PLOS ONE* pela cientista de dados Renata Curty, da Universidade Estadual de Londrina (UEL), no Paraná.

Com base nas respostas de 595 pesquisadores de diferentes disciplinas e países, ela e seus colaboradores avaliaram o grau de disseminação do reúso

de dados e a percepção dos fatores que estimulam ou desencorajam a prática. Uma curiosidade é que essa análise se baseou em dados reutilizados. A fonte original foi mais de mil questionários respondidos por pesquisadores entre outubro de 2013 e março de 2014 no âmbito do Data Observation Network for Earth (DataONE), projeto da National Science Foundation (NSF).

Renata estuda a percepção acerca do reúso de informações científicas desde o doutorado na Universidade de Syracuse, nos Estados Unidos. À época, ela verificou que os pesquisadores das ciências sociais enxergam consequências potencialmente prejudiciais associadas à reutilização de dados. “Muitos têm receio de violar questões éticas ou de confidencialidade estabelecidas entre os sujeitos da pesquisa e investigadores”, ela explica. Outra preocupação diz respeito ao risco de má interpreta-



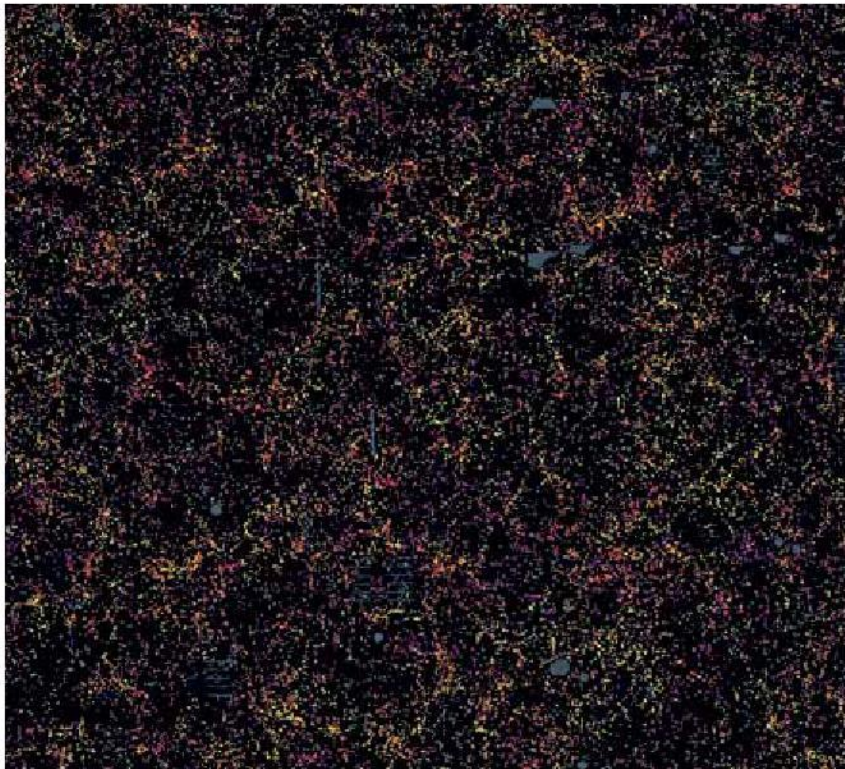
ção ou uso incorreto das informações originais. A opinião dos cientistas sociais sobre reaproveitamento de dados também é fortemente influenciada pelo próprio campo de atuação, que privilegia a produção de conhecimento novo. “Trabalhos que reutilizam dados são considerados menos autênticos e de menor impacto”, afirma.

