



## Uso de dados disponíveis em repositórios abertos aprimora pesquisas e estimula colaboração entre cientistas

Cristiane Miglionara, Simone Dias Marques | 25 de julho de 2024 | Reportagens

### Transparência | Embora ainda em implantação, a prática tem ganhado força no Brasil a partir de políticas públicas, universidades e iniciativas comprometidas com a Ciência Aberta

\*Foto: Nede Makiko Furukawa/Embrapa

Toda pesquisa – seja científica, para implantação de políticas públicas, mapeamento de negócios ou de qualquer outro tipo – necessita de dados. Esses dados são coletados de várias formas, entre elas entrevistas, levantamentos e censos ou, ainda, derivados de outras pesquisas. Nesse sentido, o uso de um repositório de dados pode ser imprescindível, ainda mais quando ele reúne e disponibiliza de forma aberta informações coletadas por pesquisadores e pesquisadoras das mais diversas áreas.

Essa é a proposta dos repositórios de dados no modelo da Ciência Aberta. São iniciativas como o repositório do Sistema Nacional de Assistência Social (SNAS), do Ministério do Desenvolvimento Social, que levam em conta as diretrizes expressas pelo acrônimo FAIR: *Findable, Accessible, Interoperable, Reusable* – ou, em português, Encontráveis, Acessíveis, Interoperáveis e Reutilizáveis – para promover uma maior disseminação e popularização da produção científica.

A mestranda em Políticas Públicas na UFRGS e especialista em Gestão Pública pela Universidade Estadual do Rio Grande do Sul (UERGS) Jéssica Krislei Costa de Neque utiliza bancos de dados abertos desde 2022. “Minha primeira experiência de pesquisa com bancos de dados abertos foi durante a escrita do meu Trabalho de Conclusão de Curso (TCC), enquanto cursava Políticas Públicas. Descobri esses bancos de dados através de recomendações de professores e colegas e em pesquisas online sobre fontes de dados confiáveis para estudos na área”, conta.

Para Jéssica, a relevância de dados como os disponibilizados pelo SNAS está em sua capacidade de fornecer uma base detalhada para análise quantitativa, permitindo identificar fatores que influenciam o desempenho dos municípios na política de proteção social especial. Sua pesquisa de mestrado tem como foco a análise da efetividade da Proteção Social Especial (PSE) nos municípios brasileiros. “Esses bancos de dados fornecem informações sobre o funcionamento dos conselhos, do Pacto de Aprimoramento, além de dados sobre infraestrutura, recursos humanos e financiamento dos Centros de Referência Especializados de Assistência Social (CREAS)”, explica.

Conforme a cientista da computação, especialista em gerenciamento de dados científicos e membro titular da Academia Brasileira de Ciências (ABC) Claudia Maria Bauzer de Medeiros, o acesso aberto por meio de repositórios de dados possibilita tanto a auditoria em termos de qualidade quanto a apuração de possíveis erros, resultando na melhora da pesquisa como um todo. Além disso, é uma forma de proporcionar a colaboração e a cooperação entre pesquisadores e pesquisadoras.

“Quando dados são acessados e reutilizados, há uma quantidade não desprezível de novos projetos que surgem com a parceria de quem os reutiliza. Outra vantagem, principalmente para a nova geração de cientistas, é a da preparação desses dados conforme os bons práticas de pesquisa. Para que os dados possam ser disponibilizados, eles têm de ser preparados”

— Claudia Maria Bauzer de Medeiros

Como a pesquisa ainda está em andamento, Jéssica ainda não fez o depósito de seus dados em repositórios abertos. Reconhece, no entanto, a importância de compartilhá-los para promover a transparência e permitir que outros pesquisadores possam replicar e expandir os estudos realizados. “Tenho a intenção de depositar os dados coletados durante minha pesquisa de mestrado em repositórios apropriados, conforme as diretrizes e políticas de compartilhamento de dados da minha instituição e dos próprios repositórios”, declara.

