



## Apresentação

### Foreword

**Luana Farias Sales \***

**Luis Fernando Sayão \*\***

Nos cenários da ciência contemporânea, as coleções de dados deixam de ser meros subprodutos da pesquisa e se tornam protagonistas de novos conceitos de investigação científica e de geração de conhecimento. A pesquisa orientada por dados se alinha e se contrapõe, ao mesmo tempo, aos métodos tradicionais de pesquisa criando perspectivas inimagináveis para o progresso da ciência. Esse movimento dialético, que tem como catalisador as tecnologias digitais, as redes de computadores e os algoritmos sofisticados, vai de uma forma disruptiva e avassaladora, redefinindo a ideia secular de metodologia científica e de construção de hipóteses. Seus postulados em constante transformação redesenham os fluxos de geração de conhecimento e criam uma cartografia inédita para os percursos da comunicação científica, tornando, dessa forma, a ciência e seus processos mais transparentes e reprodutíveis, mais interdisciplinares e transversais, mais abertos e participativos, tudo isso sob o signo da velocidade.

Nesse contexto científico em transição, cujo desdobramento ainda não se pode prever com precisão, os dados são requalificados como ativos informacionais de primeira grandeza que guardam o potencial de serem compartilhados, reinterpretados e reusados para novas pesquisas. A partir dessa constatação começa a se esboçar a noção de que o avanço da ciência, para muitos domínios, está proximamente relacionado à curadoria, processamento, análise e disseminação de grandes coleções de dados de pesquisa e de seus códigos, instrumentos e métodos. Nessa direção, há disciplinas híbridas inteiramente orientadas por dados. Porém, esta nova ordem não se limita a essas disciplinas, mas vai se instalando de forma pervasiva em todos os domínios do universo científico, das ciências exatas às humanidades, artes e cultura.

A importância dos dados de pesquisa, bem como as possibilidades que eles oferecem na formação de uma nova ecologia de conhecimentos interligados, impõe, entretanto, o desafio de gerenciá-los para que eles se tornem finalmente úteis e inteligentemente compartilháveis e que criem memórias científicas dinâmicas e confiáveis para o futuro reuso. Nessa direção, as plataformas de gestão de dados necessitam de: 1) implementação de conceitos importantes e sofisticados como

---

\* Doutora em Ciência da Informação pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Professora do Programa de Pós-Graduação em Ciência da Informação do Ibict. Endereço: Rua Lauro Muller, 455, 4º andar, sala 408, CEP: 22.290-160. Telefone: (21) 97112-7411 / (21) 3873-9450. E-mail: luanasales@ibict.br.

\*\* Doutor em Ciência da Informação pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict). Tecnologista em Ciência e Tecnologia da Comissão Nacional de Energia Nuclear (CNEN). Professor do Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro (Unirio). Endereço: Rua General Severiano, nº 90, Botafogo, CEP 22290-040, Rio de Janeiro, RJ. Telefone: (21) 2173-2028. E-mail: lsayao@cnen.gov.br.

interoperabilidade, integridade, autenticidade, curadoria, reconhecimento de autoria, controle de versões, preservação de longo prazo; 2) espaços colaborativos e de comunicação que viabilizem as novas formas de socialização entre cientistas, como apoio a revisão por pares aberta, e de interlocuções com outros segmentos da sociedade, como escolas, hospitais e governo; 3) políticas que permeiem e ordenem seus fluxos e que estabeleçam uma necessária conformidade ética e legal para a disseminação de dados científicos; 4) serviços inéditos que atendam às novas demandas e às idiosincrasias tradicionais das comunidades acadêmicas; 5) interlocução e o alinhamento aos movimentos internacionais, posto que as pesquisas e seus dados não são limitados por fronteiras geográficas e políticas. Assim sendo, se tornam essenciais a participação do Brasil na iniciativa GO FAIR, cujo objetivo é promover por meio da sugestão de alguns princípios que os dados se tornem mais facilmente localizáveis, acessíveis, interoperáveis e reutilizáveis, ou seja, FAIR, facilitando, dessa forma, o seu reconhecimento por máquinas e pessoas; e a participação do Brasil na Open Government Partnership, através do Compromisso número 3 do 4º Plano de Ação Nacional em Governo Aberto, que tem como objetivo é estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil.

