

PERSPECTIVA

Curadoria de dados de pesquisa: notas sobre uma abordagem mais sustentável

Por **Renata Curty**
rcurty@ucsb.edu

Meu primeiro contato com a área de curadoria de dados de pesquisa foi ainda durante o doutorado. Buscava explorar os fundamentos e a literatura ainda em formação, tendo em vista que tinha interesse no comportamento de reuso.

Em 2011, vivíamos um momento muito entusiasmado com relação à ciência aberta, mas também apreensivo quanto às políticas e mandados de compartilhamento.

Poucos anos depois, intensificavam-se as diretrizes por parte das agências de fomento e dos editores científicos, que mais tarde culminaram com uma alta demanda de formação e capacitação de profissionais especializados em dados.

Data stewards, *curators* e *data managers* eram os profissionais mais requisitados nos anúncios das listas de discussão da área. Escolas de Ciência da Informação (CI) e Biblioteconomia passaram a dar mais atenção ao nível mais elementar da conhecida pirâmide de Davenport e Prusak, antes mais



Renata Curty é Research Data Specialist/Facilitator no Research Data Services Department da UCSB Library na University of California, Santa Barbara.

PERSPECTIVA

abreviado em suas grades curriculares. As universidades e centros de pesquisa começaram a formar departamentos para oferecer suporte aos pesquisadores, muitas vezes como parte da estrutura organizacional de suas bibliotecas.

Recordo-me das incertezas sobre a infraestrutura que acolheria esses dados e se, de fato, eles teriam efetivo valor de reuso. Afinal, a reprodutibilidade e o reuso de dados eram expostos como os principais motivadores e argumentos em prol das iniciativas da ciência aberta, mesmo ainda com raros exemplos de real impacto.

Os incentivos para o compartilhamento eram discutidos sob um ponto de vista mais altruísta e abstrato, e havia maior enfoque nas consequências negativas do não compartilhamento.

Também havia pouca evidência ou compreensão, sob a ótica daqueles que haviam experimentado reutilizar dados produzidos por terceiros, dos gargalos que dificultavam que novas pesquisas pudessem ser desenvolvidas a partir de dados preexistentes.

Apesar de mais suposições do que certezas, a curadoria já era apontada como essencial para a viabilização da ciência aberta e a principal via para uma melhor qualidade dos dados disponíveis em repositórios, que expandiram rapidamente.

Diante daquele cenário de intensas mudanças relacionadas ao meu tema de pesquisa, confesso ter experimentado por vezes a "síndrome do impostor". Embora me sentisse confortável com a teoria e acompanhasse e contribuísse com as discussões no tema, por vezes me questionava: como esses fundamentos e teorias se materializam na prática? Atribuo esse desassossego ao meu viés mais pragmático.

Avançando para o ano de 2020, quando ingressei na minha atual posição, em que a curadoria e a gestão de dados de pesquisa precisavam não apenas ser debatidas, mas principalmente, executadas. Fomos o primeiro campus do sistema a integrar o fluxo de curadoria do repositório Dryad e nos filiamos à Data Curation Network (DCN), o que nos possibilitou trocar experiências e aperfeiçoar nossos fluxos e processos.

Foi nesse período também que ingressei no comitê consultivo da DMPTool e passamos a oferecer consultorias, workshops e outros serviços aos pesquisadores, de modo que seus dados atingissem conformidade FAIR. Continuamos nessa mesma linha de atuação.

Ainda há muito o que ser feito e a ciência aberta ainda encontra entraves e desigualdades disciplinares, mas um avanço óbvio percebido durante as interações com os pesquisadores é que

PERSPECTIVA

não mais temos necessidade de despendar tempo para conscientizá-los sobre a importância do tema. Atingimos uma fase em que os “porquês” são claros para a comunidade científica, mas ainda temos um caminho sobre o “como”.

Sem me alongar mais na contextualização, gostaria de compartilhar algumas observações iniciais e iniciativas acerca da curadoria de dados a partir de trocas e experiências (erros e acertos), dessa vez não mais tão distantes da aplicação.

Entendo que a curadoria não deve ser tratada como uma atividade ao final do processo – elaborarei sobre isso mais adiante –, mas com a ressalva de que o suporte prestado aos pesquisadores e a nossa participação no processo geralmente não ocorrem de ponta a ponta como aprendemos na teoria.

Nas diversas representações do ciclo de vida dos dados, a etapa inicial é o planejamento. Entre os serviços que oferecemos nessa fase, destaca-se o aconselhamento e revisão dos planos de gerenciamento e compartilhamento de dados previamente à submissão às agências de fomento.

A participação nesse processo, que tem aumentado dada a nossa parceria com o *Office of Research* do campus, nos permite identificar e

antecipar os tipos, volumes e características dos dados a serem produzidos, assim como as demandas relacionadas à infraestrutura para dar suporte à pesquisa.

Isso inclui as ferramentas necessárias para coleta, análise, armazenamento e segurança dos dados, fundamentais para o desenvolvimento dos projetos.

