

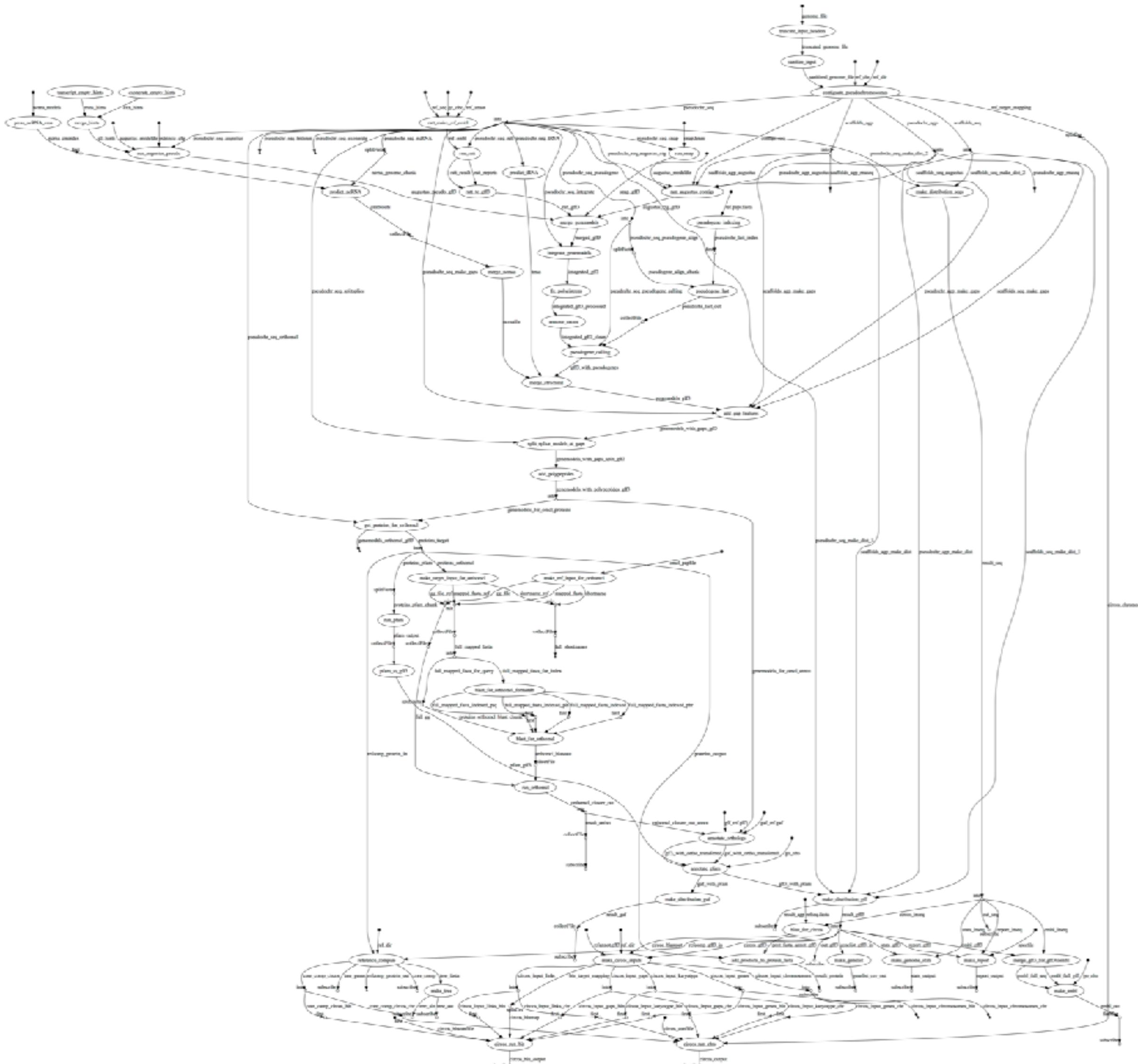


Chris Hakkaart

chris.hakkaart@seqera.io

Reproducible bioinformatics for everyone:

nextflow & nf-core



- 70 tasks
 - 55 custom scripts
 - 39 tools & libraries

Steinbiss et al., Companion parasite genome annotation pipeline DOI:10.1093/nar/gkw292

DATA ANALYSIS WORKFLOWS

- Data analysis applications perform computation to generate information from (large) datasets
- Embarrassingly parallelisation, can spawn 100s-100k jobs over a distributed cluster
- Mash-up of many different tools and scripts
- Complex dependency trees and configuration

Quantifying Reproducibility in Computational Biology: The Case of the Tuberculosis Drugome

Daniel Garijo¹, Sarah Kinnings², Li Xie³, Lei Xie⁴, Yinliang Zhang⁵, Philip E. Bourne^{3*}, Yolanda Gil^{6*}

1 Ontology Engineering Group, Facultad de Informática, Universidad Politécnica de Madrid, Madrid, Spain, **2** Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, California, United States of America, **3** Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California, United States of America, **4** Department of Computer Science, Hunter College, The City University of New York, New York, New York, United States of America, **5** School of Life Sciences, University of Science and Technology of China, Hefei, Anhui, China, **6** Information Sciences Institute and Department of Computer Science, University of Southern California, Los Angeles, California, United States of America

To reproduce the result of a typical
computational biology paper
requires 280 hours.

≈1.7 months!