O estímulo ao reaproveitamento de dados tem várias fontes. Uma delas é a preocupação crescente com a reprodutibilidade de pesquisas (ver seção *Boas Práticas na página 10*) e a importância de disponibilizar as informações primárias coletadas para que outros consigam verificar a precisão e a relevância dos resultados. No Brasil, desde 2013 a revista *Brazilian Political Science Review*, da Associação Brasileira de Ciência Política, exige que os autores de artigos cujo conteúdo utiliza métodos quantitativos disponibilizem no site da revista os da-

dos que embasaram o trabalho e também os *codebooks*, dicionários que permitem a identificação das variáveis usadas na obtenção das informações. “A proposta é tentar viabilizar a replicação dos procedimentos que levaram às conclusões dos trabalhos”, diz a cientista política Marta Arretche, do Departamento de Ciência Política da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo (FFLCH-USP) e editora da revista.

A pesquisadora lembra que a ciência só se torna replicável se os dados e ferramentas usados nos experimentos, simulações e análises forem disponibilizados de forma aberta e livre. No entanto, ela destaca, é fundamental que essa massa de informações seja acompanhada de explicações sobre sua origem. “Sem dados bem documentados não é possível reproduzir o experimento original ou reusar os dados em outra pesquisa”, completa.

Desde 2014 o grupo de periódicos científicos *PLOS* passou a condicionar a aceitação de artigos à divulgação de seus dados brutos em repositórios públicos (ver Pesquisa FAPESP nº 218). Em revistas de genética e bioinformática, cujos trabalhos geram uma enxurrada de informações sobre sequências de DNA e proteínas, essa recomendação há tempo se tornou uma exigência. Isso permitiu às geneticistas Lygia da Veiga Pereira e Maria Vibranovski, do Instituto de Biociências da USP, explicarem como ocorre o desligamento de uma das duas cópias do cromossomo X em embriões femininos. Elas analisaram informações disponibilizadas em 2013 por pesquisadores chineses e constataram que o gene *XIST*, responsável por iniciar a inativação, era expresso em embriões femininos a partir do estágio de oito células (ver Pesquisa FAPESP nº 260). “Os chineses haviam feito toda a parte laboratorial. Conseguim-



Reúso de dados sobre aglomerados de galáxias (acima) está ampliando as perspectivas de estudo de objetos celestes em astronomia

ram os embriões humanos, separaram as células, extraíram e sequenciaram o RNA, mas não olharam para a inativação do cromossomo X”, disse Lygia, que publicou o achado em setembro de 2017 na revista *Scientific Reports*.

Outro caso é o do grupo do parasitologista Marcelo Ferreira e da bióloga Priscila Rodrigues, ambos do Instituto de Ciências Biomédicas da USP. Há algum tempo eles estudam os padrões de dispersão global dos parasitas causadores da malária. “Usamos amostras do material genético dos protozoários disponíveis no GenBank, banco de sequências de DNA e de aminoácidos do Centro Nacional de Informação Biotecnológica dos Estados Unidos”, conta Priscila. Pelo menos dois artigos foram produzidos nos últimos três anos com base nessa prática: um em 2016, na *Nature Genetics*, destacando que o *Plasmodium vivax* acumulou mutações que o teriam diferenciado das cepas africanas e asiáticas depois de chegar às Américas, e outro em janeiro de 2018, na *Scientific Reports*, apresentando novas evidências sobre como as migrações humanas ajudaram a espalhar esses parasitas pelo continente americano.

O incentivo ao reúso de dados também provém das agências de fomento, interessadas não apenas no desafio da reprodutibilidade como em racionalizar a apli-

cação de recursos públicos nos projetos que financiam. “O compartilhamento de dados pode ajudar o cientista a economizar tempo e recursos e evita a duplicação de pesquisas”, destaca a engenheira eletrônica Claudia Bauzer Medeiros, do Instituto de Computação da **Universidade Estadual de Campinas (Unicamp)** e coordenadora adjunta do programa eScience da FAPESP. “Estudos internacionais mostram que a prática aumenta o número de parcerias, acelera descobertas científicas e amplia a visibilidade do conhecimento produzido”, afirma.

