

June, 2023

Nextflow best practices

Leveraging nf-core tooling and standards within your organization

Overview

01 Introduction

Phil - 20 mins

02 Live demo

Harshil - 20 mins

03 Conclusion

Phil - 5 mins

04 Q&A

15 mins



01

Introduction

Seqera, Nextflow, nf-core and Tower.



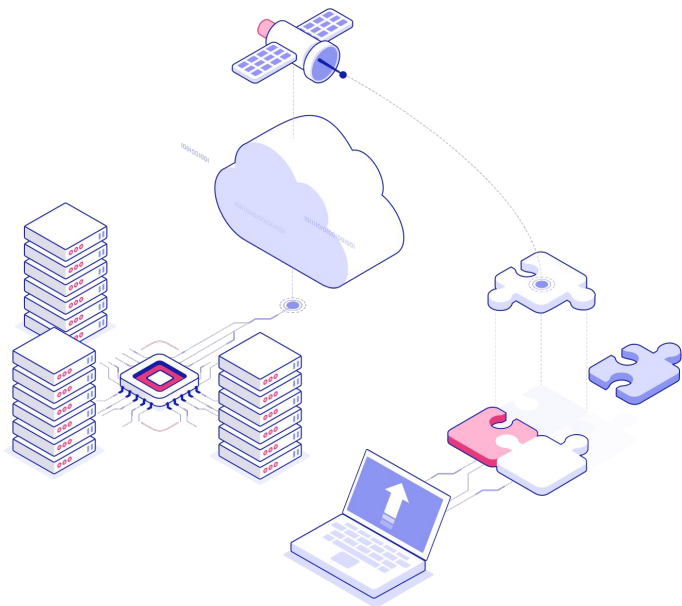
Seqera at a glance

Key Highlights

- **Founded in 2018** by Evan Floden and Paolo di Tommaso. HQ in Barcelona and Toronto
- Trusted by over **8,000 researchers** and over **150 leading pharma and biotech** companies
- **Raised over €27M** and recipient of several grants from the Chan Zuckerberg Foundation
- **Creators of Nextflow**, the global standard for workflow orchestration
- **Championing open science**, acting as a steward of over 70 open-source data pipelines available to its wider community of 10,000 researchers



The Seqera Stack



Nextflow

Scientific workflow management

Tower

Platform for collaborative scientific data analysis

MultiQC

Bioinformatics reporting

Fusion

File-system for cloud-native data pipelines

Wave

Container provisioning



Nextflow

Nextflow has a vast open source community actively using it on their day-to-day, supporting its development and contributing to new pipelines

120K+

monthly
downloads

16,000+

active developers
/month

40+

international
Workshops

150K+

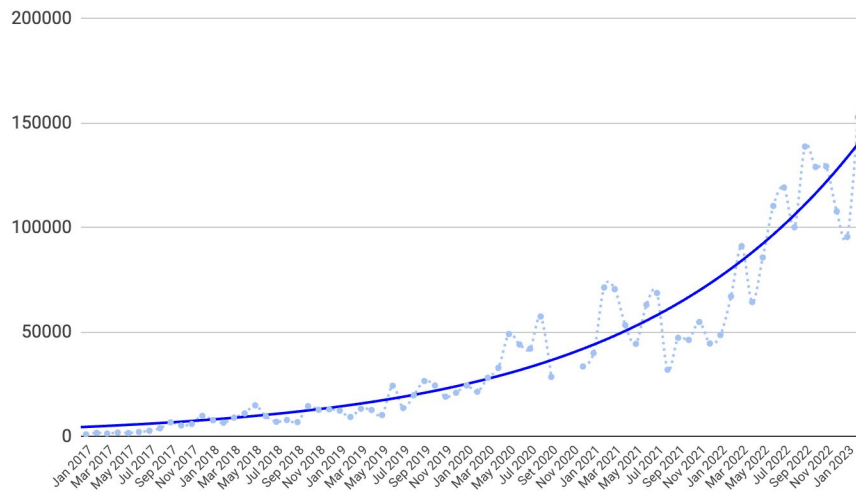
lines of
code

2K+

stars on
github

130+

contributors



Nextflow

Nextflow has a vast open source community actively using it on their day-to-day, supporting its development and contributing to new pipelines



Reproducible

Integration with code management tools, with versioned releases.