Tudo isso somado significa que necessitamos de gestão, tecnologias e padrões e institucionalização que apoiem o desenvolvimento de plataformas abertas e interoperáveis que mantenham os dados de pesquisa úteis, dinâmicos e sustentáveis no tempo e no espaço.

Os desafios são imensos e instigantes e começam com a compreensão do que é dado de pesquisa, ou melhor, de como algo em um determinado contexto científico se torna dado de pesquisa até o seu arquivamento em sistemas confiáveis. Contudo, esses desafios chegam em boa hora para a Ciência da Informação, pois abrem oportunidades extraordinárias de interlocução com outras disciplinas e se harmonizam com os pressupostos de colaboração, compartilhamento e da multi-inter-transdisciplinaridade que estão no coração da ciência aberta. Isto porque a gestão e curadoria de dados de pesquisa é uma ação colaborativa que cria um arco que une os saberes e as práticas de pesquisadores, cientistas da informação e cientistas da computação. Ela só se efetiva plenamente com uma colaboração orgânica entre esses três mundos: o do laboratório, onde o cientista na sua aventura cotidiana de erros e acertos busca novos conhecimentos, e depois se debruça na faina monótona de refinar, limpar e qualificar os seus dados para que eles deixem de ser locais e implícitos e se tornem compreensíveis e interpretáveis universalmente; o da biblioteca de pesquisa onde os cientistas da informação, bibliotecários e arquivistas de dados, ressignificam os princípios de seus domínios, antes voltados para ativos tradicionais, e recriam serviços e plataformas e modelos de gestão voltadas para as especificidades dos dados e de seus fluxos cada vez mais disciplinar; e o da computação, onde os analistas e programadores com suas máquinas, algoritmos, modelos e programas conseguindo expandir e revelar padrões, relações e conhecimentos ocultas na profusão de dados.

Como vemos, os dados de pesquisa têm muitas faces, mas nenhuma delas revela completamente a verdadeira dimensão de sua complexidade. Querer compreender os seus enigmas tomando apenas uma de suas faces pode ser inútil. A diversidade dos dados pode estar espalhada em termos de disciplina de origem, tecnologias subjacentes, nível de estruturação, modelos, modos de coleta e muito mais. Essa multiplicidade de dados pode ser analisada pelo prisma dos grandes projetos da Big Science que disseminam petabytes de dados ordenados e homogêneos, gerados e

coletados por grandes instrumentos, distribuídos globalmente por computação em grade e compartilhado por milhares de pesquisadores simultaneamente. E podem ainda ser analisados no contexto fragmentado e multidisciplinar dos pequenos laboratórios da cauda longa da pesquisa, que gera poucos gigabytes de dados, desordenados, heterogêneos e coletados manualmente, que são, porém, independentes, criativos e inovadores.

Os artigos submetidos a este número especial da Liinc espelham essa diversidade de abordagens necessárias à compreensão da natureza multifacetada dos dados e da aplicação desse entendimento na construção de sistemas abertos voltados para apoiar a gestão e compartilhamento e reuso desses ativos informacionais. Fica claro que é a ótica multidisciplinar que vai possibilitar o endereçamento correto das questões práticas e teóricas que envolvem o longo e diversificado ciclo que se inicia antes da pesquisa começar e não acaba quando o projeto se encerra. Não será possível seguir adiante sem uma convergência de experiências de diversas áreas, tecnologias e ações gerenciais e políticas. Assim sendo, convidamos nossos leitores a vivenciar um pouco mais algumas das faces dessa complexidade que cria um ponto de inflexão nos fluxos de pesquisa e na comunicação científica.