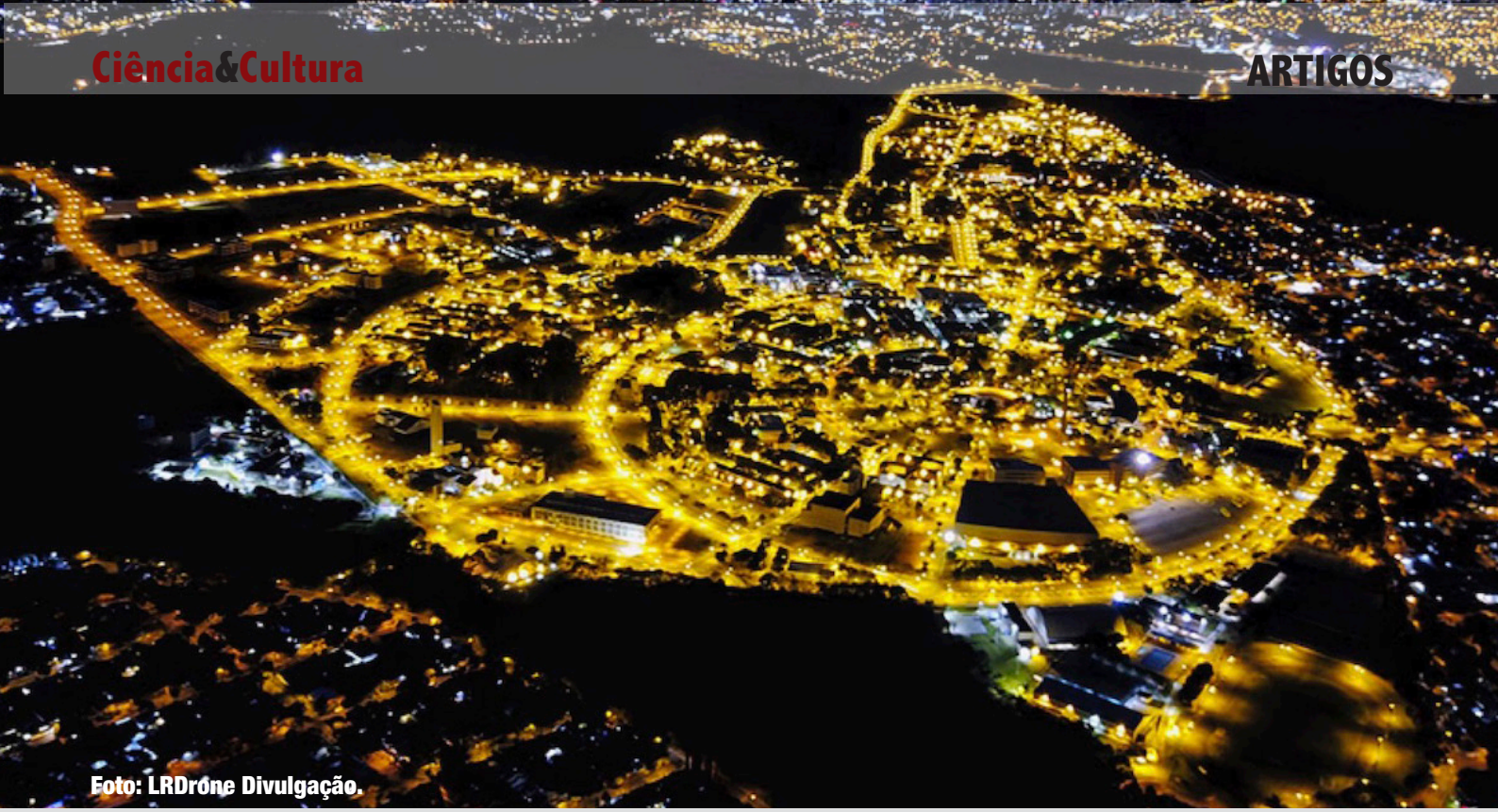

Foto: LRDrone Divulgação.

Num mundo digital, devemos valorizar menos os textos e o trabalho individual e valorizar mais o trabalho coletivo e conectado.

