Snakemake Workshop

or how to cook Spätzle

Jonathan Amme

Reiner Lemoine Institut

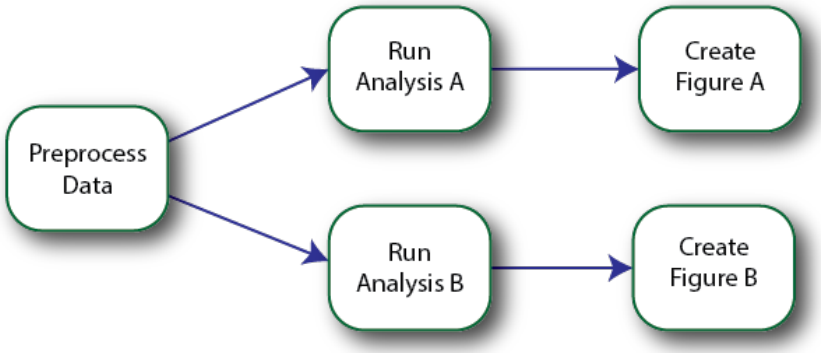
January 21, 2021

Workflow management system

Why would you need a workflow management system (WMS)?

Example workflow 1

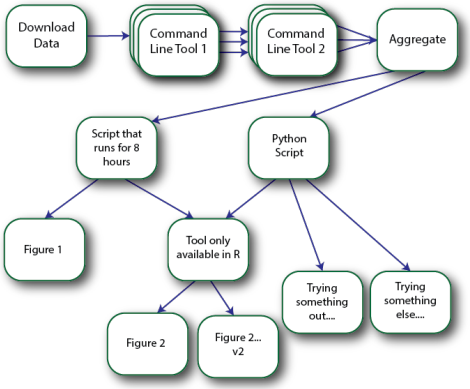
Simple workflow



Source: deto

Example workflow 2

More complex workflow (more steps, different languages, long processing time for individual steps)



Source: deto

How can snakemake help you?

REINER LEMOINE INSTITUT

- Lightweight workflow management system
- Text-based, python syntax
- Split large data-/workflow into single steps, defined by rules
- Infers dependencies and execution order (DAG)
- Reproducible and scalable data analyses
- Supported languages: BASH commands, Python, Inline python code, R script, R markdown file

Rules 1

Define how to obtain input files from output files

```
rule copy_single_file:
    input: "/path/to/file_in"
    output: "/path/to/file_out"
    shell: "cp {input} {output}"
```

Rules 2

Generalize rules using wildcards -> snakemake infers file names from dependencies (DAG)

```
rule copy_multiple_files:
    input: "/path/to/{file}"
    output: "/path/to/{file}_copied"
    shell: "cp {input} {output}"
```

Rules 3

Use named input files

```
rule annotate_multiple_files:
    input:
      files="/path/to/{file}",
      annotation="/path/to/annotation"
    output: "/path/to/output/{file}_annotated"
    shell: "cat {input.files} {input.annotation} >> {output}"
```

Rules 5

Use python script

```
rule copy_single_file:
    input: "/path/to/file_in"
    output: "/path/to/file_out"
    script: "my_script.py"
```

Rules 5

Use python inline code

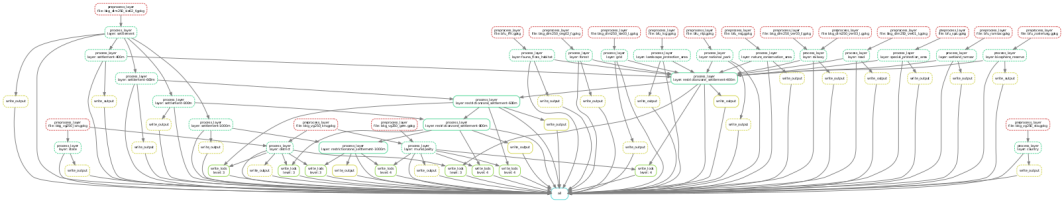
```
rule copy_single_file:
    input: "/path/to/file_in"
    output: "/path/to/file_out"
    run:
        in = open(input, "r")
        out = open(output, "w")
        for line in f:
            out.write(line)
        in.close()
        out.close()
```

More features

- Parallelization (threads, can be even run on clusters such as AWS S3)
- Resource allocation (entire workflow or per rule)
- Suspend and resume
- Logging
- Modularity
- Report generation
- …

Example 1

Part of the workflow in a current project



Example 2

Now let's cook some delicious Spätzle! (Link to Repository)