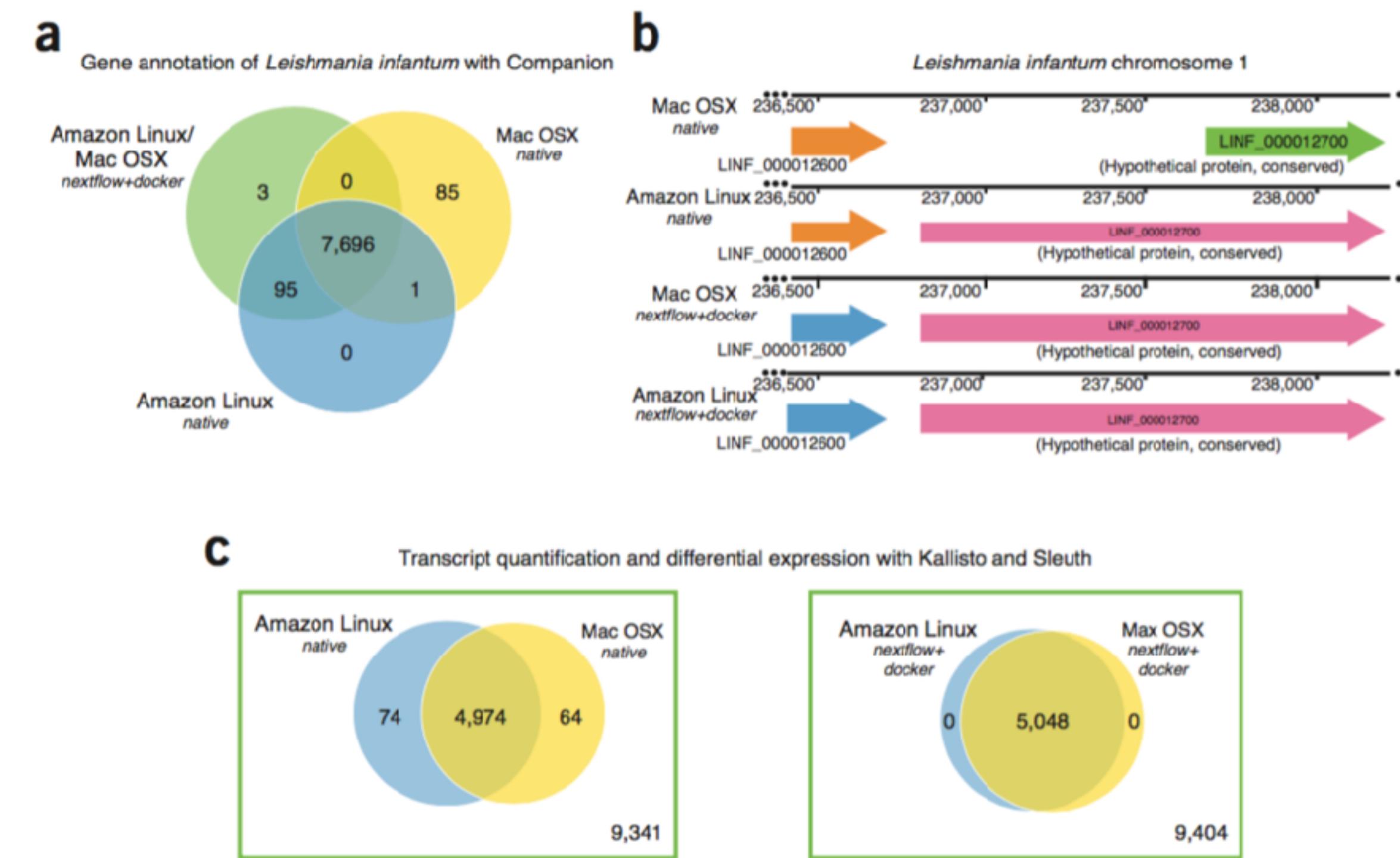
THE
SAME APPLICATION
DEPLOYED IN
DIFFERENT ENVIRONMENTS
PRODUCES
DIFFERENT RESULTS



Nextflow enables reproducible computational workflows

Paolo Di Tommaso, Maria Chatzou, Evan W Floden, Pablo Prieto Barja, Emilio Palumbo & Cedric
Notredame ✉