Dados abertos de pesquisa no Brasil

Diagnóstico e perspectivas futuras

por Thiago Lima Nicodemo

Resumo

Dados abertos de pesquisa são materiais digitais (dados brutos e metadados) gerados durante investigações científicas, publicados em repositórios com vários tipos de tecnologia, como o Dataverse. Esses dados devem seguir critérios claros de coleta e ser estruturados para garantir o reuso, alinhando-se aos princípios sustentáveis de reuso, disponibilidade e acessibilidade. No mundo inteiro, a ciência aberta está sendo impulsionada por universidades, institutos de pesquisa, agências de fomento e mesmo órgãos multinacionais (como a UNESCO). No Brasil, sua adoção está crescendo, com repositórios criados em universidades e institutos de pesquisa, e políticas praticadas por algumas agências de fomento. Contudo, os desafios persistem: resistência cultural de pesquisadores (priorização da produção de artigos como produtos), dilemas éticos (proteção de dados sensíveis) e infraestrutura precária. A publicação de dados exige distinção entre objetos de estudo (ex.: vídeos de entrevistas) e dados científicos (critérios metodológicos e metadados). Repositórios como o REDU, da Unicamp, ilustram políticas institucionais emergentes que obrigam o depósito de dados ao final de projetos. Para avançar, propõem-se políticas sistêmicas: planos de gestão de dados articulados com padrões internacionais (ex.: RDA, FAIRsharing), infraestrutura nacional soberana (evitando dependência de plataformas estrangeiras) e integração entre universidades e agências de fomento. Enquanto países como os membros da UE investem em políticas sistêmicas (ex.: European Open Science Cloud), o Brasil carece de uma estratégia nacional. Nesse sentido, a soberania de dados é crucial, especialmente com a ascensão da IA, que depende de volume de dados de qualidade e de bases interoperáveis.

Palavras-chave: Dados Abertos de Pesquisa; Ciência Aberta; Repositórios Institucionais; Soberania de Dados; Políticas de Gestão de Dados; Humanidades Digitais

Uma questão de método: o que são dados abertos de pesquisa?

Os dados abertos de pesquisa são dados brutos, metadados, dados digitais resultantes de análises cruzadas e outros materiais digitais gerados ou coletados durante atividades de pesquisa científica. Com o avanço dos repositórios e marcos regulatórios da ciência aberta no mundo, especialmente ao longo da década de 2010, tais como o Harvard Dataverse Repository (2011 e oferecido em código aberto em 2013), Zenodo (2013), Figshare (2011), Dryad (2008) e EUDAT (2011, infraestrutura europeia para dados de pesquisa multidisciplinares), foi-se convencionando que dados abertos devem ser, sobretudo, os critérios de uma determinada coleta de dados científicos e as informações estruturadas reunidas por meio desses critérios. Estabeleceu-se também que esses dados devem ser publicados em repositórios adequados para esse fim, tais como aqueles que operam com a tecnologia do Dataverse.

Inicialmente, os interessados em publicizar os seus dados de pesquisa devem estar cientes do que é compartilhável. Os dados de pesquisa não são novidades emergentes do mundo digital. Possuem suas bases nas próprias metodologias científicas de distintas áreas. O mundo digital multiplica a produção dos dados de pesquisa e acelera a necessidade de compartilhar informações em rede. Contudo, os dados de pesquisa têm sua origem nas bases de dados do mundo ainda analógico, quando os cientistas das mais diversas áreas produziam "fichários", escrevendo categorias gerais em cartões regulares e no corpo desses objetos. Nesse caso, os dados que estruturam a coleta eram o que hoje chamamos de "metadados de pesquisa" e as informações anotadas eram os dados.

Essa afirmação tão elementar pode não ser tão óbvia no olhar ainda estranho dos pesquisadores quando lhes é oferecida a possibilidade de publicar seus dados. Se a coleta é feita por meio de entrevistas gravadas, por exemplo, o vídeo ou áudio pode ser considerado como dado ou mesmo metadado, dependendo da estratégia do pesquisador ou grupo de pesquisa – por exemplo, se houver restrições éticas, o vídeo precisa ser editado ou mesmo guardado, mas não publicado. Já em outros casos, o vídeo da entrevista pode ser divulgado na íntegra, como arquivo de dados, por exemplo, quando há concordância dos entrevistados e convém como estratégia de dados abertos para aquela determinada pesquisa. Se a análise da pesquisa é feita a partir de prontuários médicos, tampouco são esses objetos que devem ser inscritos, mesmo que sejam anonimizados e que haja concordância dos interessados. Os prontuários podem entrar no repositório se for o caso, com tratamento adequado de anonimização dos dados. Mas o imprescindível é a coleta a partir desses dados, ou seja, o conjunto de informações que foram produzidas a partir da pesquisa. A mesma lógica se aplica para um(a) historiador(a) que digitaliza documentos de arquivo ou livros para depois analisá-los, a digitalização não é o principal objeto a ser incluído numa base de dados aberta – mas sim as informações coletadas (e os critérios dessa coleta) a partir da pesquisa com a massa de documentos digitalizados.

Resumindo, o que deve estar nos repositórios não são as coisas, pessoas ou objetos pesquisados, mas o produto da coleta científica e os critérios pelos quais foi realizada. Justamente por isso a resposta sempre depende da estratégia dessa determinada pesquisa. O mesmo raciocínio deve ser aplicado para a análise desses resultados, que pouco interessa para um repositório de dados abertos – mas a metodologia de análise pode interessar, para permitir auditoria e reprodutibilidade da pesquisa. Afinal, devemos deixar que nossos colegas de hoje e do futuro consigam alcançar resultados semelhantes com o mesmo material, estimulando a verificabilidade da ciência,[a] assim como devemos resguardar a possibilidade de que melhores ou outros resultados apareçam.

