CAPÍTULO 10

FERRAMENTAS DE FLUXO DE TRABALHO ABERTAS

MIRELE CAROLINA SOUZA
DIEGO JOSÉ MACEDO
MILTON SHINTAKU

10.1 INTRODUÇÃO

Ao longo da história humana, buscou-se certa padronização das atividades, de forma que até normas internacionais foram criadas para promover essa orientação, partindo da premissa de que, se um processo está padronizado, tem maior chance de evoluir para alcançar melhor qualidade. Essa busca pela padronização e, por consequência, pela qualidade, está muito presente na indústria e no comércio, mas pode ser aplicada a todas as áreas humanas.

Ao observar uma pessoa preparando uma refeição seguindo uma receita, nota-se a busca pela padronização, com o objetivo de que o resultado final seja o esperado — ou seja, um conjunto bem definido de ações para a execução de uma tarefa. Essa atividade, ainda que corriqueira, possibilita a reprodução de uma experiência desenvolvida por um especialista, permitindo que até mesmo um leigo, sem formação acadêmica, consiga realizar tal tarefa.

Em uma pesquisa científica, por sua vez, mesmo com toda a diversidade nas diferentes áreas do conhecimento, seguem-se padrões aceitos pela metodologia científica. A pesquisa científica é uma atividade humana que segue padrões aceitos e defendidos pela comunidade científica. Nesse sentido, pode-se considerá-la como trabalho? Historicamente, os cientistas não eram vistos como trabalhadores, mas, atualmente, pode-se considerá-los como profissionais que ganham seu sustento por meio da pesquisa?

Ribeiro e Leda (2004) discutem sobre o conceito de trabalho, relatando que o termo possui uma relação etimológica com um instrumento utilizado para imobilizar animais e, posteriormente, para tortura, sendo associado com um fardo ou sacrifício. Para os autores, o termo “trabalho” foi sendo ressignificado conforme a história e as vertentes, especialmente no campo da administração, sendo que grande parte da população dedica

uma quantidade significativa de seu tempo a uma atividade remunerada, considerada como trabalho.

Essa visão de trabalho como forma de subsistência tem amparo na definição de dicionário, no qual o trabalho está relacionado a ofício ou profissão. Com isso, pode-se compreender que o pesquisador é um trabalhador da ciência, um profissional que executa o ofício da pesquisa. No entanto, trabalho também pode ser entendido como um conjunto de atividades realizadas por um grupo de pessoas para alcançar um objetivo ou propósito específico. Nesse sentido, a pesquisa é um trabalho direcionado à obtenção de novos conhecimentos e à melhoria da vida humana, conforme a definição de Bjork (2005).

Nesse contexto, com a aproximação da pesquisa ao trabalho, podem-se criar sinergias para que o movimento da Ciência Aberta apoie outras atividades humanas, como as atividades produtivas e governamentais. Da mesma forma, é possível aplicar muitos dos preceitos da administração e outras disciplinas às atividades de pesquisa. Isso revela a reciprocidade entre os princípios que orientam a pesquisa e o trabalho, confirmando, cada vez mais, o papel da ciência como apoio à melhoria da sociedade.

10.2 FLUXO DE TRABALHO

Com base no entendimento de trabalho, como esse conjunto de atividades, incluindo a pesquisa, pode-se procurar compreender o fluxo de trabalho no âmbito da Ciência Aberta. Na administração, ou disciplinas correlatas, fluxo de trabalho também é conhecido como *workflow*, em uma tradução literal, visto que essa disciplina utiliza muitos termos em língua inglesa, pois grande parte das teorias são provenientes de países anglófonos ou mesmo oriundo de literatura nesse idioma.

Literalmente, fluxo de trabalho significa, como conceitua Marinescu (2002), a execução coordenada de um conjunto de atividades ou tarefas, no qual pode estar presente em uma variedade de trabalhos, como pesquisas científicas, gerenciamento de campo de batalha, suporte logístico para a fusão de duas empresas e gerenciamento de assistência médica, entre tantos outros. Zhao e Stohr (2000), por sua vez, defendem que o fluxo de trabalho requer coordenação e estudos para estruturar as atividades, identificando os processos, especificando as tarefas, assim como os seus executores. Dessa forma, é possível promover a automação com o uso de tecnologia.