Ao longo desse processo, também atuamos como mediadores entre diferentes unidades e setores do campus, que por vezes operam de modo descentralizado. Por exemplo, quando identificamos que o projeto planeja gerar dados sensíveis ou de seres humanos, podemos servir de ponte para serviços de armazenamento com maiores níveis de segurança e indicar ferramentas para a anonimização ou sanitização dos dados, bem como planejar workshops no tema.

Além disso, orientamos os pesquisadores em questões relacionadas ao planejamento quanto a documentação, organização, acesso, licenciamento dos produtos resultantes da pesquisa (dados, scripts/software, imagens, modelos, etc.), bem como opções para guarda, preservação e publicação. Tudo isso ainda na fase do planejamento, e sem a certeza de que o projeto receberá financiamento, a curadoria, portanto, nessa fase, ainda é hipotética.

PERSPECTIVA

Fato é que, a partir da aprovação, a implementação do plano estará a cargo do pesquisador principal e de seu time. O modelo de acompanhamento direto de projetos é insustentável mesmo entre as instituições mais bem equipadas. Nem sempre as consultorias são escalonáveis, pois as demandas são específicas e requerem conhecimento especializado, seja na área do conhecimento, no tipo de dados e nos sistemas utilizados.

A partir de nossa experiência e da discussão com os pares, temos concluído que a melhor alternativa seria investir em programas de letramento de dados (*data literacy*) e gestão de dados de pesquisa adaptados para os diferentes níveis e áreas do conhecimento, e em parcerias com os diferentes departamentos.

Entendemos que nosso papel e contribuição principal estão em multiplicar boas práticas por meio de programas que capacitem embaixadores e sigam o modelo de "formação de formadores", capazes de exercer a curadoria e gestão de dados tal qual planejada, com a devida especialidade, de forma contínua e acompanhada durante todo o ciclo de vida dos dados e do projeto. Portanto, temos um papel mais consultivo e educativo do que executivo no âmbito da curadoria propriamente dita.

Tratamos do início e do meio do ciclo de vida dos dados, mas e o fim? Haveria ainda um papel para as bibliotecas na curadoria dos dados de pesquisa na etapa do arquivamento dos dados? Sim, é possível, mas eu e meus pares também temos observado a insustentabilidade do modelo convencional de curadoria passiva em que coleções de dados chegam, por vezes, com falhas de documentação e que a nossa intervenção para tornar os dados mais FAIR, principalmente no quesito de reusabilidade que dependem de verificação mais aprofundada dos dados são limitadas. Isso pode acontecer devido à falta de especialização, como exemplificado anteriormente, ou porque estamos agindo muito tardiamente no processo.

Solicitar modificações e retrabalho dos pesquisadores quando muitas vezes eles já estão concentrados em seus próximos projetos tem se mostrado pouco eficaz. Portanto, ações educativas e consultivas proativas durante as etapas em que os pesquisadores estão lidando mais diretamente com a manipulação e processamento dos dados seriam mais benéficas.

Isso pode envolver fornecer orientações desde o início do projeto para garantir a qualidade dos dados e a documentação adequada. Além disso, é importante oferecer suporte

PERSPECTIVA

contínuo e acompanhamento para garantir que os dados sejam adequadamente preparados para arquivamento, possibilitando sua reutilização futura de maneira mais fácil e eficiente.

As ações educativas e consultivas mencionadas, aliadas às oportunidades decorrentes do avanço das ferramentas de inteligência artificial e aprendizado de máquina para a "FAIRtificação", tais como a validação de arquivos e a complementação da documentação dos dados, como a geração de arquivos README, prometem elevar a qualidade e a sustentabilidade da curadoria de dados a médio e longo prazo. Mais um ciclo de mudanças acompanhadas de algumas incertezas e muito entusiasmo.

Sobre a autora

[Renata Curty](#)

Research Facilitator/Data Specialist no Research Data Services Department da University of California, Santa Barbara. Atuou como Professora Adjunta (2015-2020) e Professora Assistente (2006-2015) no Departamento de Ciência da Informação da Universidade Estadual de Londrina (UEL), na área de

Tecnologia da Informação exercendo atividades de pesquisa, ensino e extensão nos seguintes temas: reuso de dados em ciência, comunicação e métricas científicas, repositórios digitais, acesso aberto, automação de bibliotecas, preservação digital, arquitetura da Informação, e usabilidade.

Graduada em Biblioteconomia pela Universidade Estadual de Londrina. É Mestre em Ciência da Informação pela Universidade Federal de Santa Catarina - Bolsista CAPES. Possui titulações de Master in Philosophy (MPhil) e Philosophy Doctor (PhD) e em Information Science and Technology pela School of Information Studies (Syracuse University, NY) - Bolsista CAPES/Fulbright.

Redação e Foto: Renata Curty

Diagramação: Herta Maria de Açucena do Nascimento Soeiro