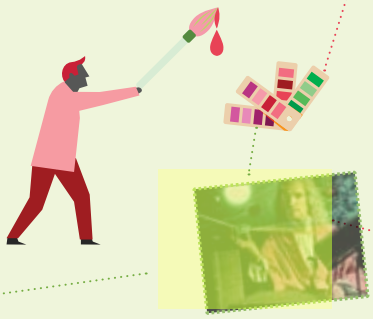


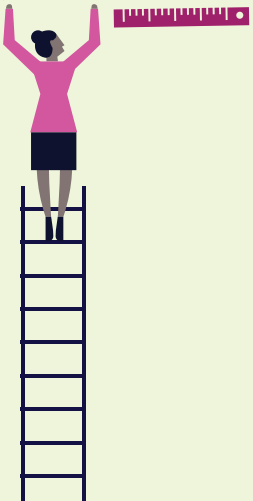
Uma estratégia para dados

Pesquisadores são estimulados a gerenciar e compartilhar as informações científicas que produzem

Bruno de Pierro

Gerenciar e armazenar grandes volumes de dados gerados em pesquisas são desafios enfrentados por cientistas em todos os campos do conhecimento. Na última década, algumas das principais agências de fomento, como a National Science Foundation (NSF) nos Estados Unidos e o Economic and Social Research Council do Reino Unido, passaram a exigir que os pesquisadores submetam, junto com as solicitações de financiamento, os chamados planos de gestão de dados, que descrevem como os dados produzidos serão gerenciados, preservados e divulgados em repositórios públicos. O objetivo é promover o compartilhamento de informações de forma a permitir a reutilização ou reprodução de experimentos, com isso acelerando novas descobertas científicas e racionalizando a aplicação de recursos.

No Brasil, não existe a obrigatoriedade da elaboração de planos de gestão. Em outubro do ano passado, a FAPESP deu um primeiro passo e anunciou que os pedidos de financiamento



Organização de um plano em cinco etapas

1. Destaque os tipos de dados que serão produzidos durante a execução do projeto. Por exemplo: registros de coleta, resultados experimentais, gráficos, mapas, vídeos, planilhas, gravações de áudio ou imagens.

2. Comunique eventuais restrições éticas ou legais para o compartilhamento de dados, além de medidas para garantir a privacidade, confidencialidade, segurança e propriedade intelectual.

3. Descreva a política de preservação e compartilhamento. Por exemplo, se os dados serão disponibilizados imediatamente ou apenas após a publicação de um artigo.

4. Apresente os métodos que serão empregados para armazenar os registros e torná-los acessíveis. Inclua os metadados (dados que descrevem conjuntos de dados) para que usuários possam reutilizar arquivos depositados em repositórios.

5. Atualize o plano sempre que necessário, incluindo correções de rumo e adoção de novas metodologias.

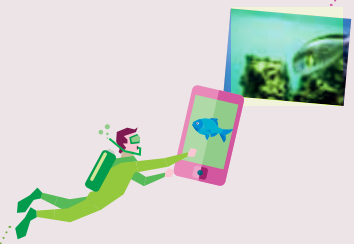
FONTE CLAUDIA BAUZER MEDEIROS

para projetos temáticos – aqueles com duração de cinco anos que se destacam por seus objetivos ousados – devem conter um documento complementar explicitando o plano de gestão de dados. A medida irá se estender gradativamente para outras modalidades de apoio ainda este ano. "Trata-se de uma iniciativa pioneira no país ao estabelecer políticas e diretrizes para o gerenciamento de dados científicos", afirma Claudia Bauzer Medeiros, professora do Instituto de Computação da Universidade Estadual de Campinas (Unicamp) e coordenadora do programa eScience da FAPESP.

O *Código de Boas Práticas* da Fundação, lançado em 2011, já estabelecia que os pesquisadores devem disponibilizar os registros resultantes de suas pesquisas. "A partir de agora, eles deverão detalhar como os dados serão gerenciados desde a coleta até a preservação, declarando de que forma e a partir de quando serão disponibilizados", diz. No Brasil, a Unicamp foi pioneira em criar formulários para planos de gestão e cadastrá-los no site

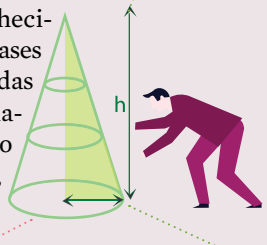
ILUSTRAÇÃO JÚLIA CHEREM RODRIGUES COM IMAGENS DA FREEPICK

É necessário detalhar como os dados serão gerenciados desde a coleta até a preservação



mundial DMPTool (dmptool.org). Essa iniciativa, comandada por Benilton de Sá Carvalho, do Instituto de Matemática, Estatística e Computação Científica (Imecc), permite que pesquisadores daquela universidade possam facilmente criar seus planos on-line e disponibilizá-los para todo o mundo. O DMPTool reúne mais de 200 instituições de pesquisa de diferentes países que oficializaram a criação e a disponibilização de seus planos de gestão de dados. No momento, apenas três universidades brasileiras estão na plataforma: a Unicamp, a Universidade de São Paulo (USP) e a Federal do ABC (UFABC).