Uma das grandes questões sobre o uso de dados abertos é a sua confiabilidade, já que não há uma auditoria dos conteúdos depositados por pesquisadores e pesquisadoras. Trata-se de um sistema que tem como base a confiança na ética científica e a construção conjunta entre pares. Para Jéssica, os dados em repositórios abertos como os do SNAS são geralmente confiáveis, pois são mantidos por instituições governamentais e atualizados regularmente.

“No entanto, é sempre importante verificar a qualidade e a consistência dos dados, bem como estar ciente das possíveis limitações ou lacunas. Além disso, a transparência e a documentação fornecidas junto aos dados são essenciais para garantir a confiabilidade e a reprodutibilidade das análises realizadas”

— Jéssica Krislei Costa de Neque

### Uma prática ainda incipiente

Em 2018, a ABC criou o Grupo de Trabalho Ciência Aberta (GT-OS). A equipe é coordenada por Medeiros e tem como objetivo refletir sobre os benefícios, os riscos, as possibilidades e os desafios da Ciência Aberta no país, considerando as implicações no uso e na forma de disponibilização de dados.

Medeiros ressalta que a divulgação aberta dos dados de pesquisas com financiamento público depende de todo um ecossistema científico. “É preciso entender que Ciência Aberta é cara, tem custos inerentes para quem publica e para quem financia. A economia vem no longo prazo, no reuso de dados e softwares, na aceleração dos processos”, afirma.

Segundo relatório do GT-OS apresentado em novembro de 2023, a existência e o uso de repositórios abertos no Brasil e na América Latina ainda são muito incipientes. “Há uma ausência, pois para fazê-los se leva tempo. É necessária uma equipe dedicada, não pode ser um projeto abandonado em casos de mudança de governo ou de direção de uma instituição”, ressalta Medeiros. Ela aponta a necessidade de uma mudança no meio científico voltada para uma cultura de curadoria e de documentação de dados e softwares – transformação que deve ser internalizada pelas instituições e pelas próprias e pelos próprios cientistas e pesquisadoras.

Entre as iniciativas no Brasil, Medeiros destaca a Rede de Repositórios de Dados Científicos do Estado de São Paulo. Trata-se de projeto lançado em 2019 e realizado por designação da Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp). Sob sua coordenação, a Rede conta com a participação da Universidade Estadual de Campinas (Unicamp), da Universidade de São Paulo (USP), da Universidade Estadual Paulista (Unesp), da Universidade Federal de São Carlos (UFSCar), da Universidade Federal do ABC (UFABC) e da Universidade Federal de São Paulo (Unifesp). Inclui, ainda, o Instituto Tecnológico de Aeronáutica (ITA) e a Embrapa Agricultura Digital (CNPq/Embrapa).

“Essas instituições criaram seus comitês internos permanentes de gestão de repositório. O único caso não levado adiante foi o do ITA, porque eram projetos específicos. Os demais já estão depositando dados constantemente”, conta. “Esta é uma exigência da Fapesp, menos nos casos de restrições éticas ou legais. As universidades de São Paulo estão tendo de cumprir para que as e os pesquisadores tenham recursos ou relatórios aprovados.”

O depósito de dados considerados sensíveis nos repositórios da Rede é realizado pelas e pelos pesquisadores conforme a política de cada instituição participante. Coordenadora do repositório do Unicamp, Medeiros relata que em 2023 a universidade realizou uma mudança no Termo de Consentimento Livre e Esclarecido (TCLE), após concordância formal do Conselho Nacional de Ética em Pesquisa (Conep). “A Unicamp fez essa consulta devido às práticas de Ciência Aberta. O novo termo prevê que cada sujeito autorize ou não a publicação livre de seus dados para pesquisa, desde que anonimizados”, pontua. “Essa é a primeira coisa, do lado da pesquisa. Do lado da Unicamp, o responsável pelo depósito dos dados tem que confirmar que não viola a Lei Geral de Proteção de Dados (LGPD) e que a Comissão de Ética em Pesquisa aprovou o projeto. Na USP, na Unesp e na Embrapa há o mesmo tipo de protocolo”, completa.