O que dificulta o compromisso da comunidade científica com os dados abertos começa, portanto, pela precariedade da formação metodológica. Com pouca consciência metodológica, pesquisadores sempre terão dificuldade em dar sustentabilidade às suas coletas. Os dados abertos começam, portanto, com a clareza de: quais dados são coletados? Como os dados são coletados (quais são os critérios da coleta)? E a

partir de quais materiais brutos esses dados são coletados? A tendência para armazenar no repositório é privilegiarmos os dois primeiros casos. Em outras situações, o último caso, o dos dados brutos a partir dos quais a coleta ocorreu, podem entrar no repositório, seja por necessidade de preservação, seja de registro, ou de garantia de reprodutibilidade, dentre outras circunstâncias possíveis.

Quão públicos são os dados de pesquisa?

A história do conhecimento aberto é mediada por marcos regulatórios, declarações e convenções como a *Budapest Open Access Initiative*, de 2002, a *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*, de 2003,[a] a *Recommendation on Open Science* da UNESCO, de 2021 e a *OECD – Enhanced Access to Publicly Funded Data for Science, Technology and Innovation*, de 2023.[b] Esses documentos sob diversos aspectos encorajam os pesquisadores a compartilharem seus dados. A principal diretriz para considerarmos os dados de uma pesquisa pública é certamente o investimento de dinheiro público em tal pesquisa. Parece óbvio pensar assim, mesmo porque esse é um princípio que ultrapassa as fronteiras da ciência. Pertence aos princípios públicos da transparência, que se manifesta, pelo menos no caso brasileiro, em leis como a Lei de Acesso à Informação de 2011, que diz no seu artigo 8º, que são públicas as informações produzidas ou custodiadas por entidades públicas, ou por entidades privadas que recebam recursos públicos para realização de ações de interesse público.

Pensando no caso brasileiro, qualquer dado de pesquisa financiado por instituições de fomento, tais como a CAPES, o CNPq, a FAPESP, ou até mesmo sem financiamento direto, mas dependentes da infraestrutura e recursos de qualquer universidade pública, são públicos e têm compromisso com a transparência. Agências de fomento à pesquisa e universidades públicas devem cobrar que os pesquisadores publiquem os dados das suas pesquisas, por princípio. Isso já ocorre em alguns casos, como no REDU da UNICAMP. O depósito de dados no repositório é obrigatório pelo menos para aqueles que concluem uma pós-graduação, a menos de restrições éticas ou legais.

> "A ciência aberta é uma realidade sem volta, uma condição para que a nossa ciência ganhe maior maturidade, amplitude e escala."

Seguindo a lógica presente na Lei de Acesso à Informação, qualquer tipo de sigilo ou restrição, quando se trata de informações públicas, deve ser a exceção e não a regra. Contudo, em quais casos os dados devem ter acesso restrito ou sigilo quando se trata de conhecimento aberto? Em primeiro lugar, dados pessoais sensíveis são mais suscetíveis a vetos ou restrições de acesso. Dados sensíveis são aqueles que podem revelar aspectos íntimos de um indivíduo, cujo uso inadequado ou não autorizado pode resultar em discriminação, ou danos ao titular. Bons exemplos são os dados referentes à saúde ou à vida sexual, bem como dados genéticos ou biomédicos. Isso não impede que os dados sejam organizados e compartilhados, mas é importante neste caso que estejam "tratados", sem qualquer tipo de identificação, e que a coleta tenha sido feita consoante a um termo de consentimento. Então nem todo dado pessoal precisa ser restrito, o que é necessário é se cumprir esses parâmetros legais e éticos. A pesquisa com seres vivos, aliás, impõe uma necessidade de articulação importante entre os dados abertos e os protocolos elaborados pelos comitês de ética em pesquisa. Os dados advindos de pesquisas com seres humanos publicados em repositórios abertos devem ter sido aprovados nos comitês de ética em pesquisa.

A proteção de patentes e de propriedade intelectual e segredos comerciais também podem ser um elemento de restrição. As dúvidas vêm frequentemente de pesquisas que têm, pelo menos em parte, financiamento privado. Imaginem por exemplo a publicização dos testes para liberação de um medicamento patenteado ou de uma nova tecnologia. Qual seria o sentido em oferecer para a concorrência os dados que permitiram a evolução e o refinamento

de um produto tão custoso? A restrição de acesso deve ser considerada pertinente neste caso. Não podemos esquecer, no entanto, que essas restrições foram levantadas no compartilhamento de dados de ensaios associados a COVID 19 – o amplo compartilhamento entre todos os grupos de pesquisa, ação pioneira na época, é reconhecido por todos como tendo permitido o desenvolvimento de vacinas em um tempo recorde. Houve restrição ao acesso aberto, mas não de acesso irrestrito entre laboratórios de pesquisa de vacinas. A mesma lógica de prerrogativa de restrição deve ser aplicada para dados relacionados com contenciosos jurídicos, para evitar interferências no andamento do processo ou outros tipos de processos decisórios em andamento.[c]

Outro caso importante de restrição de acesso é uma possível ameaça à segurança da sociedade ou do Estado. Casos nos quais os dados possam comprometer atividades de inteligência ou de fiscalização em andamento; ou dados de geolocalização de móveis, imóveis em ação ou em campanha estratégica militar ou de polícia.

