

A ciência como um empreendimento aberto

Jessica Bland

A comunicação aberta sempre fez parte da ciência, mas existem novas formas de compartilhar dados. O relatório *Science as an open enterprise*, da Royal Society, destaca a necessidade de lidar com a avalanche de dados de interesse científico disponibilizada por meio das novas tecnologias, a fim de preservar o princípio da abertura e explorar os dados de uma maneira que tenham potencial para deflagrar uma nova revolução científica.

O que em parte motivou a Royal Society a fazer esse estudo foi uma grande controvérsia que houve no Reino Unido há um ano e meio, em 2009: o *climategate*. *E-mails* enviados por pesquisadores foram *hackeados* e alguém os publicou. Foi um *wikileaks* da ciência, por assim dizer. E os *e-mails* sugeriam que alguns dos cientistas haviam tentado esconder dados dos céticos da mudança climática. A razão para isso se tornar polêmico é que nós esperamos que os cientistas sejam abertos ao debate e ao ceticismo. Esse acontecimento suscitou uma série de outras questões para a análise da ciência no século XXI. O relatório, lançado em junho passado, aborda oportunidades e desafios e foi produto das discussões com *experts*, incluindo gente da indústria, das ciências sociais, das ciências da computação, das ciências do clima.

Vejamos um caso que aconteceu em maio de 2011, em Hamburgo, na Alemanha. Houve um surto de infecção intestinal causado pela *Escherichia coli*, que se espalhou rapidamente por toda a Europa, afetando 400 mil pessoas. Todas as vítimas testavam positivo para uma determinada cepa de *E. coli*. Os médicos não sabiam o que fazer com isso lá em Hamburgo. O que acontecia com essa *E.coli*? Ela parecia tão semelhante às outras cepas – por que então provocava essa infecção? A busca de solução levou a uma forma bastante aberta de cooperação. Em primeiro lugar foram abertos os dados sobre o genoma da cepa de *E.coli*. Isso foi publicado num *site* para que qualquer pessoa em qualquer parte



Jessica Bland é conselheira de política científica da Royal Society. Formada em Física e Filosofia pela Universidade de Oxford, fez mestrado em ciências da comunicação no Imperial College

do mundo pudesse acessar. Em três semanas cerca de 200 relatórios foram publicados sobre o que deveria ser feito para impedir a epidemia e seus efeitos, e os resultados foram usados para controlar o surto. Isso foi possível graças a um método bastante aberto de fazer ciência e a *expertise* de outros países, que levaram a uma solução para uma questão de saúde pública em alguns meses. Essa é uma história muito forte, do tipo de conquista que buscamos usando uma ciência mais interativa.

O relatório foi produzido depois de um ano de debate e precisamos ser cuidadosos com a linguagem e com o entusiasmo que temos sobre essa forma de fazer ciências, que costumamos chamar de “abertura inteligente”. Nós temos que abrir os dados de uma forma em que as outras pessoas possam usá-los, seja para efeitos de saúde pública, para a indústria ou outro tipo de aplicação. A abertura não é, por si só, uma coisa útil. É preciso abrir de uma forma inteligível. Os quatro critérios que nós temos que seguir são, em primeiro lugar, tornar acessíveis os metadados. Em segundo lugar, eles precisam ser compreensíveis. Em terceiro, é necessário expor um contexto, a fim de que as pessoas que forem usá-los entendam como eles foram recuperados, o quanto são confiáveis, ou seja, isso é para o propósito de uma revisão por pares. E, por fim, os dados precisam ser reutilizáveis ou replicáveis. Só se esses quatro critérios forem cumpridos é que nós teremos uma abertura dos dados apropriada. A abertura é cara, e isso é um problema, porque precisamos ter esses critérios atendidos para cada um dos vários públicos que vão utilizar os dados.

