

# Workflow engines and workflow capable environments in the SSH

This document gives a (certainly ) non-exhaustive overview wrt workflow engines and workflow capable environments used or experimented with in the SSH domain.

Most entries are used (or developed) from organisations affiliated with the CLARIN infrastructure domain, so NLP workflows. But also broader Humanities and of course Digital Humanities use such systems.

## CLARIN-DK

### Text Tonsorium

The Text Tonsorium is an *NLP workflow management* system that automatically combines the necessary natural language processing tools to achieve your goal, in a way similar to how a trip planner computes the best route from your current position to your destination.

<https://cst.dk/WMS/> , <https://github.com/kuhumcst/DK-ClarinTools>

and also [https://cst.ku.dk/english/ansatte/?pure=en%2Fpublications%2Fworkflow-management-in-clarindk\(7f1ec585-ea2b-4ce1-a712-38bcab4a4d53\).html](https://cst.ku.dk/english/ansatte/?pure=en%2Fpublications%2Fworkflow-management-in-clarindk(7f1ec585-ea2b-4ce1-a712-38bcab4a4d53).html), CLARIN DK workflow manager ([https://www.clarin.eu/sites/default/files/clarin2020\\_p\\_2.1.3\\_jongejan.pdf](https://www.clarin.eu/sites/default/files/clarin2020_p_2.1.3_jongejan.pdf))

## CLARIN NL

**TTNWW** was developed in the CLARIN NL project and was based on Taverna. It was withdrawn in 2019 for maintenance problems.

## CLARIN D

**Weblicht** [https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main\\_Page](https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page) the best known and most populated workflow environment in the CLARIN domain

## DANS SSHOC DataVerse Cloud

In the SSHOC project the hosting of container based services and workflow creation is worked on and investigated. See

<https://www.aclweb.org/anthology/2020.lr4sshoc-1.1.pdf> for the tentative workflow implementation

## GATE (Sheffield Univ. )

long-time existing NLP software first as local installed software and afterwards as a set of services. Note that the **Parthenos platform** gives access to GATE services

See <https://gate.ac.uk/family> for software overview and <https://cloud.gate.ac.uk/shopfront> for the GATE cloud services and GATE Teamware for a workflow manager <https://gate.ac.uk/teamware/>

## PARTHENOS platform (result from the PARTHENOS community cluster project)

<https://parthenos.d4science.org>

D4Science is hosting platform for (building and operating VRE's whereof some offer workflow facilities. The most interesting for CLARIAH is PARTHENOS developed in the PARTHENOS project offering a number of

Humanities , CH and NLP (e.g. GATE) services that can be invoked in sequence

**ELG platform** (European Language Grid project), (in development)

A platform allowing for testing (also workflows) of industrial strength NLP services, WU wide consortium, mainly DFKI, CUNI, ATHENA, ...). ELG platform works with containerised services. Container orchestration with Kubernites. A prototype was demonstrated at the ELG conference Dec 1-3 2020

See "A Workflow Manager for Complex NLP and Content Curation Pipelines " [https://www.european-language-grid.eu/wp-content/uploads/2020/05/IWLTP2020\\_Moreno-Schneider.pdf](https://www.european-language-grid.eu/wp-content/uploads/2020/05/IWLTP2020_Moreno-Schneider.pdf)

LAPPS project (US) using **Galaxy** workflow manager <https://galaxyproject.org/learn/advanced-workflow/>

See <https://www.aclweb.org/anthology/L16-1073.pdf>

See also <https://www.slideshare.net/cneudecker/collaborative-workflow-development-and-experimentation-in-the-digital-humanities> for a relevant older (2013) overview. (note the early statement about preventing reinvention of the wheel)

**Taverna** (<https://taverna.incubator.apache.org>) used by many projects , some also in Humanities e.g. UPF CLARIN effort around 2010. Now promoted via EOSC-PILAR [https://eosc-pillar.d4science.org/catalogue-eoscpillar?path=/dataset/taverna\\_workflow\\_system](https://eosc-pillar.d4science.org/catalogue-eoscpillar?path=/dataset/taverna_workflow_system)

**TextFlows** workflow management web platform for text mining and NLP was developed

See <https://www.sciencedirect.com/science/article/pii/S0167642316000113>

**Nextflow** (<https://www.nextflow.io/index.html>) general data-driven pipe-lines

use in the SSH:

- RUN tool family: LaMachine, Nedrab Pipeline etc. (See <https://www.clarin.eu/resource-families/tools->