Hábitos acadêmicos de não-compartilhamento

Existe muita resistência à cultura da ciência aberta e do compartilhamento de dados de pesquisa. Talvez a mais arraigada das oposições seja originária da cultura autoral, focada na supervalorização do artigo e do livro, estabelecidos como resultados mais importantes de um projeto de investigação. No fundo, paira na universidade uma cultura de privatização das informações produzidas e/ou coletadas com dinheiro público das agências de fomento e das universidades. Podemos considerar esse conjunto de atitudes como uma espécie de patrimonialismo, ou seja, de apropriação para o espaço privado daquilo que é ou deveria ser público. Isso tem também relação com uma atomização da produção no mundo moderno, reproduzida na universidade: cada professor(a) cuida do seu laboratório como um feudo, cada aluno produz informações sem se preocupar com a sustentabilidade desses dados no futuro, somente focados em extrair conclusões e publicar seu artigo ou tese.

O mundo digital acelera a níveis sem precedentes a produção de dados, bem como sua circulação. Isso impõe em todos os campos uma mudança para uma lógica sistêmica e integrada de produção. Para isso acontecer, também são necessários critérios em comum previamente estabelecidos. A digitalização também permite a integração entre os dados comparáveis, levando a mudanças profundas na abrangência das pesquisas e na velocidade com que dados e conclusões podem ser produzidos. No meu campo, o da história, a digitalização de acervos documentais e bibliográficos permitiu a criação de bases de dados transversais, que acumulam milhões de documentos advindos de diferentes arquivos no mundo. Isso permite a expansão dos estudos e o cruzamento desses dados. O texto, enquanto resultado de um processo de pesquisa, tende a ficar mais desatualizado e mais parcial na medida em que o tempo passa, para não falar das habilidades de produção textual e de cruzamento de dados dos modelos de linguagem de inteligência artificial, com profundas consequências no fazer científico. (Figura 1)

Isso joga luz na pesquisa enquanto um processo. Estudei no meu doutorado o pensamento de Sérgio Buarque de Holanda. De forma exaustiva indexei os principais temas e autores debatidos nos seus artigos de jornal, demonstrando como alguns assuntos apareciam depois desenvolvidos nos seus livros. Para terminar o processo, fui à sua biblioteca e processei todos os livros aos quais ele fazia referência, buscando por grifos e anotações. Guardei essas informações em cadernos, fichamentos, marcadores dentro de livros (*post-its*) ou em imagens que fiz com uma câmera digital. Enquanto pesquisava, meu principal objetivo era produzir o texto do doutorado, de modo que não me preocupei em nada com a sustentabilidade e o reuso dessas informações. Apenas me preocupei de que fossem verificáveis, por meio de referências precisas, citações e notas de rodapé. Se tivesse desenvolvido uma base de dados, teria elaborado uma indexação do processo criativo daquele autor e com a tecnologias disponíveis já naquele momento poderia ter expandido esse cruzamento ao infinito, na reverberação daqueles temas em outras obras e textos. Isso ajudaria muito a entendermos, por exemplo, como certo autor