NATURE BIOTECHNOLOGY VOLUME 35 NUMBER 4 APRIL 2017



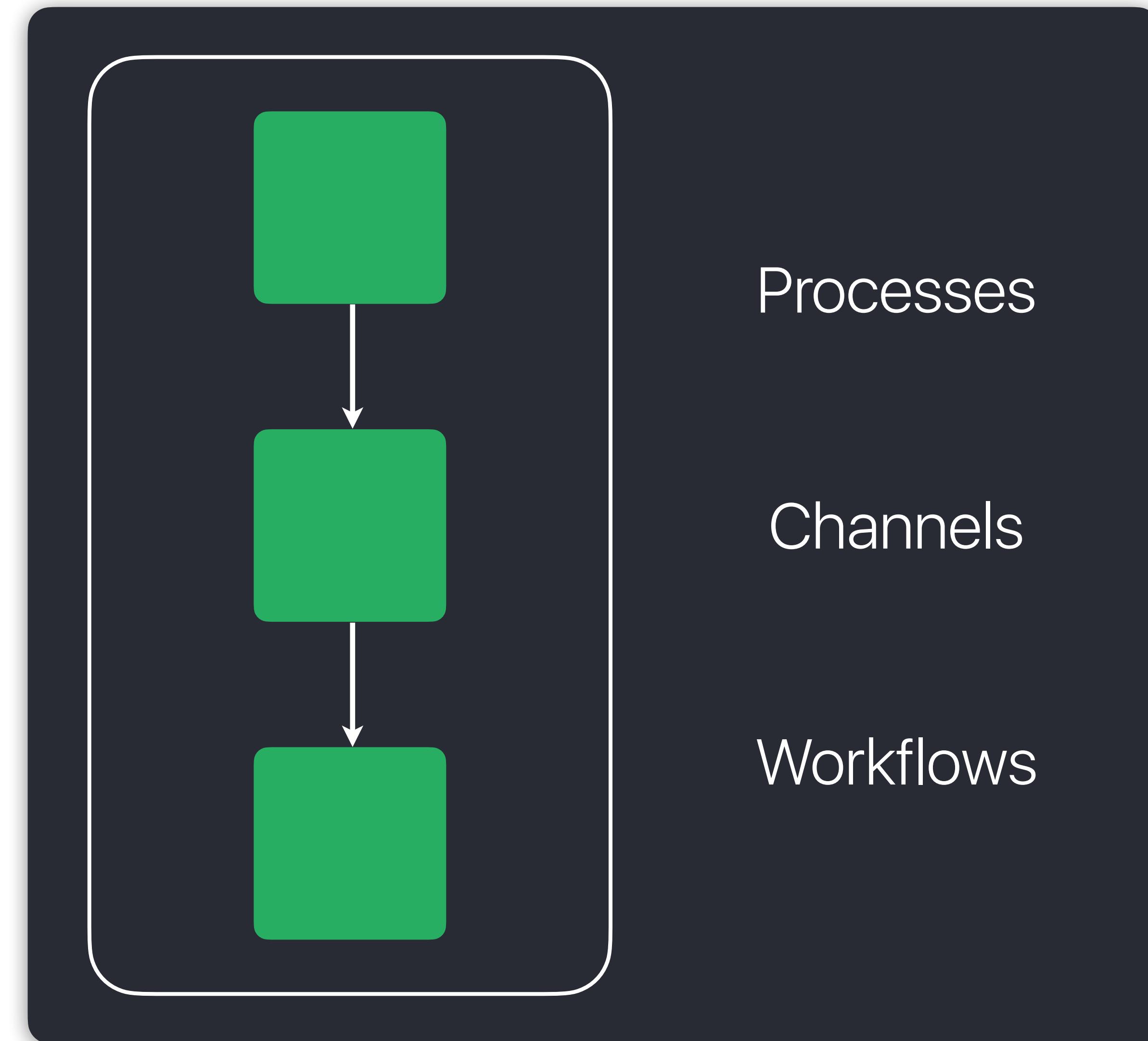
nextflow





Language

nextflow



nextflow

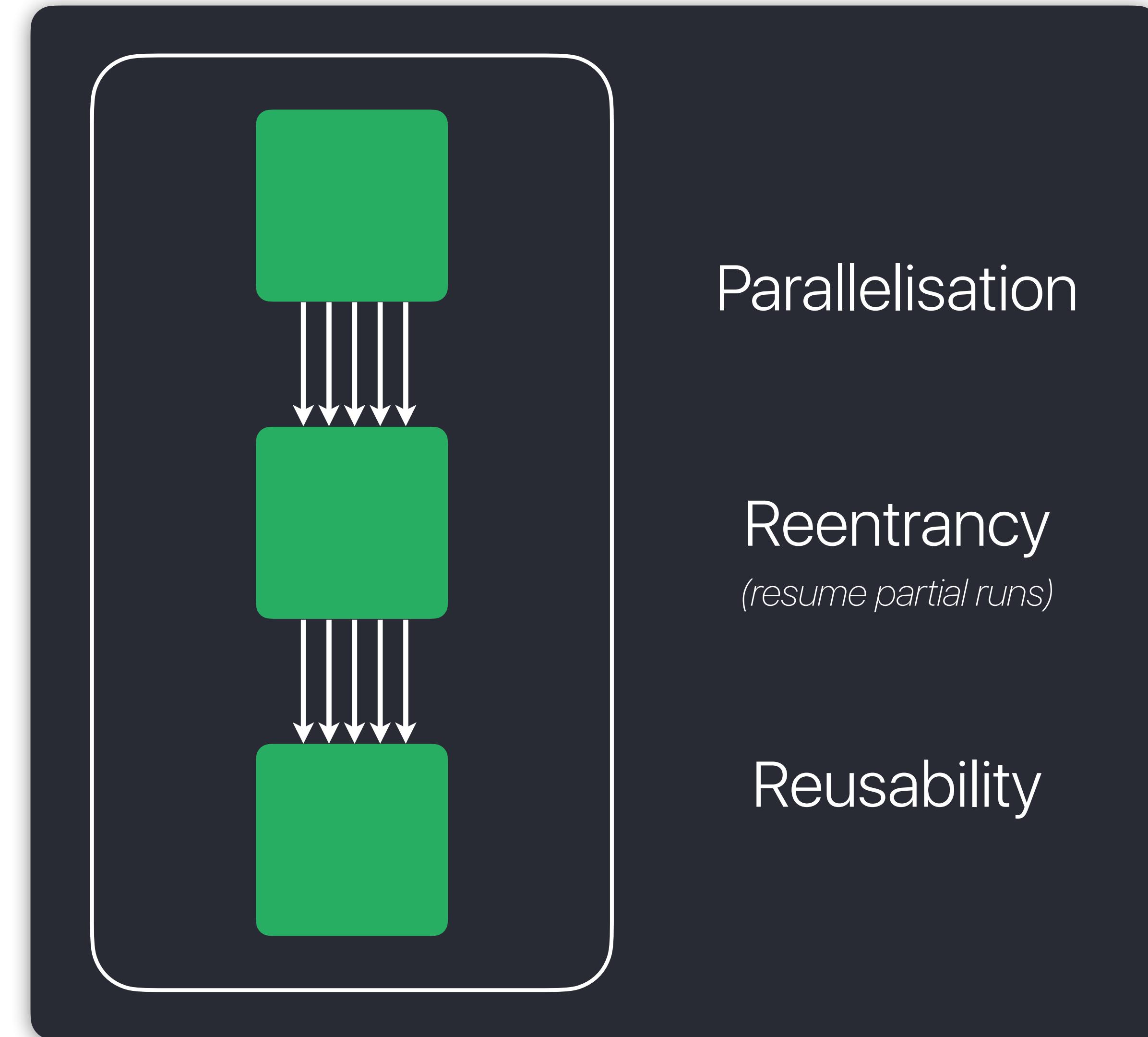
```
#!/usr/bin/env nextflow
process fastqc {
    input:
    path input

    output:
    path "*_fastqc.{zip,html}"

    script:
    """
    fastqc -q $input
    """
}

workflow {
    Channel.fromPath("*.fastq.gz") | fastqc
}
```