Portable

Docker, Singularity, Conda, works with most compute environments.



Scalable

5 samples on your laptop,
5k on an HPC or
5 million in the cloud.



Tower

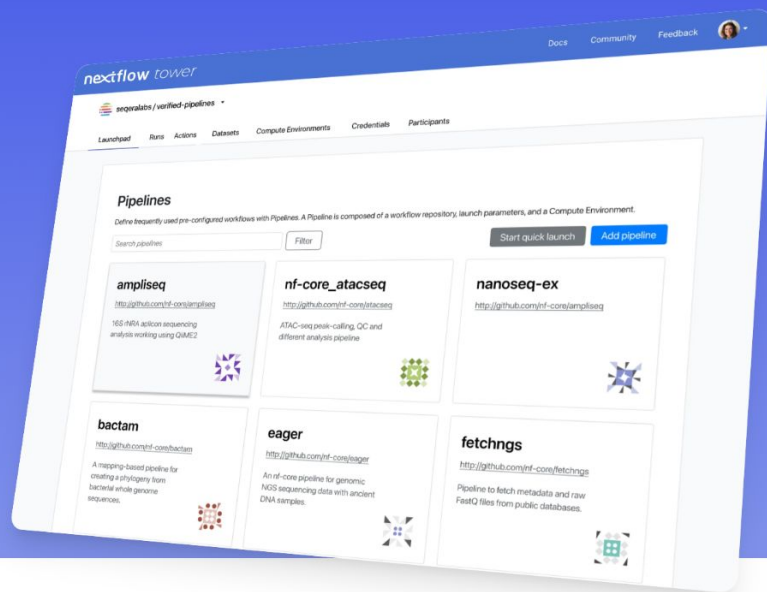
A central command post for managing data analysis pipelines at scale

nextflow tower

Maximize productivity by enabling collaborative data analysis at scale, on-premises or in any cloud.