A ideia de que pesquisas produzidas com recursos públicos devem ter seus resultados franqueados sem restrições, inclusive em relação aos dados primários coletados, também se relaciona com o conceito de ciência aberta, que envolve o acesso livre à informação e a construção colaborativa do conhecimento, observa Claudia Domingues Vargas, do Instituto de Biofísica da Universidade

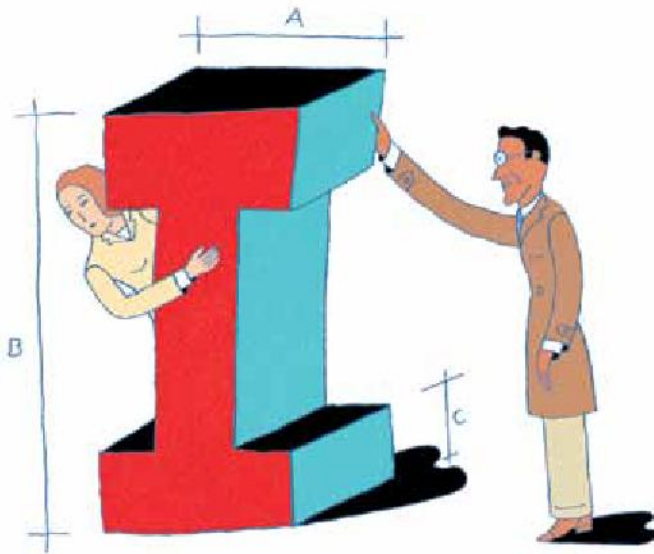
Federal do Rio de Janeiro (UFRJ). Ela é uma das pesquisadoras envolvidas no Neuroscience Experiments System (NES), que permite livre acesso às informações primárias de estudos no campo das neurociências.

A plataforma foi concebida no âmbito do Centro de Pesquisa, Inovação e Disseminação em Neuromatemática (NeuroMat), um dos Centros de Pesquisa, Inovação e Difusão (Cepid) apoiados pela FAPESP. Envolve pesquisadores de matemática, ciência da computação, estatística, neurociência, biologia, física e comunicação de universidades brasileiras e do exterior. “A proposta do NES é ser um repositório público que permita acesso aberto a um amplo conjunto de dados neurofisiológicos, clínicos e experimentais, bem como aos softwares usados na análise, processamento e geração dessas informações”, explica Claudia Vargas, que é uma das pesquisadoras principais do NeuroMat.

O compartilhamento de dados científicos avança em ritmos distintos a depender da área do conhecimento. Na astronomia, é algo corriqueiro, como observa a física Marcelle Soares-Santos, professora da Universidade Brandeis e pesquisadora no Fermi National Accelerator Laboratory, um dos mais importantes centros de física de partículas do mundo, nos Estados Unidos. “Beneficiei-me muito dessa prática no doutorado”, comenta. À época ela desenvolveu algoritmos para encontrar aglomerados de galáxias a partir de informações do Sloan Digital Sky Survey sobre 500 milhões de objetos celestes. Marcelle explica que os dados primários em astronomia são ricos e raramente explorados de forma completa. “Muitas questões em astronomia só podem ser estudadas a partir de análises de diferentes dados científicos obtidos por outros grupos de pesquisa.”

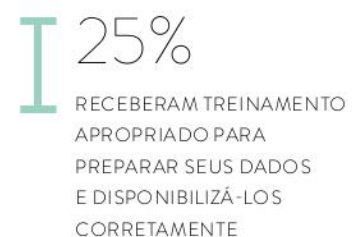
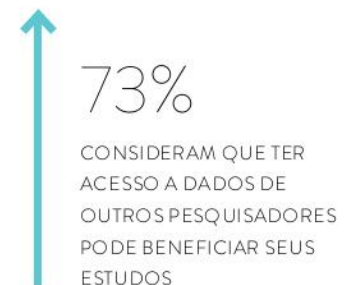
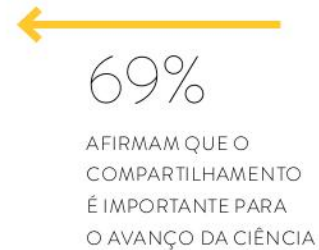
PARADOXO

