



ARTIGO

SOFTWARE LIVRE: PRÉ-REQUISITO PARA A CIÊNCIA ABERTA

POR

Christina von Flach e Fabio Kon
flach@ufba.br, kon@ime.usp.br

A ciência sempre se pautou pela disseminação do conhecimento, com destaque para a reprodutibilidade de experimentos. É claro que há limites práticos – e mesmo éticos – para o que pode ser compartilhado e reproduzido, mas expandir esses limites é um objetivo de interesse para cientistas de todas as áreas. Esse é o foco do movimento pela ciência aberta, que tem tido bastante repercussão nos últimos anos.

A Ciência Aberta requer que as ferramentas e os instrumentos necessários para a prática científica estejam dis-

poníveis para todos, de modo que os experimentos possam ser reproduzidos e os resultados verificados por terceiros. No Século 21, o software consolidou-se como onipresente no conjunto de ferramentas usado por pesquisadores das Ciências Exatas e Biomédicas e tem uso crescente nas Ciências Sociais e Humanas. Software é usado para limpeza, processamento e visualização de dados, bem como para criar modelos e realizar previsões. Algoritmos especializados são codificados na forma de bibliotecas, scripts e metadados. Sem compartilhar todos esses artefatos, é muito difícil e custoso reproduzir a pesquisa científica e validar sua correção.

Assim, para alcançar a abertura plena na Ciência, é essencial que bibliotecas de software, scripts, pacotes e aplicativos usados e desenvolvidos por cientistas estejam prontamente disponíveis para qualquer pesquisador interessado em reproduzir ou estender uma determinada parte da pesquisa. Em particular, quando se trata de financiamento público, é altamente desejável que o software produzido no contexto da pesquisa (**research software**) seja **Software Livre** para garantir que seus benefícios sejam disponibilizados para uma vasta comunidade e de forma abrangente.

O **Movimento do Software Livre** teve início em 1983 quando o projeto GNU foi fundado como um meio de desenvolver um corpo robusto de software, incluindo um sistema operacional, ferramentas e aplicativos associados que estariam completamente disponíveis para todos os interessados. Em 1998, o termo **Open Source Software** passou a ser utilizado como uma estratégia para difundir, para empresas, essa nova forma colaborativa de desenvolver e compartilhar software. Atualmente, o software livre é reconhecido e adotado pelas principais empresas de TI do mundo, que usam e produzem software sob licenças de software livre/aberto. Milhares de componentes de software de alta qualidade estão disponíveis como software livre, desde drivers e sistemas operacionais até bibliotecas, arcabouços e aplicativos.

A **Free Software Definition** (<https://www.gnu.org/philosophy/free-sw.html>) estabelece o que é Software Livre em termos de quatro liberdades dadas aos seus usuários: (1) para executar o software

para qualquer propósito, (2) para estudar e modificar o software (o acesso ao código fonte é uma pré-condição para isso), (3) para redistribuir cópias do software e (4) para distribuir suas versões modificadas do software. Essas quatro liberdades alavancam os princípios da **Ciência Aberta**. Por exemplo, a permissão para executar o software para qualquer finalidade permite que pesquisadores utilizem o software existente sem ter que comprá-lo ou construí-lo do zero para realizar seus próprios estudos. A permissão para estudar e modificar o software na forma de código-fonte oferece suporte à reprodutibilidade e replicabilidade, divulgando software de pesquisa, artefatos relacionados e o conhecimento embutido neles. Também aumenta a transparência (**fluxos de trabalho visíveis**), auditabilidade e confiabilidade (os resultados podem ser verificados por terceiros e qualquer pessoa pode detectar e corrigir um erro ou um recurso malicioso). A permissão para redistribuir cópias possibilita o compartilhamento de **pacotes de replicação** que, além dos dados brutos, fornecem o código necessário para seus experimentos de análise e interpretação em diferentes ambientes. Finalmente, a permissão para distribuir versões modificadas do software permite que os pesquisadores desenvolvam seu próprio trabalho, reutilizando e expandindo o fluxo de trabalho, a base de código ou a ferramenta de alguém, de modo a compartilhar o novo conhecimento para o benefício de outros.

As **licenças de software livre** correspondem à definição mencionada anteriormente e, como tal, alavancam a Ciência Aberta. Para que seja

compartilhado como Software Livre (*Free and Open Source Software*), um trecho de código deve idealmente ser distribuído sob uma licença formalmente aprovada pela **Open Source Initiative** (<https://opensource.org/licenses>), de preferência uma que seja rotulada como “popular e amplamente utilizada ou com comunidades fortes”, tais como a *GNU General Public License* (GPL) ou a *Mozilla Public License 2.0*.

Projetos de código aberto bem organizados adotam um modelo de desenvolvimento que resulta em código de alta qualidade que se adapta rapidamente a diferentes situações. Várias práticas estabelecidas são usadas e reconhecidas como contribuições importantes para a manutenção de software de alta qualidade. Estas incluem o uso de repositórios públicos, sistemas de controle de versão, colaboração por pares, revisão de código, testes automatizados, formatos e interfaces padrão e documentação relevante, e podem ser consideradas boas práticas para a Ciência Aberta.