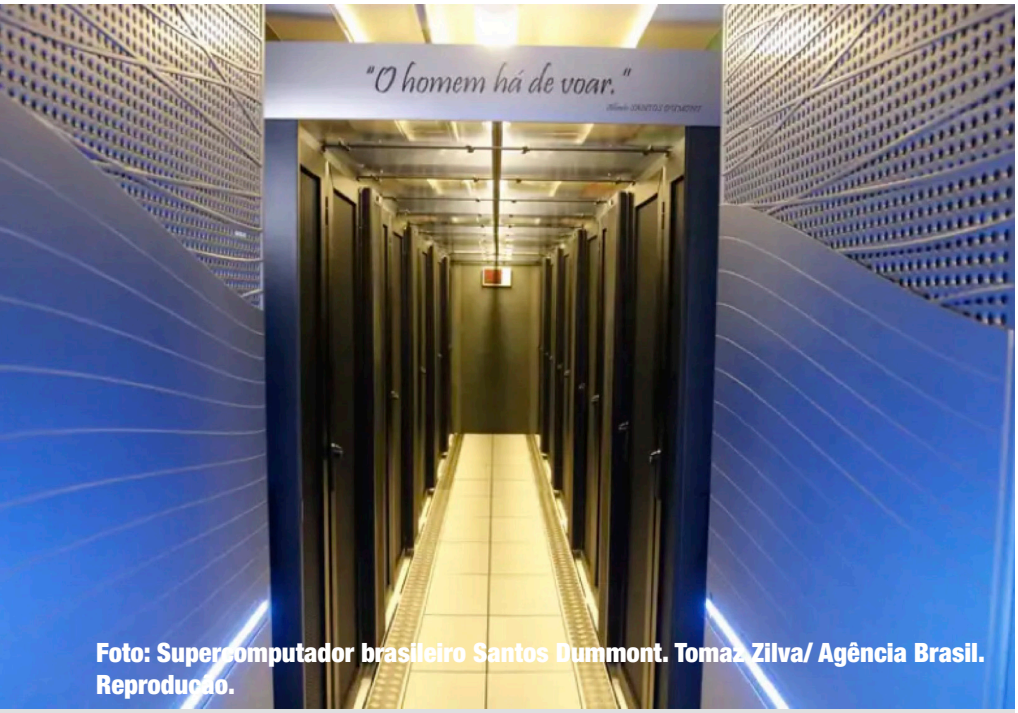

Foto: Supercomputador brasileiro Santos Dummont. Tomaz Zilva/ Agência Brasil. Reprodução.

Figura 1. A digitalização integra os dados comparáveis, permitindo mudanças profundas nas pesquisas e na velocidade com que os dados e contribuições são produzidos.

ou certa ideia foram lidos, recebidos e entendidos na cultura brasileira do século XX. Se tivesse estabelecido uma política de coleta de dados com outros colegas que estudavam outros autores, poderíamos ter chegado numa constelação cruzada, o que geraria uma grande plataforma colaborativa. Infelizmente não foi o que ocorreu porque só eu entendo dos dados de pesquisa que produzi, foi uma organização muito pessoal, o que os tornam inacessíveis ou pouco úteis para qualquer colega no presente e no futuro.

Como deve ser uma política sistêmica de dados abertos?

Produzir dados melhores e com a consciência de que serão compartilhados não é uma tarefa exatamente fácil, especialmente para quem está começando. Para um jovem pesquisador, saber o que é de comum interesse no seu campo mostra-se necessário no mínimo a tutela de alguém mais experiente. O que seria necessário saber? Requer-se um conhecimento muito grande da área científica na qual se produz o conhecimento. O que é de comum interesse entre a maioria dos pesquisadores? O que já foi comprovado em uma região ou de um modo e agora deve ser do outro?

Na maioria das vezes, são necessários padrões de metadados ou pelo menos diretrizes, que podemos aqui considerar "políticas", que sejam de preferência coordenadas por agrupamentos ou associações de pesquisadores. No mínimo, cada laboratório ou centro de pesquisa deve ter a sua e esta deve estar articulada em rede com outros laboratórios dedicados a uma área específica.

Também é necessário seguir padrões propostos por consórcios de instituições dedicadas ao assunto, tais como o grupo Research Data Alliance (RDA), da *Go Fair Initiative* e da *Faisharing.org* que procura oferecer catálogo de padrões de metadados para interoperabilidade e compartilhamento.[d] Esta última oferece um catálogo de 1830 padrões, 2292 banco de dados e 342 políticas de compartilhamento de dados.

Existem áreas científicas mais consolidadas no que se refere as práticas de compartilhamento, como, por exemplo, aquelas relacionadas a dados genéticos, como o GenBank.[e] O caso do mapeamento genômico da COVID-19, realizado pelo NCBI Virus[f] da National Library of Medicine e pelo GISAID é um bom exemplo.[g] Graças a aplicação de inteligência artificial à coleta de dados abertos de mapeamentos de perfis genéticos e variantes dos vírus foi possível o desenvolvimento de uma geração de vacinas mais abrangente e polivalente. Os dados abertos são práticas mais consolidadas em áreas como a astronomia e a física, graças ao Open Data Portal do CERN ou o Sloan Digital Sky Survey;[h] bem como na área de ciências climáticas (Coupled Model Intercompatison Promessa),[i] neurociência (ConnectomeDB),[j] biodiversidade (Global Biodiversity Information Facility – GBIF),[k] dentre outras.

A política de dados da Universidade Estadual de Campinas (Unicamp), por exemplo, é estruturada por uma Comissão de Gestão de Dados de Pesquisa (CGDP),[l] que tem representantes de todas as áreas do

conhecimento e a incumbência de sustentar e promover a política institucional de dados abertos de pesquisa,[m] observando as melhores práticas em âmbito internacional e estimulando uma cultura de dados abertos e de compartilhamento sustentável na universidade. A CGDP é também responsável pela gestão do Repositório de Dados de Pesquisa (REDU),[n] instrumento oficial incumbido de armazenar conteúdos digitais na forma de software, dados brutos de pesquisa, gravações de áudio e vídeo, questionários, códigos computacionais, fotografias e imagens, planilhas, entre outros. O REDU utiliza a infraestrutura Dataverse e é articulado com a rede de repositórios de dados de pesquisa do estado de São Paulo,[o] criada sob coordenação da A Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp), e que inclui os repositórios de natureza análoga da Universidade de São Paulo (USP), da Universidade Estadual Paulista (Unesp), da Universidade Federal de São Carlos (UFSCar), da Universidade Federal de São Paulo (Unifesp), da Universidade Federal do ABC (UFABC), e da Embrapa. Esta última participa da rede desde sua concepção (de 2017 a 2019), tendo em vista que o repositório foi concebido e implementado pela Embrapa Agricultura Digital, sediada no estado de São Paulo.