O estudo publicado na *PLOS ONE* destaca um fato curioso relacionado à percepção sobre o reúso: os pesquisadores que mais se preocupam com a credibilidade dos dados que pretendem utilizar são os que se mostram mais dispostos a reaproveitar registros produzidos por terceiros. Já os que quase nunca reutilizam têm mais dificuldade para entender os benefícios dessa prática e avaliar a qualidade das informações disponíveis.



DISPOSIÇÃO PARA COMPARTILHAR

Atitudes relacionadas ao reuso de informações científicas de acordo com 1.162 pesquisadores de várias áreas e países



No estudo *How and why researchers share data (and why they don't)*, desenvolvido em 2014 pela editora John Wiley & Sons com quase 3 mil pesquisadores de diferentes áreas e países, verificou-se que os alemães são os mais dispostos a compartilhar dados, com o objetivo de aumentar a visibilidade e garantir a transparência de suas pesquisas. Já os chineses são menos propensos a dividir com outros informações de pesquisa, sobretudo porque isso não é um requisito para o financiamento. Os brasileiros reclamaram do trabalho extra para organizar essa massa de dados, dos custos para hospedá-la e das dificuldades para encontrar repositórios adequados.

Em estudos sobre o reuso de dados científicos, pesquisadores frequentemente alegam que se sentem receosos em fornecer suas informações porque ainda querem explorá-las em novos estudos ou temem não receber os créditos pela cessão. Esses e outros temores também foram verificados no relatório *Open Data: The research perspective*, da editora Elsevier. Mas o mesmo estudo constatou que 73% dos entrevistados julgavam que o acesso a dados científicos de terceiros poderia beneficiar suas próprias pesquisas e que 64% deles se mostravam dispostos a compartilhar informações com outros pesquisadores.

O principal desafio, segundo Claudia Bauzer Medeiros, é promover o reuso de informações científicas, mostrando aos pesquisadores os benefícios da prática, e ao mesmo tempo agir para combater os

casos de apropriação indevida de dados. Outra estratégia desejável, segundo ela, é a criação de cursos que ensinem pesquisadores e alunos a preparar dados e experimentos para compartilhamento. “Esse tipo de treinamento já é padrão em vários países do mundo, tendo em alguns casos se tornado uma exigência na formação de pesquisadores”, completa.

Renata Curty argumenta que é preciso investir em sistemas que verifiquem a qualidade dos dados disponibilizados e em recompensas para os pesquisadores que adotarem essa prática. Nos Estados Unidos já existem algumas iniciativas nesse sentido. Uma delas é a Plataforma Global de Informações sobre Biodiversidade (GBIF), que reúne quase 850 milhões de registros de espécies, 6 milhões deles oriundos do Brasil (ver Pesquisa FAPESP nº 263). Ao cadastrar os dados primários de suas pesquisas na GBIF, os pesquisadores podem gerar um *Data Paper*, um documento que pode ser publicado on-line em plataformas de acesso aberto voltadas para descrever conjuntos de informações de pesquisas consideradas valiosas. Existem publicações dedicadas a disseminar esses artigos de dados, segundo a pesquisadora da UEL, como o *Biodiversity Data Journal*, o *Data in Brief*, da Elsevier, e o *Scientific Data*, do grupo Nature. ■

Artigo científico

CURTY, R. G. *et al.* Attitudes and norms affecting scientists' data reuse. *PLOS ONE*. v. 12, n. 12, p. 1-22. jan. 2018.

FONTE: OPEN DATA: THE RESEARCH PERSPECTIVE, DA EDITORA ELSEVIER