

BioFlow-Insight: Leveraging Graph representations to Facilitate the Reuse of Bioinformatics Workflows

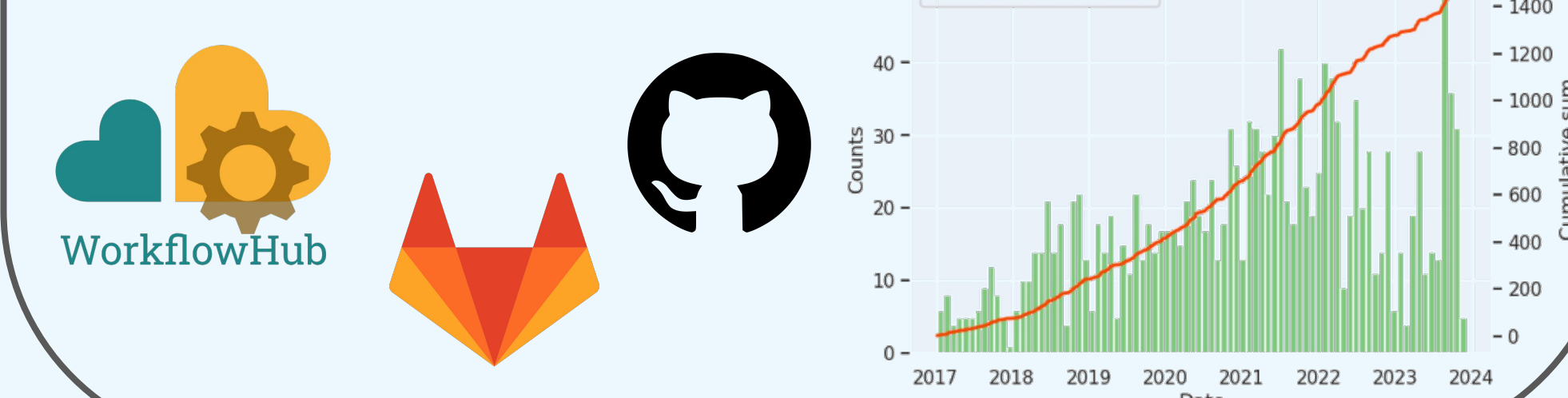
CONTEXT

Scientific Workflows enable Reproducible analyses

With the generation of an immense amount of raw biological data, notably with NGS technology, **scientific workflow systems** have become invaluable tools for bioinformaticians. These systems allow to easily **describe, share, and execute** bioinformatics analyses. Notably, scientific workflows enable the **development of reproducible analyses** through the inherent use of software containers and easy modularisation [1].

Many Workflows available Online

Making workflows available online is becoming common practice, with platforms like **GitHub** and **GitLab** simplifying this process. Additionally, collections like **WorkflowHub** compile shared workflows. As a result, **more workflows are becoming accessible online** [3]. The figure below illustrates this trend by showing the cumulative sum of Nextflow workflows created per month on GitHub since 2017.



Steep Learning Curve

Modern workflow systems like **Nextflow** [2] are code-based. These powerful tools, can be daunting for bioinformaticians who are not comfortable with programming, resulting in a high learning curve [4].



Can be difficult to Understand

Workflows describe complex and sophisticated bioinformatics analyses. Since these workflows are code-based and often spread across multiple files, **understanding their functionality and steps can be challenging** [1]. This task becomes even **harder when the documentation is unclear** which can be the case for many workflows due to the heterogeneous nature of software documentation.