A disponibilização de dados de experimentos ou coletados em campo tem o potencial de impulsionar parcerias e acelerar descobertas científicas ao ampliar a visibilidade da pesquisa. Em 2016, um consórcio internacional envolvendo mais de 30 organizações, entre elas a Fundação Oswaldo Cruz, a Academia Chinesa de Ciências e os Institutos Nacionais de Saúde (NIH) dos Estados Unidos, estimulou pesquisadores a compartilhar dados coletados durante o surto do vírus zika. A medida surtiu efeito e em poucos meses foram publicados estudos evidenciando a relação entre o zika e a microcefalia. Na área de biodiversidade, o armazenamento de dados científicos em repositórios garante o acesso a milhões de registros sobre espécies de plantas e animais, facilitando a produção de novos conhecimentos. A rede speciesLink, uma das bases digitais de biodiversidade desenvolvidas no país, permite a seleção de informações sobre a ocorrência e a distribuição de espécies de microrganismos, algas, fungos, plantas e animais. A plataforma reúne registros de 470 coleções do Brasil e de outros países. Essas coleções compartilham cerca de 9 milhões de registros de 125 mil espécies, das quais 2.756 ameaçadas de extinção.



Contudo, fazer um plano de gestão não se restringe a depositar dados em uma base on-line. De acordo com o Data Curation Center, órgão do Reino Unido responsável pela preservação de dados de pesquisa, o plano deve conter informações sobre como e por que os dados foram produzidos e armazenados. Por isso, é fundamental explicar como serão organizados os chamados metadados – dados que descrevem outros dados. "Trata-se de fornecer descrições sobre os conjuntos de dados, detalhando como eles foram produzidos, quando, onde e como podem ser reutilizados e também quem os gerou", esclarece a cientista da informação Márcia Teixeira Cavalcanti, professora da Universidade Santa Úrsula, no Rio de Janeiro, e membro do grupo de pesquisa Informação, Memória e Sociedade do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict). "Isso significa conferir identidade e padronização para os dados científicos, para que possam ser facilmente acessados nas buscas em repositórios e reutilizados em outras pesquisas", diz.

Em 2016, Márcia Cavalcanti foi uma das responsáveis pela curadoria de dados da plataforma CarpeDIEN (carpedien.ien.gov.br) do Instituto de Energia Nuclear (IEN), que realiza pesquisas em áreas como radiofármacos e inteligência artificial. "Levou um tempo para conseguirmos adequar os modelos de metadados mais apropriados para o tipo de informação com o qual estávamos lidando", conta. Segundo a pesquisadora, o processo de curadoria deve começar antes de os dados serem produzidos. "No plano de gestão, é importante

ILUSTRAÇÃO JÚLIA CHEREM RODRIGUES COM IMAGENS DA FREEPICK

que o pesquisador descreva até mesmo quais softwares ou equipamentos serão utilizados para gerar informações como imagens ou algoritmos." Para Claudia Medeiros, esse tipo de informação é fundamental. "Muitas vezes, ter acesso aos dados não é suficiente para reproduzir um experimento. É preciso também ter os mesmos programas de computador ou sistema operacional para repetir as mesmas condições do estudo original", destaca.

Durante o período no IEN, Márcia Cavalcanti realizou um levantamento sobre os repositórios de dados na Europa. Publicado no ano passado na revista do Instituto de Ciências Humanas e da Informação da Universidade Federal do Rio Grande (FURG), no Rio Grande do Sul, o estudo analisou a situação de 33 países, dos quais apenas nove declararam ter repositórios de acesso aberto a dados de pesquisa em 2016. De acordo com o trabalho, isso revela que em muitos países europeus a implementação de políticas de compartilhamento de dados de pesquisa se encontra em estágio inicial. O Horizonte 2020, o principal programa de apoio à pesquisa e à inovação da União Europeia, que entrou em vigor em 2007, lançou em 2016 um documento descrevendo os passos para a preparação de um plano de gestão de dados, que passaram a ser obrigatórios em todos os projetos submetidos a partir de 2017. Um dos pontos mais importantes do guia é chamar a atenção para situações em que a divulgação de dados brutos pode deflagrar problemas éticos. Por exemplo, ensaios clínicos que utilizam dados pessoais e precisam garantir a privacidade dos pacientes.

"Salvo exceções desse tipo, não existe argumento para justificar o não fornecimento de dados por pesquisadores financiados com dinheiro público", afirma Gilberto Câmara, pesquisador do Instituto Nacional de Pesquisas Espaciais (Inpe) e membro da coordenação do Programa FAPESP de Pesquisa sobre Mudanças Climáticas Globais. Segundo ele, muitos pesquisadores evitam depositar dados de experimentos antes de publicar o estudo em um periódico científico, alegando que as informações podem ser apropriadas por outros e publicadas sem receber o devido crédito. "Isso é conversa fiada", critica Câmara. Ele explica que o compartilhamento de dados independe da publicação do *paper*. Isso porque as informações depositadas em repositórios recebem um código identificador conhecido como Digital Object Identifier (DOI), permitindo a rastreabilidade do dado. "O fato é que, infelizmente, muito pesquisador não quer que alguém publique uma análise antes que ele, que coletou os dados, divulgue seu trabalho primeiro", diz Câmara.

"Todos os dados relativos aos meus trabalhos são depositados em bases abertas conforme são coletados", afirma o pesquisador, que publica dados gerados por análises de imagens de satélite na Pangaea, plataforma que reúne dados georreferenciados. Recentemente, informações guardadas por Câmara nessa base digital foram reutilizadas por pesquisadores do Restore+, um consórcio internacional com sede na Alemanha para promover estudos sobre o uso da terra. Câmara celebra a iniciativa da FAPESP de exigir o plano de gestão de dados dos pesquisadores. "Essa ação pode ajudar a combater hábitos perversos praticados no meio científico ao difundir boas práticas de gerenciamento de dados", aponta. "Há pesquisadores que se sentem donos dos dados e só os cedem a colegas se obtiverem algo em troca, como a coautoria do artigo. Essa conduta, infelizmente, é bastante frequente", diz. ■

Pesquisadores financiados com dinheiro público não podem se furtar a compartilhar informações, diz Câmara