Também sob a coordenação de Medeiros, o repositório da Unicamp obteve um crescimento sensível em 2024, com o depósito de dados oriundos de dissertações e teses. “Estamos com cerca de 330 conjuntos de arquivos e 6 mil downloads. Isso quer dizer que cada conjunto de arquivos registrou em média 20 acessos externos”, revela.

### Desafios para a implementação na UFRGS

A Pró-reitoria de Pesquisa (Propeq) da UFRGS ainda não tem uma ação específica para um repositório centralizado de dados abertos. De acordo com o vice-pró-reitor de Pesquisa, Jefferson Simões, no entanto, “é intenção induzir tal ação em diferentes unidades, mesmo que o número de projetos sobre o tema ainda seja restrito. Por outro lado, é necessário que haja [um volume] maior de demandas específicas sobre temas a serem disponibilizados”.

Ele enfatiza que a criação de um repositório para disponibilização dos dados gerados pelas pesquisas depende de recursos financeiros específicos dos órgãos de fomento à pesquisa (CNPq, Finep e Fapergs). Ressalta, ainda, que falta iniciativa por parte de pesquisadores e seus centros, núcleos e grupos de pesquisa.

“Ou seja, não adianta simplesmente dizer que é importante, tem que haver demanda e proposta. Os colegas têm que entender que essa é uma demanda crescente. Algumas áreas de conhecimento, tanto por questões políticas quanto dos financiadores, são mais atuantes. Dou até meu exemplo: o Tratado da Antártica exige que toda a pesquisa antártica seja aberta (ou seja, tenha os dados divulgados) após alguns anos, geralmente cinco a seis anos após o estudo ser realizado e os trabalhos, publicados”, relata Simões, que é professor titular de Glaciologia e Geografia Polar da UFRGS e pioneiro da ciência glaciológica no Brasil.

“Para deixar bem claro e não haver dúvida, nós temos aqui uma definição de Ciência Aberta que dá acesso a informações, aos periódicos, aos sete mil projetos de pesquisa registrados na Propeq, todos são abertos ao público, assim como todas as publicações geradas pelos pesquisadores da UFRGS. O que nós não damos por falta de recursos é o acesso aos dados, o que depende de uma estruturação que envolve TI, hardware, gestão e manutenção de dados e metadados ao longo do tempo. Isso é algo impossível sem recursos novos”

— Jefferson Simões

Atualmente, os pesquisadores da UFRGS enviam seus dados para repositórios nacionais e internacionais, gratuitos e consagrados, como descreve a professora do departamento de Ciências da Informação da Universidade Samile Vanz, cuja tese de doutorado versou justamente sobre a identificação dos requisitos internacionais de avaliação da confiabilidade de repositórios de dados de pesquisa.

“Cada um busca uma alternativa e submete. Mas seria muito importante para a universidade ter seu repositório próprio, pois os dados de pesquisa hoje são produtos gerados pela instituição, que é avaliada pelo volume de artigos publicados, e o volume de dados disponibilizados é um indicador em breve também será mensurado”, afirma. Para a professora, quanto antes a universidade criar o seu próprio repositório, mais terá condições de determinar qual é a criação dos seus pesquisadores e alunos em termos de volume de dados de pesquisa produzidos e disponibilizados.

O Lume foi criado há cerca de 15 anos e disponibiliza as publicações científicas produzidas na UFRGS, como artigos, teses e dissertações. Logo, é possível ter o controle documental da produção e a preservação desses documentos, mas não se sabe quantos conjuntos de dados ou quantos terabytes de dados de produtos da UFRGS existem, porque não há um repositório para esse fim.