Goldmine for Reuse

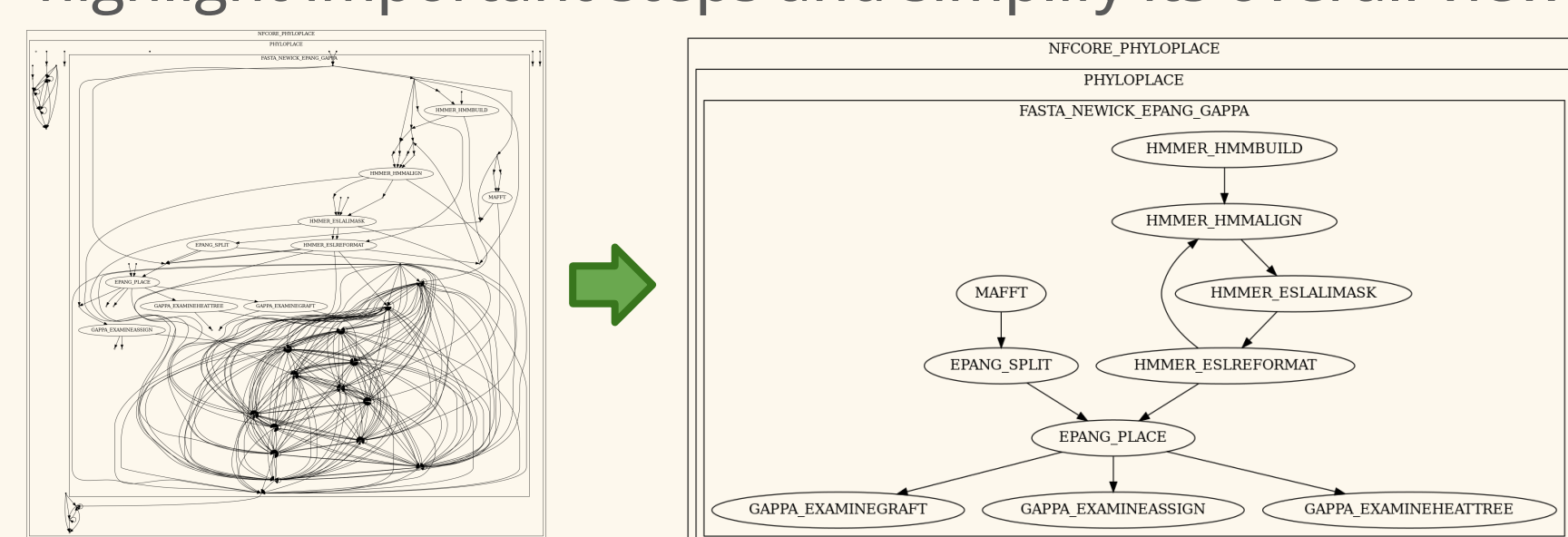
The large collection of workflows available online is an **invaluable resource for developers**. They can opt for a 'reuse-in-part' approach, where a part of a workflow (e.g., a subworkflow) is treated as a black box, and integrated into a new workflow. Alternatively, developers may choose to 'repurpose'. In this scenario, the extracted subworkflow is modified and adapted before integration into a new workflow. **By leveraging existing workflows in this manner, developers can accelerate and simplify the development of new workflows.**

Need for a tool to simplify workflow understandability, thus facilitating workflow reuse.

CONTRIBUTIONS

Graphical Representation

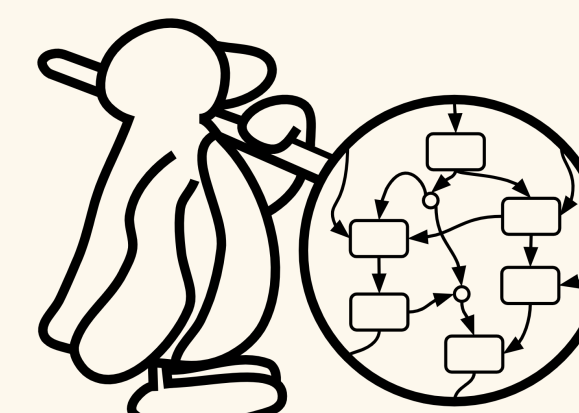
BioFlow-Insight generates multiple graphical representations of a workflow, **depicting its different steps at various levels of simplification**. The lowest level graph shows the entire workflow, while the highest simplification level depicts only the **major steps and their dependencies**. These representations are particularly useful for large workflows, as they highlight important steps and simplify its overall view.



Graph structures generated by BioFlow-Insight and representing the phyloplace nf-core workflow.

BioFlow-Insight

BioFlow-Insight is a tool that analyses the source code of Nextflow workflows without executing them and without any configuration. From the analyses it then generates multiple outputs, notably **graph structures depicting the steps of the workflow, at different simplification levels**. It also provides a detailed description of the workflow utilising a common standard (RO-Crate). Finally it is also capable of detecting programming errors in the code.



Error Detection

By analysing the integrity of the workflow code, BioFlow-Insight is **capable of detecting programming errors in the code**, not necessarily caught by Nextflow during Execution. In this case, BioFlow-Insight stops the analyses and gives a detailed description of the error, **allowing the developer to easily fix it**.

Usable from Website or Command Line



BioFlow-Insight is an **open source tool**. It is accessible via **command line** (installable through pip) or via its **web service**.

RESULTS

Facilitates the Reuse of Workflows

BioFlow-Insight is a freely accessible bioinformatics tool available as both a command-line interface and a web service. It generates the **workflow structure at several levels of granularity**, providing users with a clear understanding of the various workflow steps and their interactions. By utilising the graph structure, individual **workflow steps become apparent, making the functionality easier to grasp**. If my poster is designed correctly, it should convince you that representing a complex process through a graphical representations makes it easier to understand. In the case of workflows, **a better and faster understanding facilitates the reuse of code-based workflows**.