nextflow





Source Code
Management

Container
Technologies

Compute
Platforms



git



GitHub



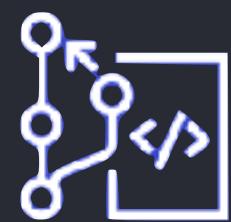
Bitbucket



GitLab



Gitea



AWS CodeCommit



Azure Repos

Container
Technologies

Compute
Platforms

nextflow



git



GitHub

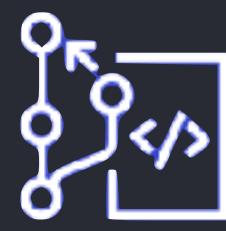
Bitbucket



GitLab



Gitea



AWS CodeCommit



Azure Repos



docker



Singularity



CONDA

Compute
Platforms

nextflow



git



GitHub



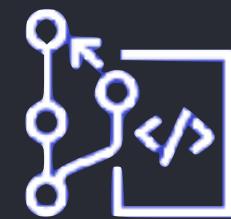
Bitbucket



GitLab



Gitea



AWS CodeCommit



Azure Repos



docker



Singularity



CONDA

SGE



Microsoft Azure



slurm
workload manager



aws

LSF

PBS



Google Cloud

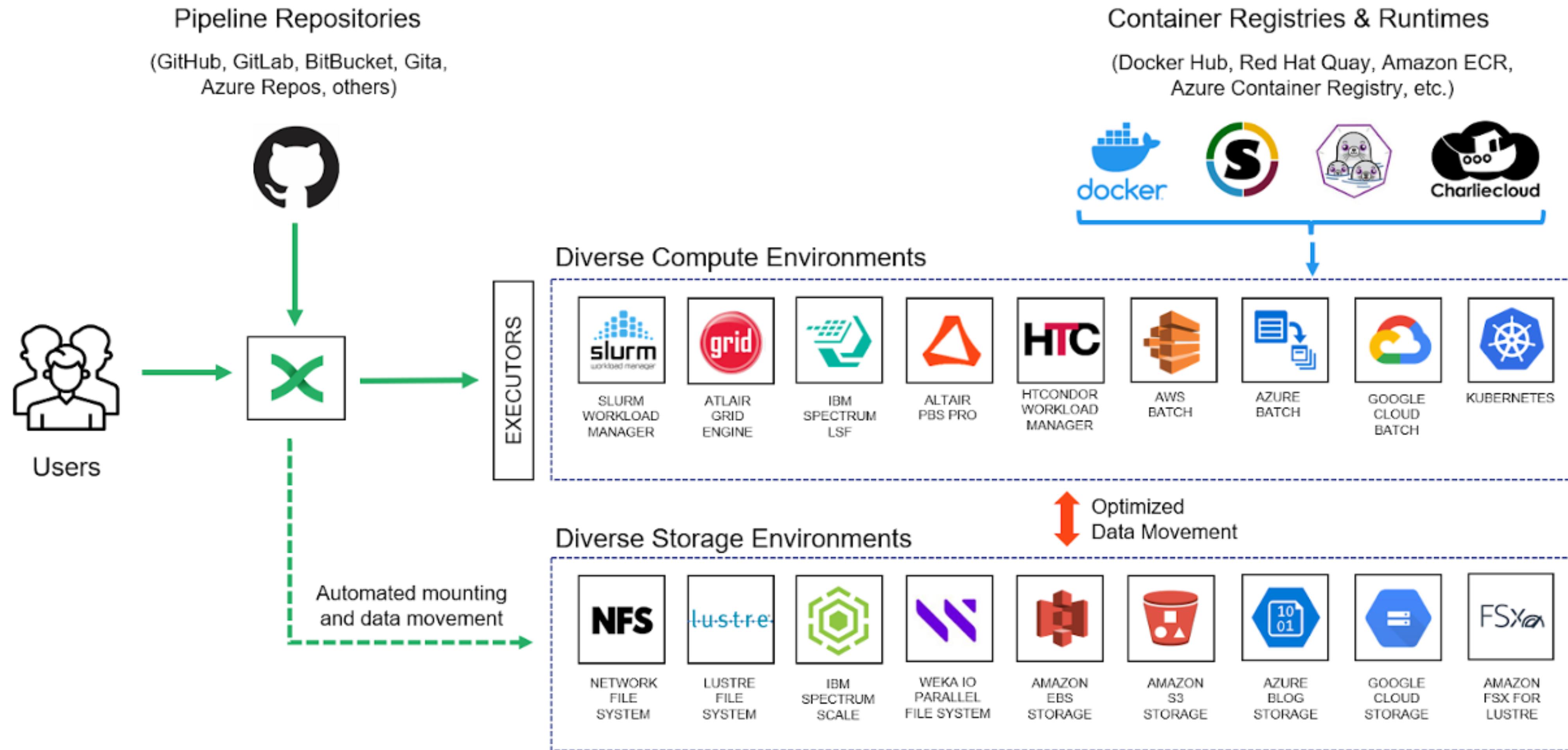


kubernetes



- **Fast prototyping** ⇒ custom DSL that enables tasks composition, simplifies most use cases + general-purpose programming language for corner cases
- **Easy parallelisation** ⇒ declarative reactive programming model based on dataflow paradigm, implicit portable parallelism
- **Self-contained** ⇒ functional approach, a task execution is idempotent ie. cannot modify the state of other tasks + isolate dependencies with containers
- **Portable deployments** ⇒ executor abstraction layer + deployment configuration from implementation logic

