

Remote data usage

Maxime U Garcia
SciLifeLab | National Genomics Infrastructure
nf-core

ANF - workflow et reproductibilité

2025-11-25..27



The Challenge of Reproducibility

Hidden reproducibility issues are like an iceberg



First, we tried to re-run the analysis with the code and data provided by the authors.

Second, we reimplemented the whole method in a Python package...



Experimenting with reproducibility:
a case study of robustness in bioinformatics
<https://doi.org/10.1093/gigascience/giy077>

The five pillars of computational reproducibility:
bioinformatics and beyond
<https://doi.org/10.1093/bib/bbad375>





Introducing Nextflow

Nextflow is a language, a runtime, and a community

 **nextflow**
pipeline



 **nextflow**
runtime

Write code
in any language



Orchestrate tasks with
dataflow programming



Define software
dependencies via containers



Built-in version control
with Git



Task orchestration
and execution



Nextflow training

Working with Files: https://training.nextflow.io/latest/side quests/working_with_files/

- How files work in Nextflow?
- Difference between local and remote files
- HTTPS - FTP
- S3 - AZ - GS (cloud)
- Using a `base_path` parameter