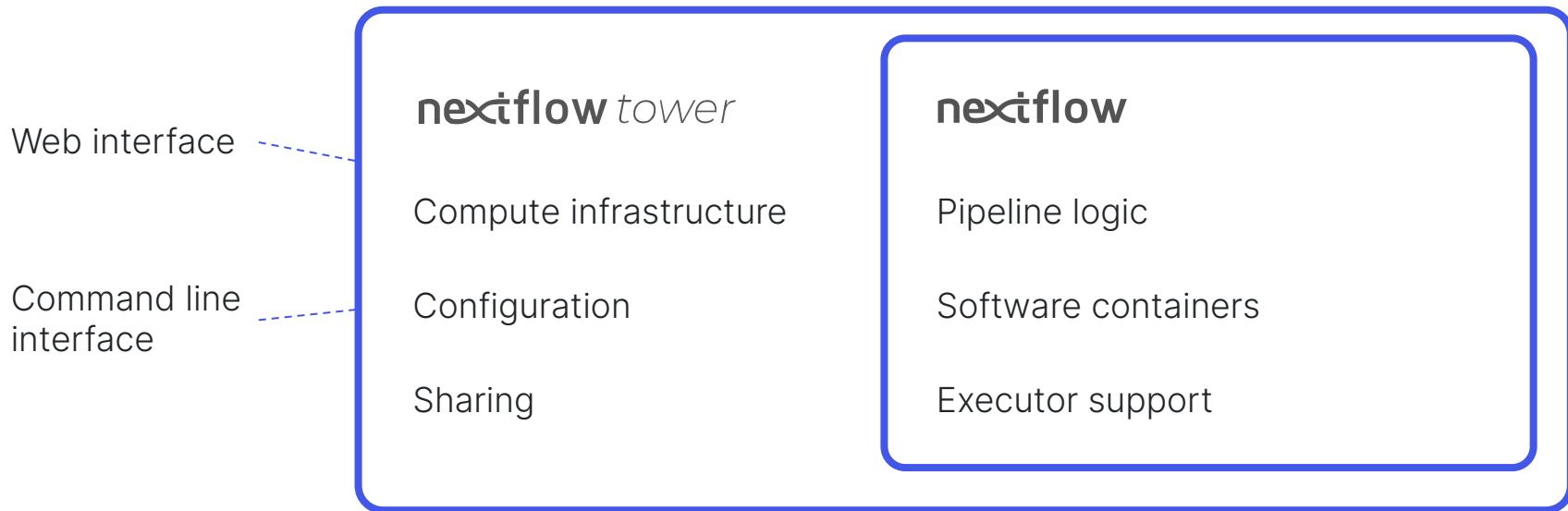
Join for free

Book a demo



Tower

A central command post for managing data analysis pipelines at scale



Tower

A central command post for managing data analysis pipelines at scale



Reduce cloud expenses

Instance selection, control & monitoring, automated resource optimization



Ensure reproducibility

Data provenance, auditability and compliance with versioned flows, containers & datasets



Automate manual tasks

Boost efficiency and avoid errors with Tower API and sequencer platform integrations



Flexible provisioning

Run locally, on HPC clusters, or auto-provision resources on your choice of cloud



Secure collaboration

Share pipelines, compute environments, and data with fine-grained access controls



Boost productivity

Access in-house or production-proven pipelines through an intuitive web UI



nf-core

A community effort to collect a curated set of analysis pipelines built using Nextflow.

6k+

Slack members

2k+

GitHub contributors

100+

GitHub repositories

46K+

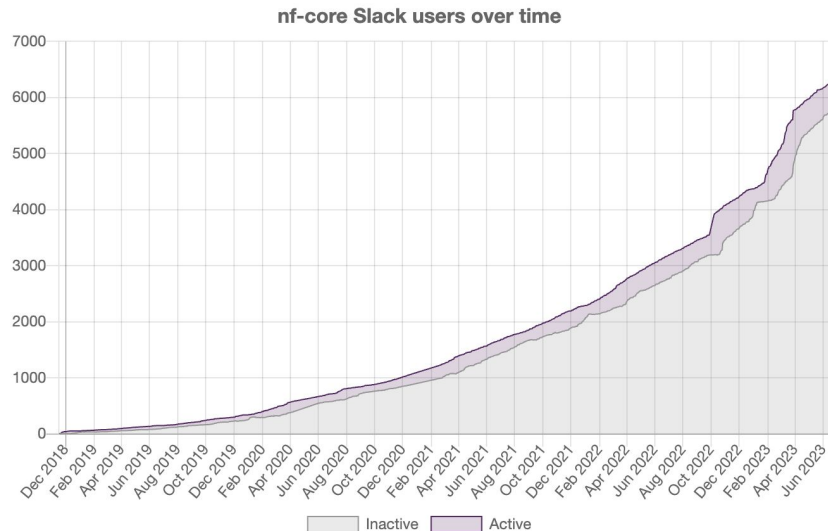
GitHub commits

15K+

Pull requests

6.7k+

GitHub issues



nf-core: Principles

A community effort to collect a curated set of analysis pipelines built using Nextflow.



Cooperation

Develop with the community



Standards

Use a common template



Collaboration

No duplicate pipelines within nf-core



Helper tools

Tools built for everyone



Compatibility

Tools work for any Nextflow pipeline



Components

Collaborate on component-level



nf-core

A community effort to collect a curated set of analysis pipelines built using Nextflow.

