

I CONGRESSO NACIONAL DE CIÊNCIA & TECNOLOGIA ÂNIMA 2025

IV SIMPOSIO DE PESQUISA ECOSISTEMA ÂNIMA



APLICAÇÃO DOS PRINCÍPIOS FAIR NA GESTÃO DE VERSÕES DE SOFTWARES CIENTÍFICOS: MELHORES PRÁTICAS DO PROJETO NF-CORE

Marianna Pereira Silva Ramalho¹; Maria do Céu Pontes Vieira¹; Ana Luiza Cruz Gomes¹; Gabriela Porto Gaspar¹; André Salustino Cabral¹; Fernando Sérgio de Macêdo Caldas Segundo¹; Marcel da Câmara Ribeiro Dantas² (Dr.) 1 Universidade Potiguar. 2 Programa de Pós-Graduação em Biotecnologia, Universidade Potiguar – UnP marcel.câmara@unp.br

Introdução

A ciência atual depende fortemente de dados digitais e softwares complexos, o que torna essenciais os princípios FAIR para garantir transparência e reproduzibilidade. Embora criados para dados, esses princípios têm sido aplicados também a softwares e workflows, cuja rastreabilidade é fundamental. Apesar de ferramentas como Nextflow e padrões do projeto nf-core facilitarem a criação e manutenção de pipelines, o controle de versões de softwares ainda é um desafio e pode comprometer a confiabilidade dos resultados. O nf-core avançou significativamente ao fornecer centenas de pipelines e módulos reutilizáveis, usados em pesquisas amplamente citadas. Com a versão 25.04 do Nextflow e a introdução de topics, torna-se possível gerenciar versões de ferramentas de modo mais simples e rastreável. Este trabalho apresenta como adaptar módulos e práticas do nf-core para integrar topics, fortalecendo a aplicação dos princípios FAIR na gestão de ambientes computacionais.

Objetivos

O objetivo geral desta pesquisa é discutir e aprimorar a transparência, a rastreabilidade e a reproduzibilidade de pipelines de análise de dados desenvolvidos em Nextflow, por meio da aplicação dos princípios FAIR na gestão de versões de softwares e dependências, utilizando o novo recurso *topics* do Nextflow.

Metodologia

O estudo baseou-se na análise e adaptação de módulos e participação em discussões das convenções do projeto nf-core, uma comunidade internacional que mantém pipelines de bioinformática com foco em padronização, qualidade e reproduzibilidade. O método adotado pode ser descrito em quatro etapas principais: 1) análise dos padrões existentes; 2) estudo da nova funcionalidade *topics*; 3) adaptação de alguns módulos do nf-core; e 4) avaliação de transparência e reproduzibilidade.

Resultados

Os testes revelaram que a adoção do recurso *topics* no contexto nf-core oferece benefícios significativos em termos de transparência, organização e controle de versões. A rastreabilidade das versões foi aprimorada; a modularidade e o reuso foram otimizados; e a governança técnica se tornou mais clara. Além disso, o uso de *topics* reforça todos os pilares FAIR, promovendo interoperabilidade e reusabilidade entre pipelines e módulos. O recurso *topics* permite que as tarefas, que podem ser muito numerosas e ocorrer ao mesmo tempo, possam reportar de forma simples as versões das ferramentas utilizadas naquela tarefa para uma estrutura de dados central, com grande potencial para transparência das análises e reproduzibilidade das mesmas. É importante ressaltar que essa adoção está em fase final de discussão no projeto nf-core, de modo que ainda não é oficialmente adotada pelo projeto, com projeções para ser o formato recomendado de rastreabilidade de versões para o final do mês de novembro. Dito isso, as discussões e desafios técnicos relativos à implantação dessa nova abordagem vão além do que foi discutido e testado pelos autores deste trabalho.

Resultados continuação

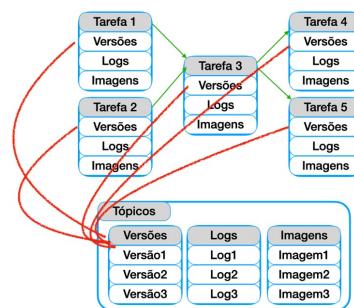


Figura 1: Diversas tarefas são executadas durante um pipeline de análise de dados. As setas verdes ilustram a dependência entre as tarefas, onde em muitos casos uma atua sobre o resultado da tarefa anterior. Na parte inferior, é possível ver uma estrutura chamada tópicos, com vários deles: Versões, Logs e Imagens. As curvas vermelhas ilustram o envio automático de informações de um determinado tipo, em cada tarefa, para a estrutura central de fácil acesso e controle a qual é chamada de tópicos. É importante ressaltar que o tópico de versões armazena não apenas a versão do software, mas também um identificador da tarefa, o nome do software, e como a informação da versão foi obtida.

Conclusões

A integração dos princípios FAIR com o controle de versões de softwares representa um passo essencial para consolidar práticas de ciência aberta e reproduzível. O uso dos *topics* do Nextflow na versão 25.04, aliado às boas práticas do projeto nf-core, constitui uma abordagem eficaz para aumentar a transparência, rastreabilidade e sustentabilidade técnica dos pipelines científicos. Como próximos passos, propõe-se expandir a adoção dos *topics* para pipelines complexos e mais módulos.

Bibliografia

- WILKINSON, M. D. et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, v. 3, n. 1, p. 1-9, 2016.
SEQERA LABS. Migration notes: Migrating to 25.04. Barcelona: Seqera Labs, 2025. Disponível em: <https://nextflow.io/docs/latest/migrations/25-04.html>. Acesso em: 15 nov. 2025.
EWELS, P. A. et al. The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*, v. 38, n. 3, p. 276-278, 2020.
LANGER, B. E. et al. Empowering bioinformatics communities with Nextflow and nf-core. *Genome Biology*, v. 26, n. 1, Art. 228, 2025.
CHUE HONG, N. P. et al. FAIR principles for research software (FAIR4RS principles). Zenodo, 2022. DOI: <https://doi.org/10.5497/RDA00068>. Disponível em: <https://zenodo.org/records/6623556>. Acesso em: 15 nov. 2025.

Agradecimentos

Este trabalho é vinculado ao Programa Pró-Ciência do Ecossistema Ânima e financiado pelo Instituto Ânima (EDITAL N° 14/2025).