Empowering HPC Workflows with **Snakemake**: A Comprehensive Training Program

The Snakemake Teaching Alliance: Malte Petersen[1], Christian Meesters[2,3], Florian Boecker[1], Fabian Brand[1], Lukas Hellmann[2], Johannes Köster[4,5], Aasish Kumar[6], Martin Paleico[6]

We present a **unique training program** tailored for researchers and administrators who want to use the **Snakemake workflow system** in an **HPC environment**.
We are developing the material in a **collaborative effort** and are **inviting contributions**.

We explain the benefit of **using wrappers**, demonstrate the **Snakemake wrappers repository**, and show how to include a wrapper in your workflow.

We show **how to configure Snakemake** to adapt to your local HPC cluster setup.

We also include a **checklist** for giving the course at other locations.

The material is **modular** and **configurable**.

It also includes **helper scripts** to generate presentations, handouts, cheatsheets, and attendance certificates.

SNAKEMAKE WRAPPERS

HPC

SITE CONFIG

BEST PRACTICES

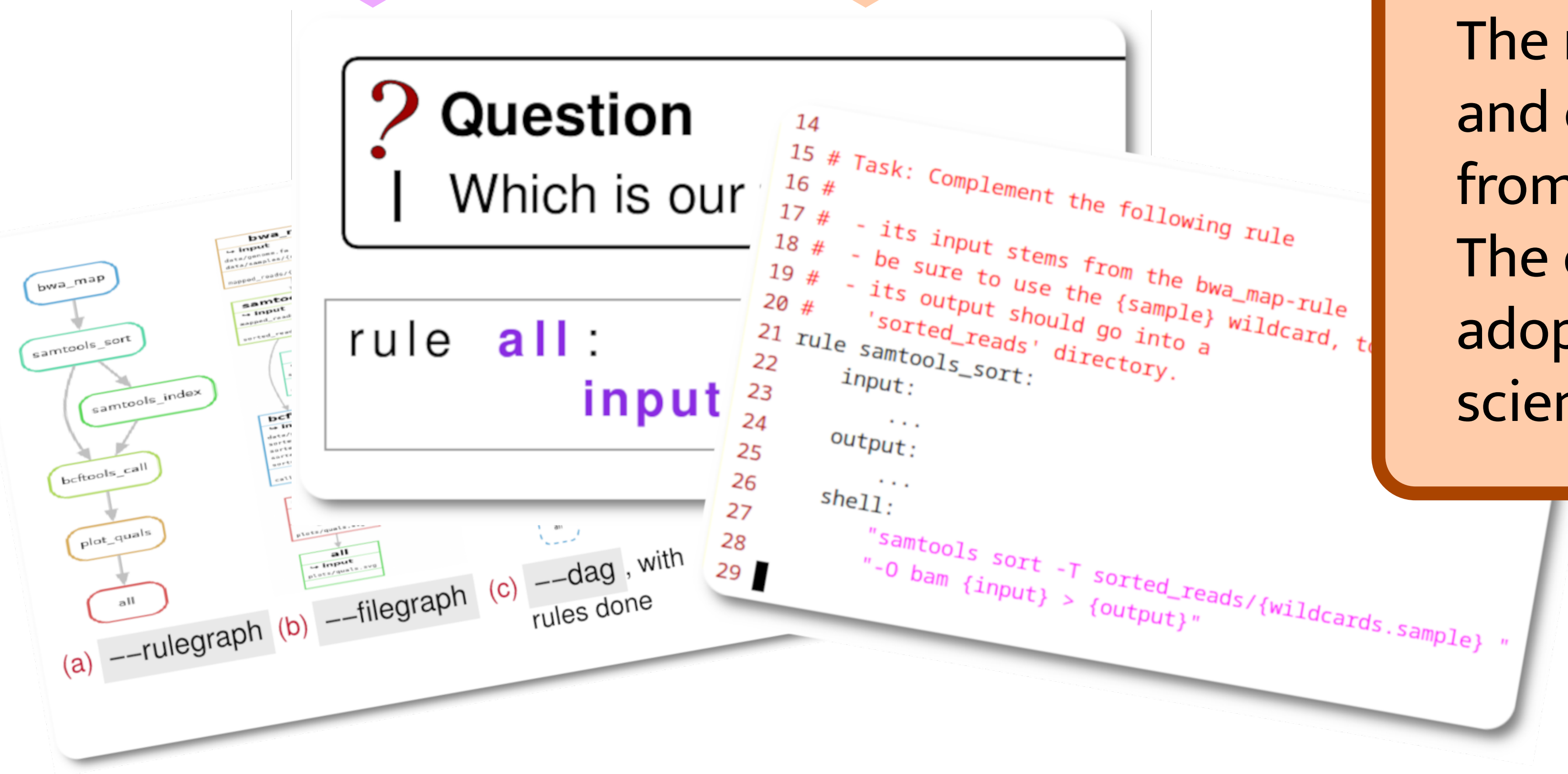
ADAPTABLE MATERIAL

FIELD SPECIFICS

https://s.gwdg.de/hYz6BZ

We present **best practices** in working on a HPC system, and in running your workflow on an HPC cluster.

The material is **field agnostic** and comes with examples from the biological sciences. The examples can easily be adopted for different scientific fields.



Question

Which is our

```
rule all:
        input
```

```
# Task: Complement the following rule
#
#   - its input stems from the bwa_map-rule
#   - be sure to use the {sample} wildcard, t
#   - its output should go into a
#     'sorted_reads' directory.
#
rule samtools_sort:
    input:
        ...
    output:
        ...
    shell:
        "samtools sort -T sorted_reads/{wildcards.sample} "
        "-O bam {input} > {output}"
```

For Users

Using the well curated workflows from the **Snakemake workflow catalog** on HPC clusters requires attention to detail.

Our material is tailored to first time HPC users and answers the relevant questions by example.

For Creators

Our course introduces workflow programmers to HPC conformant programming.

We teach **distributed execution**, optimal **resource parameterisation**, and how to **avoid I/O contention**.

For Admins

Workflow managers offer a solution to the Big Data problem for numerous scientific domains. We teach Snakemake's cluster-wide configuration options - **helping administrators** support their target user groups.

Affiliations: 1: University of Bonn; 2: University of Mainz; 3: NHR Südwest; 4: University of Duisburg/Essen; 5: IKIM, University Medicine Essen; 6: GWDG, Göttingen
Github repository: https://github.com/cmeesters/snakemake-hpc-teaching-material | **Snakemake:** https://snakemake.github.io

deRSE24 - Conference
for Research Software
Engineering in Germany
(March 5-7, 2024)