nextflow



nextflow



nextflow

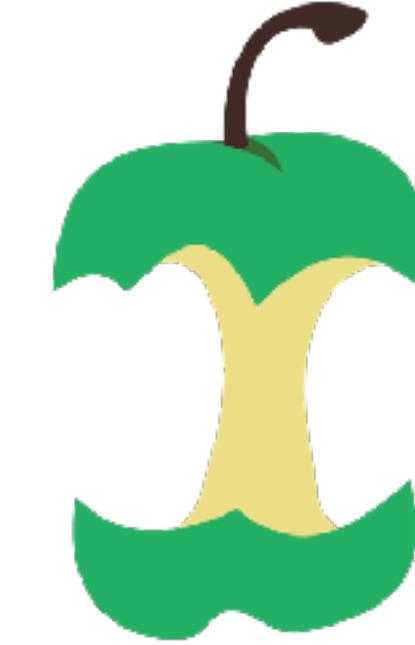
nf-core 



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

<https://nf-co.re>

nf-core



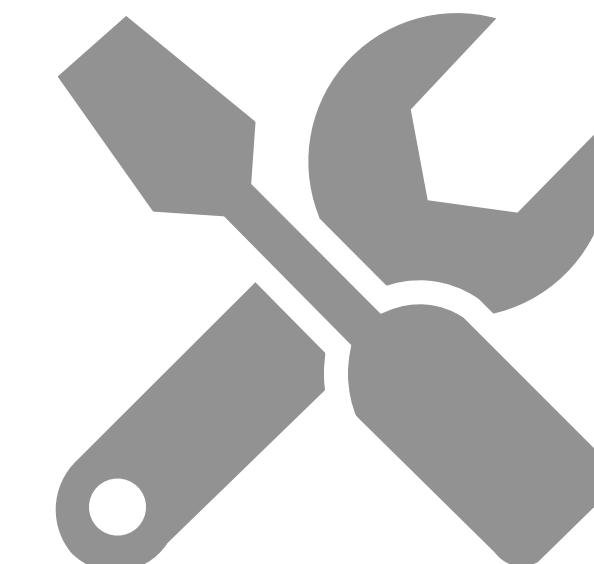
76

PIPELINES

<https://nf-co.re>



TOOLS



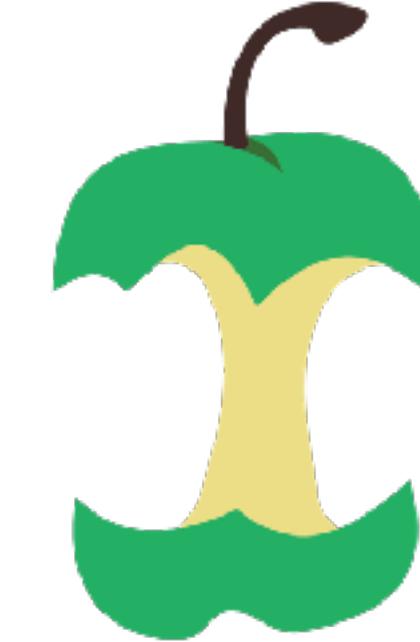
Running pipelines

Writing pipelines

Testing / automation

<https://nf-co.re>

nf-core



>800

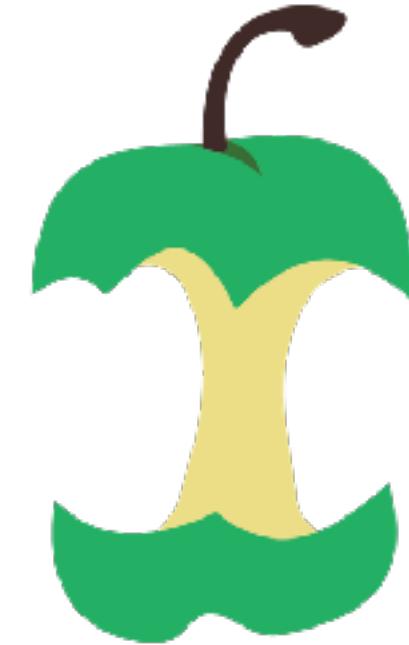
MODULES

>30

SUB-WORKFLOWS

<https://nf-co.re>

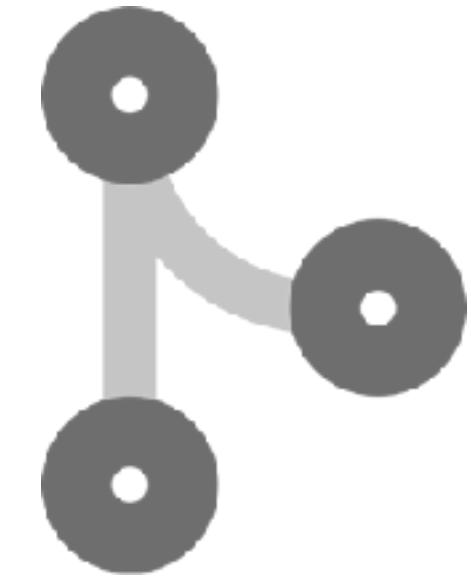
nf-core



Develop with
the community



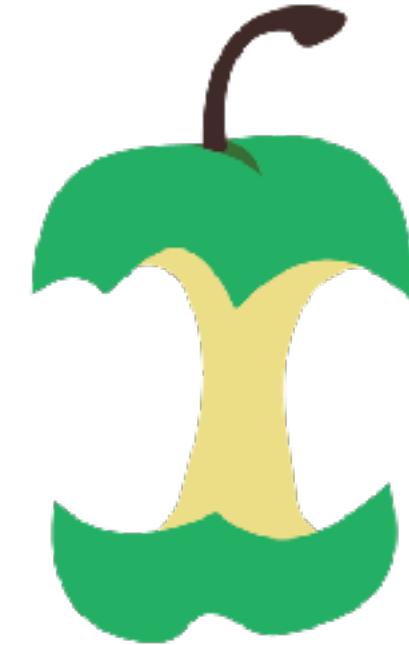