“Não nos faltam opções de onde depositar de maneira gratuita, mas a universidade perde porque deixa de controlar o que produz, mensurar o quanto ela disponibiliza em termos de dados, coleta e disponibilização a outros pesquisadores. A gente já começa a ver a entrada em jogo desses e outros indicadores. Nos editais do CNPq e da Fapesp já está em vigor a obrigatoriedade de depósito dos dados de pesquisa coletados para projetos financiados pelas agências”

— Samile Vanz

O Lume, com 15 anos de operação bem-sucedida, já se estabeleceu como um dos repositórios institucionais de maior relevância entre as instituições federais de ensino superior no Brasil e é reconhecido mundialmente, conforme descrito em reportagem do JIJ em 2022. No entanto, é evidente que a UFRGS ainda precisa desenvolver uma política robusta para planejar, criar e manter um repositório de dados abertos. Tal política não só atenderia às necessidades institucionais, mas também valorizaria as pesquisas realizadas, ampliando a circulação do conhecimento científico e a visibilidade e o impacto dos trabalhos de seus pesquisadores.

### O cenário no Brasil e no mundo

Por enquanto ainda não há obrigatoriedade em depósito aberto de dados de pesquisa no Brasil e na América Latina, diferentemente de países como a França, por exemplo. Claudia Medeiros esteve em visita técnica ao país em fevereiro deste ano para acompanhar com a maioria das universidades está procedendo em relação ao assunto e como isso é controlado pelas agências públicas de fomento.

“Lá, todo projeto que tem financiamento das agências de fomento públicas obrigatoriamente precisa depositar seus dados em repositórios abertos. Não tem como não fazer, porque senão não há financiamento de pesquisa. No momento não são todas as universidades no Brasil, mas a grande maioria tem esse controle por universidade, por departamento, pelo governo”, ressalta. Ela frisa que há exceções para casos que envolvam privacidade, projetos sigilosos por questões de segurança nacional ou que envolvam pedido de patente.

Esse tipo de sistema não é novidade, e outros países já tratam a questão como política de estado: nos Estados Unidos, por exemplo, a obrigatoriedade do depósito de dados em repositórios abertos já começou a ser implementada. De acordo com a pesquisadora, é obrigatório para a área de saúde desde 2023. “Antes havia uma obrigatoriedade a partir de certo valor de financiamento e, a partir de 2023, inclusive os projetos submetidos à patente têm três meses para os dados serem publicados”, relata.

Alemanha, Holanda e Canadá estão implementando paulatinamente políticas nesse sentido. Na Europa, Medeiros destaca, além da França, a Alemanha, o Reino Unido, a Holanda, os países escandinavos e parcialmente Espanha e Portugal. “Mas eles têm muito mais dinheiro que a gente. A União Europeia está investindo bilhões de euros nisso. Lá, essa obrigatoriedade é implantada com esse financiamento. Na Ásia tem o Japão, Vietnã, Taiwan e Malásia, que estão em diferentes etapas de implementação. A China tem uma coisa enorme, mas é difícil de a gente entender. A Índia, até onde se sabe, é bem fechada. E na África, temos certamente a África do Sul.”

No Brasil, a única política de quase obrigatoriedade é a da Fapesp. “Desde 2020 é obrigatório ter um plano de gestão de dados. Como são de 20 mil a 22 mil projetos de pesquisa por ano, tem de ter esta gestão de dados”, explica Medeiros. No entanto, outras agências públicas de fomento de pesquisa, como a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) ainda não possuem essa exigência.

Sobre o desenvolvimento de repositórios próprios por outras instituições de pesquisa brasileiras, Medeiros reforça que “não adianta querer incorporar se não há dinheiro”. Ela reforça a necessidade de uma mudança cultural que envolva o treinamento de pesquisadoras e pesquisadores, a elaboração de políticas institucionais e o investimento na criação e manutenção de repositórios. Como alternativas às instituições que ainda não desenvolveram esse recurso, Cláudia recomenda que pesquisadores usem repositórios como Zenodo e FigShare e a estrutura do Open Science Framework (OSF). “Agora, se chegarmos a um volume muito grande de dados, não há instituição no mundo que dê conta disso” alerta.