A expansão de uma prática nesse sentido deveria obedecer a uma lógica sistêmica de modo que cada universidade ou instituto de pesquisa (ou até mesmo unidades dentro deles) teria um órgão responsável pela política de dados abertos, inclusive pela sua difusão e treinamento. Esse órgão deveria assessorar cada instituto ou unidade em elaborar uma política de dados abertos, que por sua vez serviria de matriz para que cada departamento e cada laboratório criado dentro do órgão tenha sua política de dados abertos. A política normalmente é um documento com diretrizes, que não se confunde com o plano de ação ou com o conjunto de parâmetros e metadados. Em muitos casos, as políticas podem também incluir essas informações, mas o fundamental é que explicite uma direção: como os dados devem ser organizados e qual é o pressuposto para interoperabilidade, ou pelo menos, para aproximação dos dados de diferentes pesquisas.

As agências de fomento deveriam por sua vez obrigar a todos que recebem dinheiro público a propor nas suas pesquisas um plano de gestão de dados e ao final do projeto o depósito dos dados e metadados em um repositório. Os planos de dados abertos deveriam estar em consonância com as políticas das comunidades acadêmicas e laboratórios com os quais a pesquisa dialoga. (Figura 2)

Um plano de gestão de dados deve recapitular muitos dos temas tratados aqui: quais são as políticas e padrões de dados e metadados compartilháveis produzidos pela comunidade com o qual a investigação dialoga; quais são os limites de publicização dos dados, ou seja, quais dados não devem ser publicados; como os dados podem ser reutilizados por outros pesquisadores; qual é a estratégia de publicização do que será produzido; quais são as instâncias que vão supervisionar e validar o processo; e finalmente em qual infraestrutura os dados vão ficar e qual é o grau de sustentabilidade dessa infraestrutura. A pouca consciência do processo deveria ser um elemento importante para avaliar se uma pesquisa pode ser aprovada ou não.

Soberania de dados - onde e como devem ficar com os dados?

Muitas pesquisas financiadas no Brasil publicam seus dados em portais internacionais, tais como o Harvard Dataverse Repository, o MIT Libraries Data Management, o Stanford Digital Repository, o Oxford University Research Archive, o Zenodo ou o FigShare, dentre outros. No entanto, é fundamental nos atentarmos ao fato de que o conhecimento tem uma dimensão soberana e não podemos contribuir com o enfraquecimento das instituições de ensino e pesquisa brasileiras, ainda mais quando a pesquisa é financiada nacionalmente. Com o desenvolvimento de tecnologias e da própria IA, será possível cruzar muitos dados e obter resultados incríveis. É importante que o Brasil esteja em condições de participar dessa corrida. Também é importante ressaltar que a legislação, especialmente a Lei Geral de Proteção de Dados (Lei 13.709/2018), que impõe restrições rigorosas para o compartilhamento

Figura 2. Planos de dados abertos devem ser coerentes com as políticas das comunidades acadêmicas e laboratórios com os quais a pesquisa dialoga.

internacional de dados, contém ressalvas específicas para dados de pesquisa.

Do ponto de vista da infraestrutura, a pergunta mais comum é sobre a legitimidade e a segurança do armazenamento em nuvem. No entanto, antes da infraestrutura, é necessário ter governança na organização que custodia os dados, pois os dados abertos de pesquisa são somente uma parte dos dados fundamentais ligados às atividades finalísticas da instituição. É essencial que os dados abertos estejam em um plano da organização, da universidade, na maior parte dos casos, de gestão e de infraestrutura, separando os dados finalísticos dos serviços, tais como servidores e provedores de e-mails, comunicação corporativa, dentre outros. Isso para haver não somente backup e redundância, mas preservação digital de fato. A nuvem é um problema falso, desde que este espaço esteja organizado com padrões claros e maturidade institucional, além de, claro, de um plano de saída sustentável, quando se tratar de contratos. Os contratos são sempre finitos, mas é necessário que os dados sejam entregues ao final de forma sustentável, sem comprometimento da sua integridade e reprodutibilidade, garantindo a migração de plataformas.