82 Pipelines

Ready to use, covering most techniques in NGS

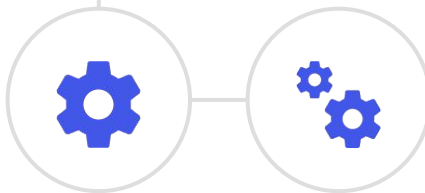


Helper tools

Running pipelines, writing pipelines, testing & automation

951 Modules

Tool process wrappers, with software and CI testing



44 Subworkflows

Shared subworkflows for common analysis pathways



nf-core: Anatomy of a pipeline

Building pipelines with units of reusability.

nf-core/ pipeline

Pipeline parameters

Resource requests

Tool arguments

Input channels

nf-core/ subworkflow

Channels

Pipeline logic

nf-core/ 
module

nf-core/ 
module

nf-core/ 
module

nf-core: Smörgåsbord

Pick and choose which smörgåsar you need

nf-core/  pipeline

fork

Your custom pipeline

nf-core/  subworkflow

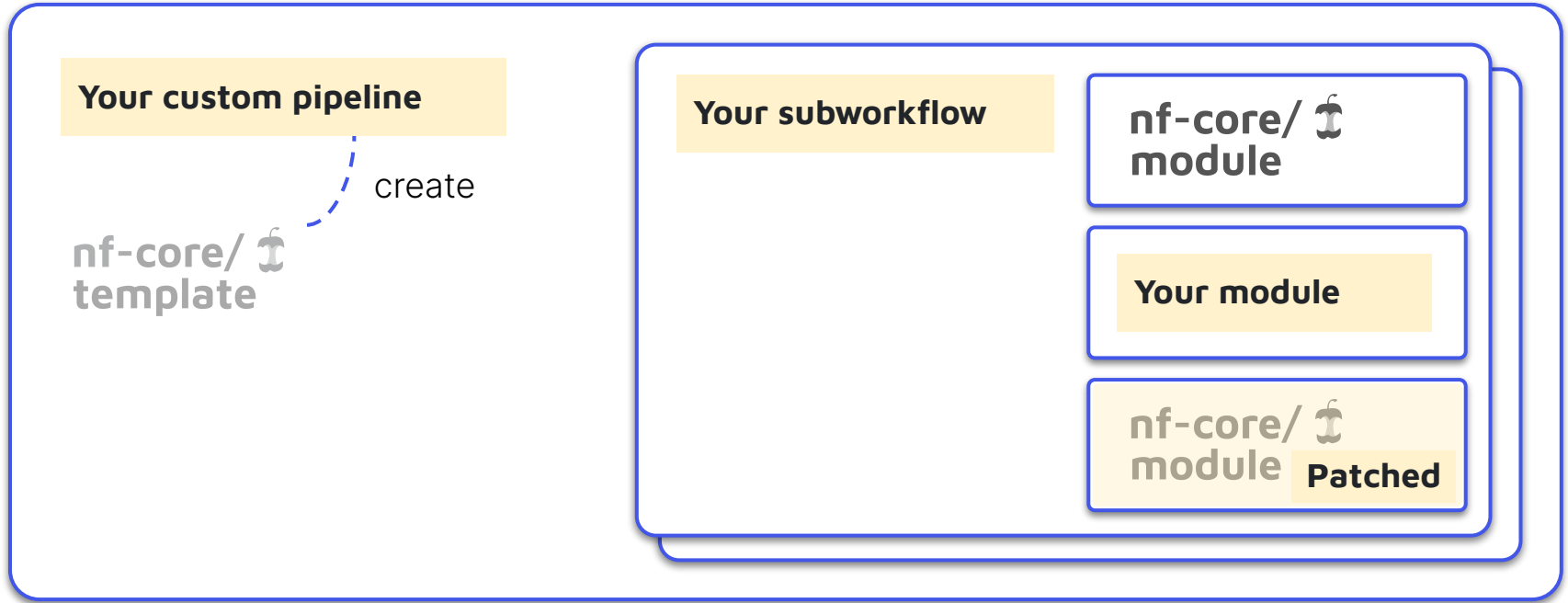
nf-core/  module

nf-core/  module

nf-core/  module

nf-core: Smörgåsbord

Pick and choose which smörgåsar you need



nf-core: Anatomy of a pipeline

Building pipelines with units of reusability.