Entretanto, Joostene e Brinkkemper (1995) relatam a dificuldade de descobrir, na prática, o que é fluxo de trabalho, especialmente em um cenário em que os sistemas de informação se tornam cada vez mais presentes nas organizações. Vale destacar que os sistemas de gerenciamento de fluxo de trabalho, desenvolvidos de forma padronizada e flexível, alteram como uma organização opera, quase como se as organizações se ajustassem aos sistemas em busca de eficiência.

Conforme o glossário do *Workflow Management Coalition* (WMC), organização que agrega estudiosos de gestão de fluxo de trabalho, este é definido como:

> [...] a sequência de tarefas, etapas e decisões que precisam ser seguidas para concluir um processo específico. Pode ser pensado como um conjunto de instruções que descrevem como um processo deve ser executado, incluindo a ordem em que as tarefas devem ser concluídas, quem é responsável por executar cada tarefa e o que deve acontecer em seguida com base no resultado de cada tarefa[1] (Workflow Management Coalition, 2024, p. 1).

[1] *Workflow Management Coalition* (WMC). Disponível em: https://wfmc.org/glossary/#W. Acesso em: 22 out. 2024.

Seguir um fluxo de trabalho garante que as atividades sejam executadas na ordem correta e pelas pessoas certas, de forma a obter os melhores resultados. Por isso, parte dos estudos sobre esses fluxos se voltam às ferramentas informatizadas de gestão. Da mesma forma, há estudos sobre sistemas de gestão de fluxo de trabalho, com abrangência maior, envolvendo não apenas as ferramentas, mas todos os elementos que compõem o processo.

No âmbito das ciências, o fluxo de trabalho tem sido conhecido há séculos com o estabelecimento das suas metodologias científicas. Entretanto, o movimento da Ciência Aberta tem alterado processos secularmente solidificados, como no caso da abertura dos dados de pesquisa ou da avaliação pelos pares. Com isso, pode-se falar de fluxo de trabalho aberto, de forma a transformar os processos mais colaborativos e transparentes.

10.3 FLUXO DE TRABALHO ABERTO

No âmbito da Ciência Aberta, um fluxo de trabalho aberto refere-se ao compartilhamento de todas as etapas de um processo ou à pesquisa científica de forma transparente e acessível. O uso de padrões abertos pode assegurar que esses fluxos sejam executados em diferentes plataformas e ferramentas. No contexto científico, o fluxo aberto significa que as etapas envolvidas na pesquisa — incluindo dados, códigos e metodologias — estejam transparentes, acessíveis e compartilháveis. Quando um fluxo de trabalho é aberto, outros pesquisadores têm a oportunidade de replicar as etapas realizadas. Essa abordagem facilita a colaboração e a melhoria contínua das metodologias pela comunidade científica. Nesse contexto, os autores Fecher e Friesike (2014, p. 35) conceituam que:

> [...] o movimento da Ciência Aberta promove maior transparência, acessibilidade e colaboração nos processos de pesquisa, garantindo que dados e métodos sejam compartilhados abertamente e facilmente reproduzíveis.

Os autores Vicente-Sáez e Martínez-Fuentes (2018) caracterizam "diferenciais" que delimitam a Ciência Aberta: transparência, acessibilidade, compartilhamento e colaboração. Essas características estão diretamente alinhadas com o conceito de fluxo de trabalho aberto, que visa garantir que todas as etapas do processo de pesquisa sejam transparentes, ou seja, visíveis e compreensíveis para que outros pesquisadores possam verificar, entender e reproduzir as etapas.

Os componentes do fluxo de trabalho, como dados e ferramentas, devem ser disponibilizados de forma aberta e acessível a todos, sem restrições de propriedade ou custos. Além disso, o compartilhamento de dados, códigos e metodologias de forma pública permite que outros pesquisadores utilizem esses recursos para gerar novos conhecimentos ou replicar as etapas de uma pesquisa. O fluxo de trabalho aberto, portanto, facilita a colaboração entre pesquisadores.

Em muitos projetos de pesquisa, as ferramentas computacionais são essenciais e, por isso, a abordagem de código aberto é considerada crucial para a implementação de fluxos de trabalho abertos. Ferramentas de código aberto possuem o código-fonte disponibilizado gratuitamente para consulta, análise, modificação e redistribuição.