Use a common
template



Collaborate,
don't duplicate

<https://nf-co.re>

nf-core



Tools built for
everyone



Works with any
Nextflow pipeline



Collaborate on
components

<https://nf-co.re>

5154

Slack users

583

GitHub organisation
members

1754

GitHub contributors

3511

Twitter followers

93

Repositories

13.07K

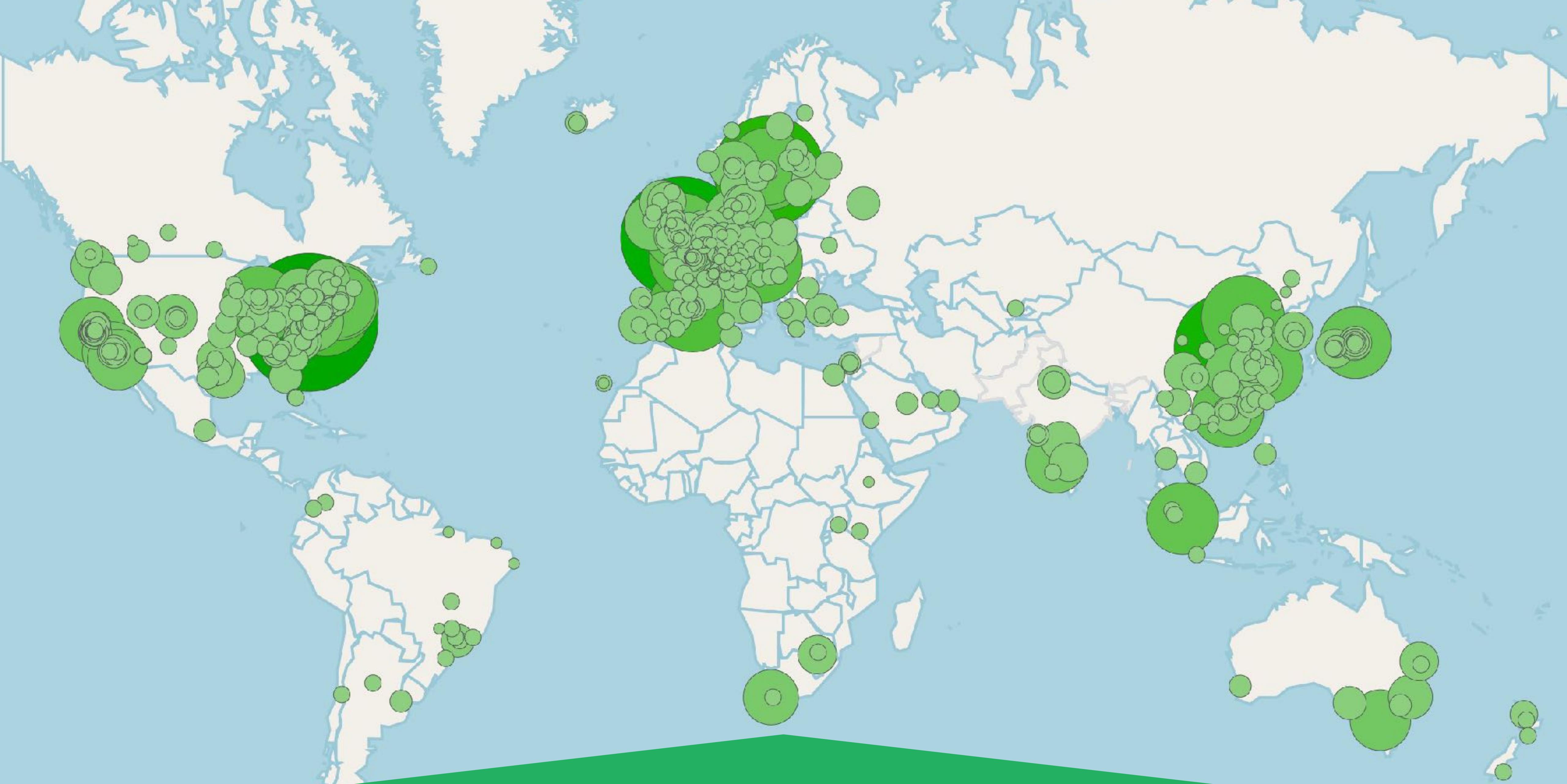
Pull Requests

37.35K

Commits

5.88K

Issues



<https://nf-co.re>

Correspondence | Published: 13 February 2020

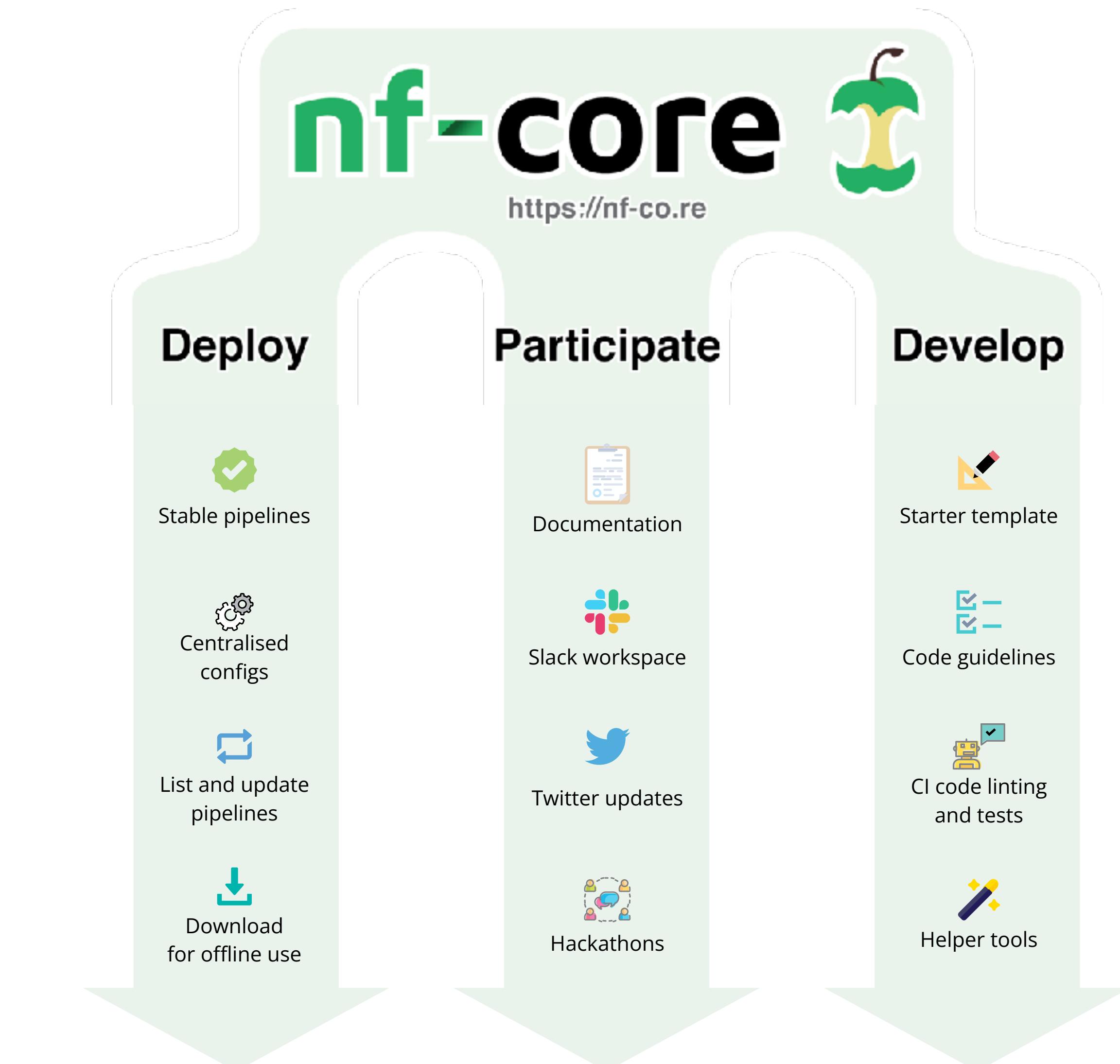
The nf-core framework for community-curated bioinformatics pipelines