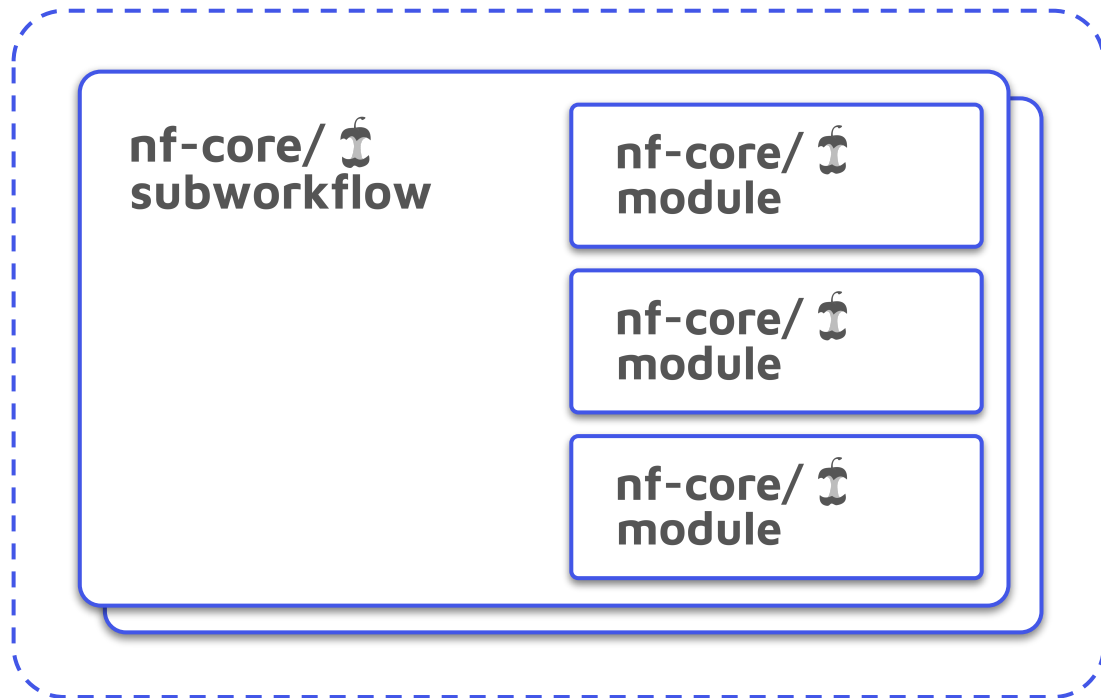
Local to pipeline



nf-core/modules



Your custom repository



nf-core/tools

Command line package to help you build your pipeline with ease

nf-core/ 
tools



Pipelines

Create from template,
sync to get updates.



Subworkflows

Create, install and
update.



Modules

Create, install, update,
patch, test.



Schema

Build your pipeline
schema with a GUI.



Linting

Test nf-core standards
and best practices.



Download

Fetch with singularity
images for offline use.



nf-validation

Nextflow plugin to natively handle schema files



Parameter validation

Functions to validate pipeline parameters and print CLI help text.



Sample sheet validation

Validate input sample sheets - formatting and contents.



Sample sheet channels

`fromSamplesheet()` channel factory, including meta maps.



[nextflow-io/nf-validation](https://github.com/nextflow-io/nf-validation)



<https://nextflow-io.github.io/nf-validation/>



nf-core: Smörgåsbord

Pick and choose which smörgåsar you need



Pipelines / template

82 pipelines and a base template



Linting

Choose conventions to test for



Subworkflows

44 subworkflows



Schema

Validation, channels and user interface



Modules

951 modules



Tooling

Development and deployment



nf-core: Examples

Battle tested examples out in the wild



Pipelines / template



Subworkflows



Modules



Schema



Linting



Tooling



**Boehringer
Ingelheim**

Custom modules repository

ATLASSIAN
Bitbucket



nf-core: Examples

Battle tested examples out in the wild



Pipelines / template



Subworkflows



Modules



Schema



Linting



Tooling



Wyoming Public Health
LABORATORY

Pushes modules back to central
nf-core repository



nf-core: Examples

Battle tested examples out in the wild



Pipelines / template



Subworkflows



Modules



Schema



Linting



Tooling



Leveraging Nextflow for the analysis of anything, anywhere, by anyone

Chris Wright
Senior Director, Customer Bioinformatics,
Oxford Nanopore Technologies