Os sistemas de controle de versão permitem que várias versões do mesmo software sejam mantidas e, possivelmente, referenciadas por experimentos de pesquisa e artigos científicos. Repositórios abertos podem ser usados para compartilhar vários tipos de ativos de pesquisa, como algoritmos, dados, código, relatórios e fluxos de trabalho, apoiando a reprodutibilidade, reduzindo a redundância e promovendo a colaboração científica aberta. A colaboração entre pares é frequente e habilitada por meio de acesso compartilhado ao código-fonte e vários canais de comunicação. A revisão de código promove a melhoria da qualidade do software

por meio de compartilhamento, colaboração e revisão por pares e pode ser aplicada a outros ativos de pesquisa. Os testes automatizados aumentam a confiabilidade e a facilidade de manutenção, além de promover agilidade no desenvolvimento de novos recursos. O uso de formatos de dados e de interfaces padronizados facilita a integração com outros sistemas e desencoraja a dependência de fornecedores específicos. A documentação frequente e contínua é uma prática recomendada para manter os guias do usuário, os manuais e os outros documentos relevantes atualizados em relação à versão mais recente do software. Essas, e outras práticas usadas pelas comunidades de software livre, podem ser consideradas boas práticas para a Ciência Aberta, uma vez que oferecem suporte contínuo à disponibilidade e reúso de ativos, colaboração entre pares, transparência do fluxo de trabalho e confiabilidade.

O software desenvolvido ou utilizado no contexto de uma pesquisa científica deve ser facilmente localizável, acessível, interoperável e reutilizável (<https://bit.ly/3mGZass>). Nesse contexto, os ecossistemas de software livre também promovem princípios da Ciência Aberta. Software livre pode ser localizado em repositórios com base em identificadores e descritores, utilizando diversos critérios como palavras-chave, linguagem de programação, versão do software, entre outros. A acessibilidade é encorajada em software livre disponível em repositórios abertos, com licenças de compartilhamento explícitas e bem definidas e documentação associada. A definição de interfaces de programação e formatos de entrada/saída e o uso de padrões promovem a interoperabilidade

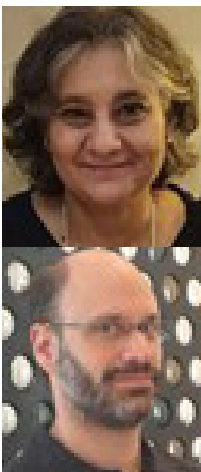
e o reuso por vários grupos de pesquisa em todo o mundo. Além disso, no ecossistema da Ciência Aberta, o software deve ser sustentável, citável e reconhecido como uma valiosa contribuição da pesquisa científica, tão importante quanto os artigos, dados e metadados. A adoção do modelo de software livre promove a sustentabilidade do software no longo prazo. No entanto, esforços e incentivos são necessários para que os créditos para o software se tornem mais específicos e rastreáveis na ciência.

O Software Livre tem sido o principal componente da pesquisa em Ciência da Computação nas últimas décadas. Há centenas de exemplos de sistemas, de bibliotecas e de ferramentas que têm promovido o rápido desenvolvimento da pesquisa em Computação. Exemplos conhecidos incluem o sistema operacional Gnu/Linux, a ferramenta de composição tipográfica LaTeX e bibliotecas e linguagens para computação estatística, como R. Por fim, atualmente, há uma grande coleção de bibliotecas livres como

Scikit-learn, TensorFlow e PyTorch que contribuem para o rápido desenvolvimento dos campos de Inteligência Artificial e Aprendizagem de Máquina.

Outras Ciências também se beneficiam com o software livre. A genômica, por exemplo, desenvolveu o software de código aberto EMBOSS, um pacote de análise para atender as necessidades de biólogos moleculares. Já ROOT é uma infraestrutura de suporte à análise de dados usada em pesquisas de Física Experimental de Altas Energias. É um software livre que nasceu no CERN e tem recebido contribuições de desenvolvedores de todo o mundo. Atualmente, mais de 1 Exabyte de dados científicos são armazenados em arquivos ROOT. O bóson de Higgs foi encontrado com o suporte do software livre ROOT.

O Software Livre é fundamental para a Ciência Aberta. Esta mensagem deve ser transmitida a cientistas, agências de fomento à pesquisa, organizações científicas e governamentais.



CHRISTINA VON FLACH é Professora Associada do Instituto de Computação da Universidade Federal da Bahia. Atua na área de Engenharia de Software tendo publicado mais de 100 artigos científicos em congressos e periódicos. Recentemente tem trabalhado com temas relacionados à Educação em Engenharia de Software, Sustentabilidade e Aspectos Sócio-técnicos de Ecossistemas de Software, e Ciência Aberta.

FABIO KON é Professor Titular de Ciência da Computação na Universidade de São Paulo. Atua na área de Sistemas Distribuídos, Engenharia de Software e Ciência de Dados tendo publicado cerca de 200 artigos científicos em congressos e periódicos. Recentemente tem trabalhado na área de Cidades Inteligentes e gostaria de ter mais tempo para tocar seu vibrafone.