O nosso problema infraestrutural maior não está na "subida" dos dados em repositório de dados abertos, mas na forma com que os dados são produzidos e armazenados. Ao propor um grande projeto para uma agência de fomento, não se cogita na pesquisa brasileira qual é o grau de sustentabilidade da infraestrutura proposta. Afinal, praticamente todo projeto de pesquisa hoje resultará numa base de dados e numa estratégia de divulgação digital dos dados. Fomentar infraestruturas isoladas, sem sustentabilidade, a longo prazo, corresponde à outra face do mesmo problema que produzir dados sem critérios claros para compartilhamento e interoperabilidade. O cenário ideal é que cada instituição de ensino e pesquisa ou um consórcio entre elas tenha um laboratório/infraestrutura de digitalização, e que, ao ser contemplado, um projeto científico deva então participar desse "condomínio" com recursos e governança. Essa parte laboratorial desta infraestrutura poderia facilitar a produção responsável de dados, ajudando a compactuar formatos e metadados. Isso é exatamente o que ocorre no mundo com os grandes *hubs* ou "supercomputadores" dedicados ao processamento de dados. Também a mesma lógica aplica-se em pesquisas com grandes infraestruturas de aceleradores de partículas.

Além de economizar centenas de milhões de reais das agências de fomento, uma medida como essa evitaria a perda de dados e mitigaria a obsolescência desses projetos depois que o financiamento acabar (coisa que acontece frequentemente). Também criaria condições favoráveis para o processamento e cruzamento dessas informações em larga escala, inclusive (mas não exclusivamente) com modelos de inteligência artificial.

Num mundo no qual tudo é produzido digitalmente e em profusão e em que a inteligência artificial é uma realidade, devemos de fato valorizar menos os textos e o trabalho individual e valorizarmos mais

o trabalho coletivo e conectado. A ciência aberta é uma realidade sem volta, uma condição para que a nossa ciência ganhe maior maturidade, amplitude e escala. A inteligência artificial só será aproveitada de forma séria sendo alimentada com dados consistentes e interoperáveis e com algoritmos abertos. A universidade, nesse caso, deveria ser um espaço para promover soluções tecnológicas, de governança e defender seus próprios interesses, os interesses da ciência e do desenvolvimento. A reflexão crítica e a tarefa formadora de uma ciência aberta também devem ter como espaço privilegiado a universidade, criando condições para uma transformação de mentalidade. Então não é necessário apenas ações em dados abertos, mas também que a universidade encampe a pesquisa e reflexão sobre os dados. Contudo, isso não basta. Os gestores da ciência, professores e pesquisadores que ocupam cargos estratégicos e que atuam nas agências, devem ter compromissos claros com uma agenda soberana que envolva a otimização de recursos públicos e a ampliação da capacidade de armazenamento e processamento de dados científicos em larga escala e em interconexão.

Thiago Lima Nicodemo é diretor do Arquivo Público do Estado de São Paulo, professor do Departamento de História da UNICAMP e diretor do Centro de Humanidades Digitais do IFCH--UNICAMP

NOTAS

[a] WILKINSON, M. D. et al. FAIR Digital Twins for Reproducible Research. Scientific Data, v. 10, n. 1, p. 104, 2023. DOI: 10.1038/s41597-023-01999-2. Disponível em: https://www.nature.com/articles/s41597-023-01999-2. Acesso em: 18 fev. 2025.

[b] UNESCO. Recommendation on Open Science. Paris: UNESCO, 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000379949. Acesso em: 18 fev. 2025.

OECD. Enhanced Access to Publicly Funded Data for Science, Technology and Innovation. OECD Publishing, 2023. DOI: 10.1787/9b6d8e2c-en. Disponível em: https://www.oecd-ilibrary.org/science-and-technology/enhanced-access-to-publicly-funded-data-for-science-technology-and-innovation_9b6d8e2c-en. Acesso em: 18 fev. 2025.

[c] ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). OECD Principles and Guidelines for Access to Research Data from Public Funding. Paris: OECD Publishing, 2007. Disponível em: https://www.oecd.org/science/sci-tech/38500813.pdf. Acesso em: 18 fev. 2025.

[d] RESEARCH DATA ALLIANCE. How the RDA works. Disponível em: https://www.rd-alliance.org/how-the-rda-works/. Acesso em: 18 fev. 2025.

GO FAIR. GO FAIR Initiative. Disponível em: https://www.go-fair.org/. Acesso em: 18 fev. 2025.

FAIRSHARING. FAIRsharing: Connecting data policies, standards & databases. Disponível em: https://fairsharing.org/. Acesso em: 18 fev. 2025.

[e] NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. GenBank Overview. Disponível em: https://www.ncbi.nlm.nih.gov/genbank/. Acesso em: 18 fev. 2025.

[f] NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. NCBI Virus - Viral Sequence Search Interface (VSSI). Disponível em: https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/. Acesso em: 18 fev. 2025.

[g] GISAID. Global Initiative on Sharing All Influenza Data. Disponível em: https://gisaid.org/. Acesso em: 18 fev. 2025.

[h] SLOAN DIGITAL SKY SURVEY. Data Release 18 (DR18). Disponível em: https://www.sdss.org/dr18/. Acesso em: 18 fev. 2025.