10.4 FERRAMENTAS DE FLUXO DE TRABALHO ABERTAS

Nesse caminho, os fluxos de trabalho científicos têm adotado cada vez mais a tecnologia para a sua gestão, na medida em que suas atividades são padronizadas e bem definidas, com a evolução das metodologias adotadas pela comunidade científica. McPhillips et *al.* (2009) defendem que o uso de tecnologias para gestão do fluxo de trabalho tem aumentado vertiginosamente, visando melhorar a sua produtividade e possibilitando que cientistas com pouca experiência em tecnologia possam utilizar ferramentas para otimizar as suas pesquisas.

Nas últimas décadas, a necessidade de automatizar e gerenciar tarefas complexas de pesquisa tornou-se importante. Ferramentas simples, como shell scripts, eram frequentemente usadas para automatizar tarefas, mas logo percebeu-se que uma abordagem mais sistemática era necessária. Durante os anos 90 e início dos anos 2000, foram desenvolvidos diversos sistemas de gestão de fluxo de trabalho, principalmente para atender a áreas específicas, como bioinformática e física computacional. Nesse período, surgiram ferramentas de fluxo de trabalho de código aberto, como *Kepler* (Altintas et *al.*, 2004), *Taverna* (Oinn et *al.*, 2004), e *Pegasus* (Deelman et *al.*, 2005), que facilitaram a criação de pipelines científicos.

Conforme conceitua Talia (2013), alguns sistemas de fluxo de trabalho científico desenvolveram seus próprios modelos de fluxo de trabalho, resultando em uma diversidade de formalismos, como *Unified Modeling Language* (UML) e linguagens baseadas em *Extensible Markup Language* (XML). Essa variedade surgiu porque cada sistema foi criado para atender a necessidades específicas de diferentes áreas científicas. Além disso, muitos desses sistemas foram desenvolvidos antes da popularização do

Business Process Execution Language (BPEL) que, embora útil para processos de negócios, não atendia adequadamente às demandas dos fluxos de trabalho científicos, levando os desenvolvedores a adotar soluções personalizadas. Como consequência, os cientistas enfrentaram desafios, pois dificultou a interoperabilidade e o compartilhamento eficiente de fluxos de trabalho entre diferentes ferramentas.

Diante da crescente necessidade de padronização e compartilhamento em pesquisas científicas, foi desenvolvida uma linguagem de descrição de fluxo de trabalho, o *Common Workflow Language* (CWL) é um padrão aberto criado para descrever fluxos de trabalho e ferramentas científicas de maneira que possam ser interoperáveis e executados em diferentes plataformas. Amstutz *et al.* (2016) conceituam o CWL como uma ferramenta desenvolvida para promover a transparência, permitindo que pesquisadores descrevam de forma precisa as etapas de seus experimentos e análises, garantindo que essas descrições possam ser reutilizadas e adaptadas por outros.

Entre as ferramentas populares que oferecem suporte ao CWL para fluxos de trabalho aberto, destacam-se: *Toil* (Vivian et *al.*, 2017), *CWL-Airflow* (Kotliar; Kartashov; Barski, 2019) e *Arvados* (2024), entre outras. Além disso, ferramentas como *Pegasus* e *Galaxy* (2024) têm buscado incluir o suporte. Essas ferramentas permitem a execução, o compartilhamento e a colaboração de *workflows* complexos em diversas áreas científicas, facilitando a automação de processos em ambientes de pesquisa e computação distribuída.

Portanto, ferramentas baseadas em padrões abertos podem ser integradas a repositórios públicos e permitem aos pesquisadores criar, executar e compartilhar fluxos de trabalho de maneira padronizada. Os padrões abertos simplificam o acesso, a modificação e a reutilização dos fluxos pela comunidade científica, além de possibilitar melhorias contínuas e a implementação de novas funcionalidades. Em uma pesquisa científica, as ferramentas de fluxo de trabalho podem desempenhar um

papel fundamental, pois oferecem a capacidade de descrever de forma detalhada todas as etapas envolvidas no fluxo. Dependendo do contexto, conforme Deelman *et al.* (2015) o fluxo de trabalho pode ser desde uma simples operação de acesso a dados ou até uma sequência complexa de etapas interdependentes, como análises e transformações.