Philip A. Ewels, Alexander Peltzer, Sven Fillinger, Harshil Patel, Johannes Alneberg,
Andreas Wilm, Maxime Ulysse Garcia, Paolo Di Tommaso & Sven Nahnsen 

Nature Biotechnology 38, 276–278(2020) | Cite this article

3253 Accesses | 3 Citations | 172 Altmetric | Metrics

To the Editor — The standardization, portability and reproducibility of analysis pipelines are key issues within the bioinformatics community. Most bioinformatics pipelines are designed for use on-premises; as a result, the associated software dependencies and execution logic are likely to be tightly coupled with proprietary computing environments. This can make it difficult or even impossible for others to reproduce the ensuing results, which is a fundamental requirement for the validation of scientific findings. Here, we introduce the nf-core framework as a means for the development of collaborative, peer-reviewed, best-practice analysis pipelines (Fig. 1). All nf-core pipelines are written in Nextflow and so inherit the ability to be executed on most computational infrastructures, as well as having native support for container technologies such as Docker and Singularity. The nf-core community (Supplementary Fig. 1) has developed a suite of tools that automate pipeline creation, testing, deployment and synchronization. Our goal is to provide a framework for high-quality bioinformatics pipelines that can be used across all institutions and research facilities.



Join the community



<https://nf-co.re/join>



nextflow tower



nextflow tower

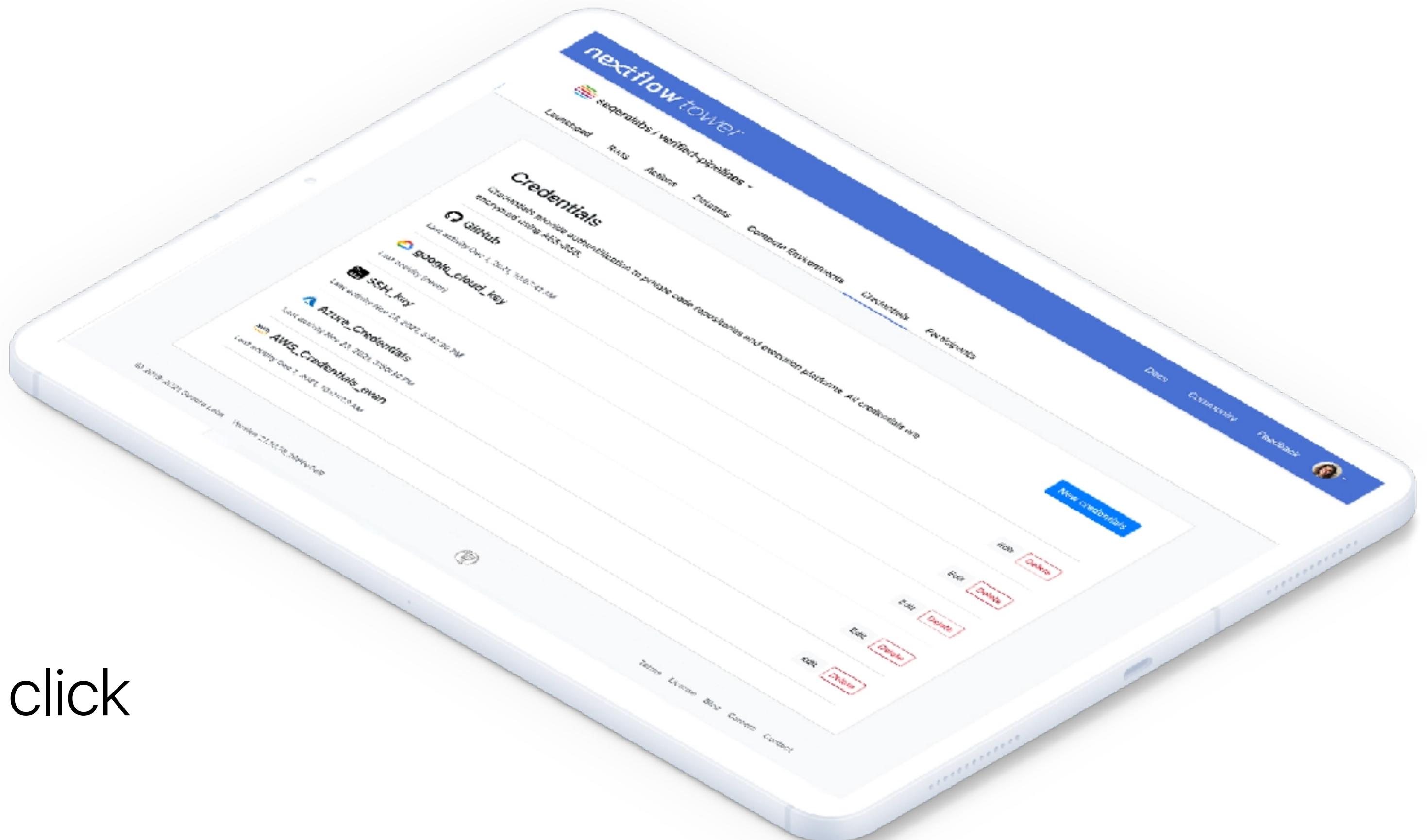
 seqera labs

Intuitive launchpad interface

Launch, manage, and monitor

Share runs and work in teams

Create cloud infrastructure with a click

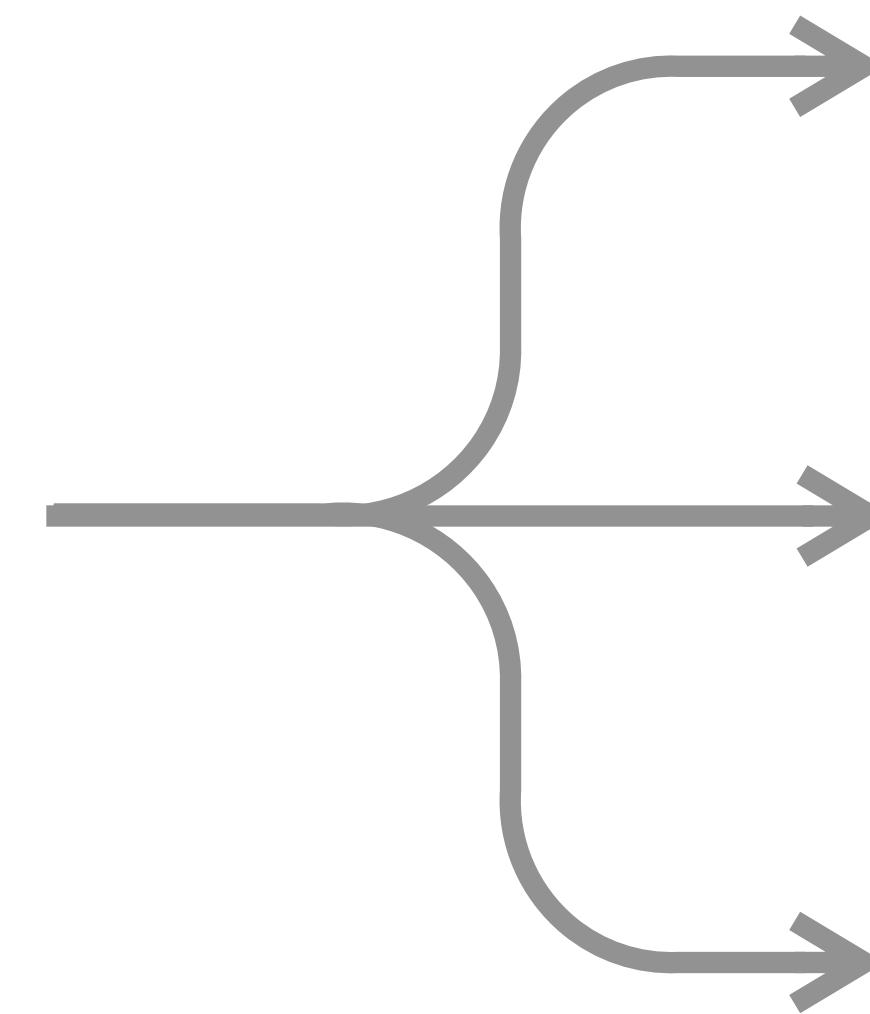
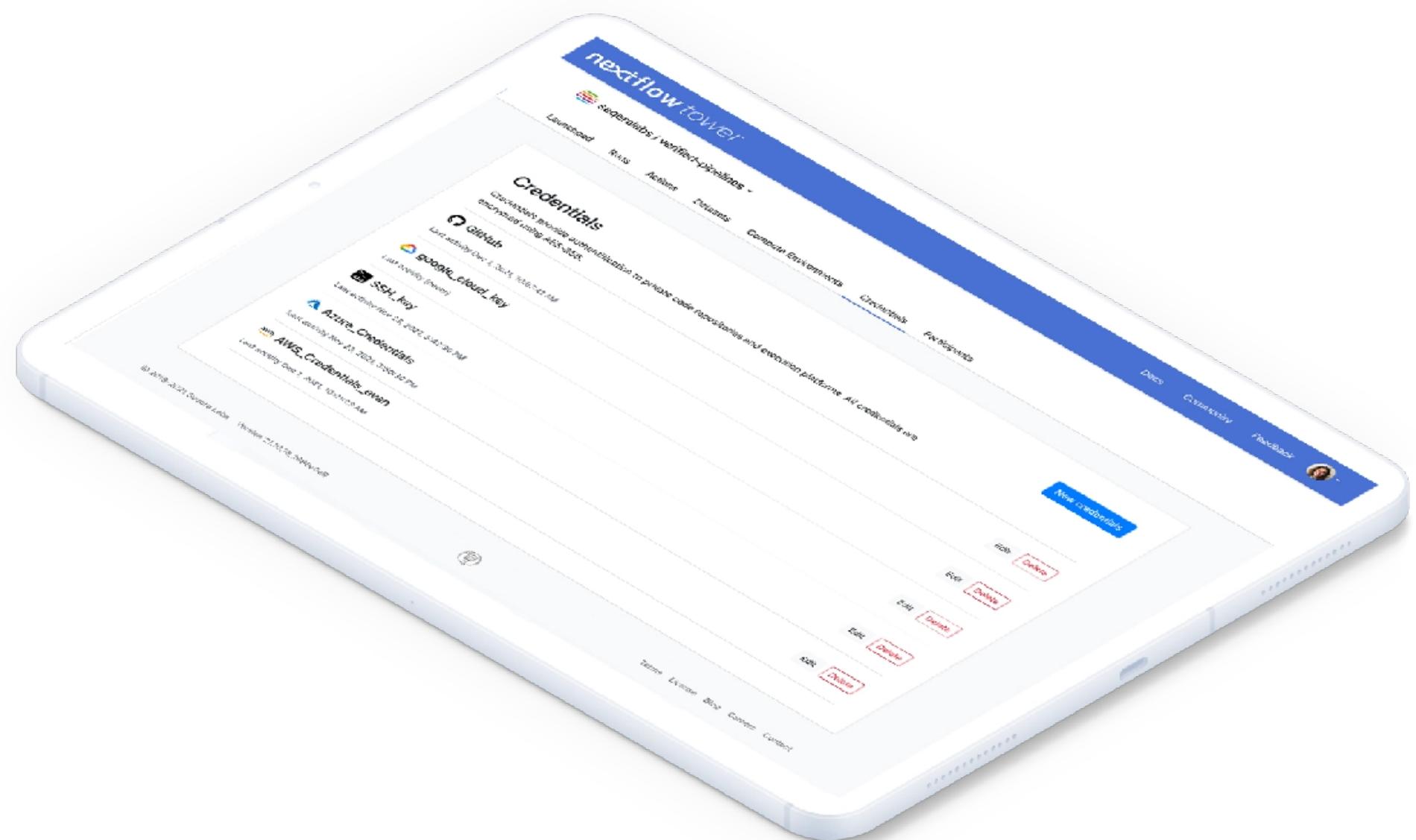


<https://tower.nf>



nextflow tower

 seqera labs



Community: Open source

Cloud: Free & paid tiers

Enterprise: Commercial

<https://tower.nf>

Chris Hakkaart

chris.hakkaart@seqera.io



chris_hakk



christopher-hakkaart



seqeralabs

<https://seqera.io>

Chan Zuckerberg
Initiative



nextflow
SUMMIT 2023

COMING SOON!



<https://nf-co.re>

<https://nextflow.io>

Training environment

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