LC
LONDON CALLING
2023



nf-core: Examples

Battle tested examples out in the wild



Pipelines / template



Subworkflows



Modules



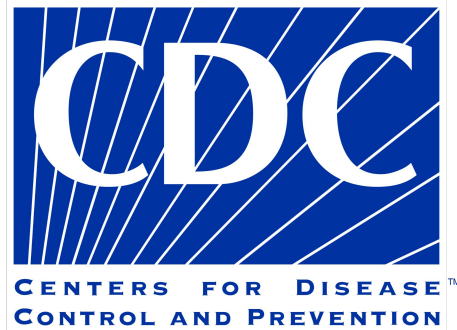
Schema



Linting



Tooling



nf-core: Examples

Who is using the template anyway?

~ 1000

Pipelines built with the nf-core template
in public GitHub repositories



02

Live demo

Build and run a pipeline in 20 minutes



03

Conclusion

Future directions



nf-core: Future directions

Improving the developer experience



Flexible template

Fine-grained control over what's included, improved support for GitLab, BitBucket.



New website

Rebuilt from the ground up, better navigation and content.



Sample sheet schema

Tooling to help create sample sheets and manage their schema.



Better testing

Moving to nf-test
Release testing with Azure
AWS runners for CI tests



Nextflow: Future directions

Coming soon to a runtime near you*



Clean intermediates

Delete intermediate files during pipeline execution



Better error messages

Tell you what went wrong and where (precisely)



Fusion for HPC

Lightning fast file I/O available on both Cloud *and* HPC



Spack / multi-arch

Nextflow + Spack, containers on the fly via Wave, multi-arch pipeline deployments.

*hopefully



Seqera Platform

Streamlining the scientific data analysis life cycle

Data Developer

Tools & workflows
IDEs - VSCode
CI/CD
Containers

Data Explorer

Instruments
Cloud storage
File systems
LIMS
Public resources

Nextflow

Scientific workflow management

MultiQC

Bioinformatics reporting

Fusion

File-system for cloud-native data pipelines

Wave

Container provisioning

Data Analytics

Reporting
Metrics
Provenance
Compliance

Data Studio

Interactive
Notebooks
RStudio
RShiny Applications
Genome Browsers

Data Pipelines

Batch Compute
Orchestration
Performant Storage



Take-home message

You can build *with* nf-core,
even if you're not building *for* nf-core

We can do great things together!



Hackathon and Summit 2023

Join us at the next nf-core hackathons!

 **BARCELONA**

October 16-20, 2023

 **BOSTON**

November 28-30, 2023

<https://summit.nextflow.io/>



Thanks: the nf-core community

Taking the phrase “team effort” to a whole new level



Schedule your demo today



Schedule a personalized demo

Experience how you can streamline your research workflows and accelerate scientific discover.

<https://seqera.io/demo/>



Resources

Links to find out more information

- **Websites**

- nf-core: <https://nf-co.re/>
- nf-validation: <https://nextflow-io.github.io/nf-validation/>
- nf-core on GitHub: <https://github.com/nf-core>
- Seqera: <https://seqera.io/>
- Nextflow: <https://nextflow.io/>
- Nextflow Tower: <https://tower.io/>

- **Content**

- Blog: [Best Practices for Deploying Pipelines with Nextflow Tower](#)
- Podcast: [Infrastructure team: nf-validation and nf-core components](#)
- Nextflow training: <https://training.nextflow.io/>
- nf-core documentation: <https://nf-co.re/docs>



Thank you



Phil Ewels, PhD

Lead Developer Advocate

phil.ewels@seqera.io

Barcelona | Stockholm



Harshil Patel, PhD

Head of Scientific Development

harshil.patel@seqera.io

Barcelona | London