10.5 CONSIDERAÇÕES FINAIS

As ferramentas de fluxo de trabalho aberto desempenham um papel crucial na modernização da pesquisa científica, facilitando a gestão e automação de processos complexos. O uso dessas ferramentas no contexto científico tem evoluído significativamente com o surgimento de tecnologias padronizadas e abertas, como o CWL, que facilitam a reprodutibilidade e o compartilhamento de fluxos de trabalho.

A integração dessas ferramentas com repositórios públicos e a adoção de padrões abertos são essenciais para melhorar a colaboração e a transparência nas pesquisas científicas. Ferramentas como, *Toil*, *CWL--Airflow* e *Arvados* entre outras, demonstram como diferentes áreas de pesquisa podem explorar fluxos de trabalho bem definidos. Assim, a padronização e o uso de ferramentas abertas impulsionam a inovação e a eficiência na pesquisa, atendendo à crescente necessidade de interoperabilidade e compartilhamento de conhecimento.

REFERÊNCIAS

ALTINTAS, Ilkay; BERKLEY, Chad; JAEGER, Efrat; JONES, Matthew; LUDÄSCHER, Bertram; MOCK, Stephan. Kepler: An extensible system for design and execution of scientific workflows. *In*: INTERNATIONAL CONFERENCE ON SCIENTIFIC AND STATISTICAL DATABASE MANAGEMENT, 16., New York, 2004. **Proceedings** [...]. New York: IEEE Computer Society, 2004. p. 423. DOI: 10.1109/SSDM.2004.1311241. Disponível em: https://www.computer.org/csdl/proceedings-article/ssdbm/2004/21460423/12OmNyrqzvr. Acesso em: 22 out. 2024.

ARVADOS PROJECT. **Arvados unified data and workflow management.** [*S. l.*]: Arvados project, 2024. Disponível em: https://arvados.org. Acesso em: 27 Set. 2024.

BJÖRK, Bo-Christer. A lifecycle model of the scientific communication process. **Learned Publishing,** [*S. l.*], v. 18, n. 3, p. 165-176, 2005. DOI:10.1087/0953151054636129.

DEELMAN, Ewa; SINGH, Gurmeet; SU,Mei Hui; BLYTHE. James; GIL, Yolanda; KESSELMAN, Carl; MEHTA, Gaurang; VAHI, Karan; BERRIMAN, Bruce; GOOD John; LAITY, Anastasia; JACOB, Joseph C.; KATZ, Daniel S. Pegasus: a framework for mapping complex scientific workflows onto distributed systems. **Scientific Programming,** [*S. l.*], v. 13, n. 3, p. 219–237, 2005. DOI: 10.1155/2005/128026. Disponível em: https://experts.illinois.edu/en/publications/pegasus-a-framework-for-mapping-complex-scientific-workflows-onto. Acesso em: 22 out. 2024.

DEELMAN, Ewa; VAHI, Karan; JUVE, Gideon; RYNGE, Mats; CALLAGHAN, Scott; MAECHLING, Philip J.; MAYANI, Rajiv; CHEN, Weiwei; SILVA, Rafael Ferreira da; LIVNY, Miron; WENGER, Kent. Pegasus, a workflow management system for science automation. **Future Generation Computer Systems,** [*S. l.*], v. 46, p. 17-35, 2015. DOI: https://doi.org/10.1016/j.future.2014.10.008. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0167739X14002015. Acesso em: 22 out. 2024.

FECHER, Benedikt; FRIESIKE, Sascha. Open science: one term, five schools of thought. *In*: BARTLING, Sonke; FRIESIKE, Sascha (ed.). **Opening Science.** Springer International Publishing, 2014. DOI: https://doi.org/10.1007/978-3-319-00026-8. p. 17-47.

JOOSTEN, Stef; BRINKKEMPER, Sjaak. Fundamental concepts for workflow automation in practice. *In*: INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS (ICIS'95), 12., Amsterdam, 1995. **Proceedings** [...]. Amsterdam: ICIS, 1995. p. 1-13.

KOTLIAR, Michael; KARTASHOV, Andrey V.; BARSKI, Artem. CWL--Airflow: a lightweight pipeline manager supporting Common Workflow Language. **GigaScience,** [*S. l.*], v. 8, n. 7, p. 1-8, 2019. DOI: https://doi.org/10.1093/gigascience/giz084. Disponível em: https://academic.oup.com/gigascience/article/8/7/giz084/5535758. Acesso em: 22 out. 2024.