Em relação à América Latina, Medeiros aponta que existem várias iniciativas governamentais, como se tem no Brasil com o Plano de Governo Aberto. No entanto, assim como aqui, a adesão depende de cada instituição. O Chile já conta com quatro universidades com repositórios próprios, e países como Argentina, Colômbia, Peru e Uruguai estão avançando nesse sentido. “O diferencial da Europa é a União Europeia, que abraçou isso por volta de 2017. No Canadá, nos Estados Unidos eram agências de fomento isoladamente, mas agora se conectando às universidades”, reflete Medeiros.

Já Karla Avanço, gerente de comunidade do Open Scholarly Communication in the European Research Area for Social Sciences and Humanities (Peras), entende que a Ciência Aberta desafia estruturas de poder estabelecidas. “Em primeiro lugar, porque facilita equidade, inclusão. Também porque valoriza a bibliodiversidade e o multilinguismo, o que favorece a publicação em línguas nacionais.” Ela salienta que a América Latina é extremamente desenvolvida em Ciência Aberta graças a organizações como SciELO e Redalyc, inspirando movimentos europeus e participando de projetos colaborativos.

Sobre a abertura de dados de pesquisas científicas, chama atenção para a regra “tão aberto quanto possível, tão fechado quanto necessário”. Como exemplo de cooperação e de compartilhamento de dados e do contexto da covid-19. “Este será um bom exemplo por muito tempo. Cientistas no mundo todo colaboraram compartilhando dados e resultados que nos ajudaram a entender o vírus”, lembra.

O SciELO, inclusive, disponibiliza uma relação de bancos de dados abertos classificados por áreas e subáreas do conhecimento. Trata-se de uma lista atualizada desenvolvida por que editores de periódicos científicos possam indicar a quem onde depositar e compartilhar seus conjuntos de dados, códigos, métodos e materiais utilizados na concepção e realização de suas pesquisas, que também ficam disponíveis em sites e outros pesquisadores.

### Ciência Aberta

No Brasil, a Ciência Aberta adota um plano governamental de ação iniciado em 2018 pela Controladoria-Geral da União e pelo Ministério da Transparência por meio da adesão à Parceria para Governo Aberto, ou Open Government Partnership (OGP). Estabelecida em 2011, a OGP é uma iniciativa integrada por 75 países e 104 governos locais. Formada por representantes de governos, sociedade civil, empresas e organizações sem fins lucrativos, tem apoio e supervisão de funcionários e pesquisadores do Independent Reporting Mechanism (IRM). Seu objetivo é incentivar globalmente práticas relacionadas à transparência pública, à participação social, à responsividade e à inovação.

Karla Avanço salienta que a Ciência Aberta defende a integridade de pesquisa, ou seja, a realização de pesquisas por meios de acordos com princípios fundamentais de confiabilidade, honestidade, cuidado e transparência. “Podemos garantir essa integridade por meio de treinamentos e do estabelecimento de políticas adequadas para prevenir e abordar más condutas. A ciência aberta possibilita o compartilhamento de dados, códigos, métodos, ou seja, dos diferentes tipos de resultados. Isso favorece a replicabilidade da pesquisa, além de dar maior visibilidade ao que foi produzido e maiores possibilidades de colaboração”, afirma.

### :: Posts relacionados



Pesquisa investiga os impactos do trabalho remoto em servidores da UFRGS no contexto da pandemia | Adoção da Ciência Aberta no Brasil enfrenta resistências de dentro da comunidade acadêmica | Perspectivas científicas africanas resistem à desvalorização e ao apagamento | Estudo analisa o impacto do envelhecimento populacional no crescimento econômico dos municípios bras...

### ÚLTIMAS



Carta aos leitores | 22.08.24



Carta aos leitores | 15.08.24



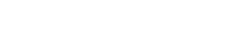
Desinformação científica é um problema público que atravessa fronteiras



Tecnologia e impactos da energia limpa H2V



Servidores com deficiência nas universidades



Prevalência de Diabetes mellitus em Angola



Adoção da Ciência Aberta no Brasil enfrenta resistências de dentro da comunidade acadêmica



Carta aos leitores | 08.08.24



A importância de recuperar o patrimônio cultural e histórico de Porto Alegre



Resíduos de alimentos podem ser utilizados para produção de embalagens biodegradáveis ativas