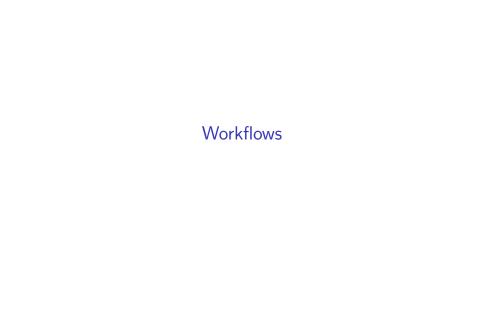
# Workflows Interest Group

November 26th, 2020



## Aims & Scope

- ► Foster discussion and knowledge sharing regarding research workflows/pipelines
- Develop and share best practices for setting up workflows, recommend solutions
  - ▶ Different audiences may require different solutions
- Create an inventory of workflow systems used in CLARIAH
- Promote interoperability between CLARIAH tools
  - More attention for underlying standards and protocols
  - This details of these are mostly in the scope of the other interest groups on Annotation, Text, Audio/Video, Linked Data.

#### What is a workflow?

There are various paradigms for workflow systems and various layers of complexity, a simple shell script or Makefile can already be considered a basic form of a workflow.

- goal oriented (GNU Make, luigi, snakemake, Airflow)
- procedural (shell scripts)
- data-flow oriented (Nextflow, SciPipe)

We can also make a distinction on the nature of the software components that are connected; are those local components (local processes) or remote networked components (webservices).

## Characteristics

- ➤ **Scheduling**: Determining which sequence of tasks leads to the desired end result
- ► **Execution**: Execution of tasks, possibly with automatic parallellisation/abstraction
  - Abstraction of pipeline logic from execution layer: local execution, computing cluster, cloud platform
- ▶ **Monitoring**: Real-time monitoring of a running workflow and ability to inspect logs on failure
- ► **Correction**: Ability to resume the workflow after failure and manual intervention; and/or automated fallbacks
- ► **Reproducibility**: Given the same input, the workflow should ideally produce the same output every time
  - This requires solutions for software deployment and versioning and is mostly in the realm of the DevOps Interest Group.
  - ► Handling of provenance data
- ▶ **Discovery**: automated discovery of possible workflows
- ► Interaction: Human intervention (steering/inspection/visualisation) in at specific points in a workflow

## **Tasks**

- Not really defined yet.
- ▶ Suggestion: Start with an inventory of solutions in use.