MARINESCU, Dan C. **Internet-based workflow management:** Toward a semantic web. New York: Wiley-Interscience, 2002.

MCPHILLIPS, Timothy; BOWERS, Shawn; ZINN, Daniel; LUDASCHER, Bertram. Scientific workflow design for mere mortals. **Future Generation Computer Systems,** [*S. l.*], v. 25, n. 5, p. 541-551, 2009. DOI: https://doi.org/10.1016/j.future.2008.06.013. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0167739X08000873. Acesso em: 22 out. 2024.

OINN, Tom; ADDIS, Matthew; FERRIS, Justin; MARVIN, Darren; SENGER, Martin GREENWOOD, Mark; CARVER, Tim; GLOVER, Kevin; POCOCK, Matthew R.; WIPAT, Anil; LI, Peter. Taverna: A tool for the composition and enactment of bioinformatics workflows. **Bioinformatics,** [*S. l.*], v. 20, n. 17, p. 3045–3054, 2004. DOI: https://doi.org/10.1093/bioinformatics/bth361. Disponível em: https://academic.oup.com/bioinformatics/article/20/17/3045/186405. Acesso em: 22 out. 2024.

RIBEIRO, Carla Vaz dos Santos; LÉDA, Denise Bessa. O significado do trabalho em tempos de reestruturação produtiva. **Estudos e Pesquisas em Psicologia,** Rio de Janeiro, v. 4, n. 2, p. 76-83, 2004. Disponível em: https://pepsic.bvsalud.org/scielo.php?script=sci_abstract&pid=S1808-42812004000300006. Acesso em: 21 out. 2024.

TALIA, Domenico. Workflow systems for science: Concepts and tools. **International Scholarly Research Notices,** [*S. l.*], v. 2013, p. 1-15, 2013. DOI: https://doi.org/10.1155/2013/404525. Disponível em: https://onlinelibrary.wiley.com/doi/epdf/10.1155/2013/404525. Acesso em: 22 out. 2024.

THE GALAXY COMMUNITY. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. **Nucleic Acids Research,** [*S. l.*], v. 52, n. W1, p. w83-w94, 2024. DOI: https://doi.org/10.1093/nar/gkae410. Disponível em: https://academic.oup.com/nar/article/52/W1/W83/7676834. Acesso em: 22 out. 2024.

VICENTE-SAEZ, Ruben; MARTINEZ-FUENTES, Clara. Open Science now: A systematic literature review for an integrated definition. **Journal of business research,** [*S. l.*], v. 88, n. c, p. 428-436, 2018. Disponível em: https://ideas.repec.org/a/eee/jbrese/v88y2018icp428-436.html. Acesso em: 22 out. 2024.

VIVIAN, John; RAO, Arjun Arkal; NOTHAFT, Frank Austin *et al.* Toil enables reproducible, open source, big biomedical data analyses. **Nature Biotechnology,** [*S. l.*], v. 35, n. 4, p. 314-316, 2017. DOI: 10.1038/nbt.3772. Disponível em: https://pubmed.ncbi.nlm.nih.gov/28398314/. Acesso em: 22 out. 2024.

ZHAO, J. Leon; STOHR, Edward A. Building workflow engines for commerce logic automation. *In*: AMERICAS CONFERENCE ON INFORMATION SYSTEMS, 140., [*S. l.*], 2000. **Proceedings** [...]. [*S. l.*]: AMCIS, 2000. p. 558-561, 2000. Disponível em: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1534&context=amcis2000. Acesso em: 21 out. 2024.

COMO CITAR ESTE CAPÍTULO:

SOUZA, Mirele Carolina Souza; MACEDO, Diego José; SHINTAKU, Milton. Ferramentas de fluxo de trabalho abertas. *In*: DRUCKER, Debora Pignatari; CIUFFO, Leandro; SAYÃO, Luis Fernando; SHINTAKU, Milton; VIDOTTI, Silvana Aparecida Borsetti Gregorio (org.) **Infraestruturas de suporte à Ciência Aberta.** Brasília, DF: Editora Ibict, 2025. p. 212-224. DOI: 10.22477/9786589167754.cap10.