[i] PROGRAM FOR CLIMATE MODEL DIAGNOSIS & INTERCOMPARISON. CMIP6 - Coupled Model Intercomparison Project Phase 6. Disponível em: https://pcmdi.llnl.gov/CMIP6. Acesso em: 18 fev. 2025.

[j] HUMAN CONNECTOME PROJECT. HCP Data - ConnectomeDB. Disponível em: https://db.humanconnectome.org/. Acesso em: 18 fev. 2025.

[k] GLOBAL BIODIVERSITY INFORMATION FACILITY. GBIF - Free and Open Access to Biodiversity Data. Disponível em: https://www.gbif.org/. Acesso em: 18 fev. 2025.

[l] PRÓ-REITORIA DA UNICAMP. Comissão de Gestão de Dados de Pesquisa. s.d. Disponível em: https://prp.unicamp.br/comissoes/gestao-de-dados-de-pesquisa/comissao/. Acesso em: 18 fev. 2025.

[m] UNICAMP. Deliberação CONSU-A-016/2020 de 6 de outubro de 2020. Disponível em: https://www.pg.unicamp.br/norma/23869/0. Acesso em: 18 fev. 2025.

[n] UNICAMP. Repositório de Dados de Pesquisa da Unicamp. s.d. Disponível em: https://redu.unicamp.br/. Acesso em: 18 fev. 2025

[o] REDE DE REPOSITÓRIOS DE DADOS DE PESQUISA – SP, s.d. https://metabuscador.uspdigital.usp.br. Acesso em: 19 fev. 2025

REFERÊNCIAS

BERLIN DECLARATION ON OPEN ACCESS TO KNOWLEDGE IN THE SCIENCES AND HUMANITIES. 2003. Disponível em: https://openaccess.mpg.de/Berlin-Declaration. Acesso em: 18 fev. 2025.

BUDAPEST OPEN ACCESS INITIATIVE. Budapest Open Access Initiative. 2002. Disponível em: https://www.budapestopenaccessinitiative.org. Acesso em: 18 fev. 2025.

FAIRSHARING. FAIRsharing: Connecting data policies, standards & databases. Disponível em: https://fairsharing.org/. Acesso em: 18 fev. 2025.

GISAID. Global Initiative on Sharing All Influenza Data. Disponível em: https://gisaid.org/. Acesso em: 18 fev. 2025.

GO FAIR. GO FAIR Initiative. Disponível em: https://www.go-fair.org/. Acesso em: 18 fev. 2025.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. GenBank Overview. Disponível em: https://www.ncbi.nlm.nih.gov/genbank/. Acesso em: 18 fev. 2025.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. NCBI Virus - Viral Sequence Search Interface (VSSI). Disponível em: https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/. Acesso em: 18 fev. 2025.

OECD. Enhanced Access to Publicly Funded Data for Science, Technology and Innovation. Disponível em: https://www.oecd-ilibrary.org/science-and-technology/enhanced-access-to-publicly-funded-data-for-science-technology-and-innovation_9b6d8e2c-en. Acesso em: 18 fev. 2025.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). OECD Principles and Guidelines for Access to Research Data from Public Funding. Disponível em: https://www.oecd.org/science/sci-tech/38500813.pdf. Acesso em: 18 fev. 2025.

PROGRAM FOR CLIMATE MODEL DIAGNOSIS & INTERCOMPARISON. CMIP6 - Coupled Model Intercomparison Project Phase 6. Disponível em: https://pcmdi.llnl.gov/CMIP6. Acesso em: 18 fev. 2025.

PRÓ-REITORIA DA UNICAMP. Comissão de Gestão de Dados de Pesquisa. s.d. Disponível em: https://prp.unicamp.br/comissoes/gestao-de-dados-de-pesquisa/comissao/. Acesso em: 18 fev. 2025.

RESEARCH DATA ALLIANCE. How the RDA works. Disponível em: ttps://www.rd-alliance.org/how-the-rda-works/. Acesso em: 18 fev. 2025.

SLOAN DIGITAL SKY SURVEY. Data Release 18 (DR18). Disponível em: https://www.sdss.org/dr18/. Acesso em: 18 fev. 2025.

UNESCO. Recommendation on Open Science. Paris: UNESCO, 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000379949. Acesso em: 18 fev. 2025.

UNICAMP. Deliberação CONSU-A-016/2020 de 6 de outubro de 2020. Disponível em: https://www.pg.unicamp.br/norma/23869/0. Acesso em: 18 fev. 2025.

UNICAMP. Repositório de Dados de Pesquisa da Unicamp. s.d. Disponível em: https://redu.unicamp.br/. Acesso em: 18 fev. 2025.

WILKINSON, M. D; et al. FAIR Digital Twins for Reproducible Research. Scientific Data, v. 10, n. 1, p. 104, 2023. Disponível em: https://www.nature.com/articles/s41597-023-01999-2. Acesso em: 18 fev. 2025.