Houve uma discussão muito grande no Reino Unido e na Europa sobre o conceito de “abertura inteligente”, sobre como as indústrias podem usar esses dados e isso gerar um grande desenvolvimento econômico. Quais são os limites dessa abertura? Há interesses comerciais legítimos que também temos que proteger? O quanto podemos



Cooperação e abertura de dados geraram artigos científicos e trouxeram solução rápida para epidemia de *E. coli*

ser abertos? Nós temos alguns exemplos, como o do Instituto Europeu de Bioinformática. Eles têm um mecanismo capaz de permitir às empresas que comparem e cruzem informações de seus bancos de dados internos com os nossos grandes bancos de dados, sem que ninguém realmente consiga ver os dados do outro. É uma abertura, mas dentro das limitações do contexto comercial.

A outra questão é quanto à utilização de seres humanos. Evidentemente, esse tipo de informação não pode ser liberada, pois seria uma invasão de privacidade. Também há questões envolvendo segurança. Houve uma controvérsia sobre o vírus H1N1, quando se encontrou uma nova forma de vírus que era altamente contagiosa. A questão era se esses trabalhos sobre o vírus H1N1 deveriam ser publicados ou não, porque bioterroristas poderiam usá-los. No final foi decidido que deveriam ser publicados, sim, porque não haveria tanta gente capaz de utilizá-los para terrorismo.

Quando falamos da transição para um ambiente de pesquisa em que os dados são mais abertos, estamos falando de um sistema bastante complexo. A ideia de uma pirâmide ajuda a entender como tentar dar conta disso. Quanto mais alta a camada dessa pirâmide, maior a responsabilidade e a demanda para acesso. Na base dessa pirâmide, estão os dados individuais e pessoais, que, para muitos pesquisadores, devem ser mantidos num arquivo – ninguém quer abrir esses dados para o mundo. Temos, então, uma grande base na nossa pirâmide, que são

as coleções individuais mantidas pelos pesquisadores. Parte disso poderia ser útil para todos, mas não podemos esquecer que isso ainda existe. Na hora que passamos para cima, na camada seguinte, podemos ver as universidades gerenciando seus bancos de coletas de dados no Reino Unido. Temos universidades que competem muito fortemente em número de artigos e em dados que

Não basta abrir os dados. Eles precisam ser inteligíveis para todos os públicos

publicam – esses dados pertencem às instituições. Então eles tentam manter isso restrito em repositórios institucionais. A camada que vem em seguida é a de coleções de dados nacionais. E no topo são os recursos de dados internacionais, como, por exemplo, o banco de dados mundial de proteínas, que abriga dados coletados ao longo dos anos.

Para simplificar, queremos ver todos os dados ativos *on-line*, queremos que isso opere de forma conjunta. Os dados são parte integral da ciência e precisam

ser comunicados dessa forma, e não só incluídos em artigos. Nossa aspiração é que toda a literatura científica esteja disponível *on-line*, todos os dados primários sejam acessíveis *on-line*. Para colocar isso em termos concretos, podemos dizer que há seis prioridades. A primeira é mudar a cultura vigente que considera dados científicos como propriedade particular. O segundo é dar crédito, no processo de

avaliação da pesquisa, à comunicação de dados úteis e a novas formas de colaboração. O terceiro é criar padrões comuns para a comunicação de dados. O quarto é fomentar o que chamamos anteriormente de abertura inteligente de dados. O quinto é o fortalecimento do grupo de cientistas que trabalham com dados. Não temos muitos engenheiros de computação capazes de fazer isso no Reino Unido, então é uma prioridade urgente no momento. E o sexto

é o desenvolvimento de novos *softwares* capazes de automatizar e simplificar a criação e a exploração de conjuntos de dados. Caso se vá investir em algo, hoje, é isso que precisa de mais investimento. Espero que uma questão como essa, que é fundamental para os pesquisadores, seja tratada em profundidade no Fórum Mundial de Ciência. E que nós possamos criar algo que seja mais do que uma ferramenta para lidar com um pequeno surto epidêmico, como aconteceu na Alemanha no ano passado. ■