

Creating Flexible and Transparent Data Processing Pipelines using Common Workflow Language

Janneke van der Zwaan, Wouter Smink,
Anneke Sools, Gerben Westerhof, Bernard
Veldkamp, and Sytske Wiegersma

netherlands

eScience center

by SURF & NWO

UNIVERSITY OF TWENTE.

Core Functionality

- **Transparently combine data processing tools**
 - **Common Workflow Language (CWL)**
 - `scriptcwl`
 - `nlppln`
 - **Use case: text processing**



Common Workflow Language

- **Standard for data processing workflows**
 - <http://www.commonwl.org/>
- **Processing step**
 - **Specification of a command line tool**
- **Workflow**
 - **Sequence of processing steps**

 Apache-2.0



COMMON
WORKFLOW
LANGUAGE



nlppln + scriptcwl

- **nlppln**
 - Provide CWL steps for existing (NLP) tools
 - Create CWL steps for new (NLP) tools
 - Convert Python to command line tools
- **scriptcwl**
 - Combine steps into pipelines

<https://github.com/WhatWorksWhenForWhom/nlppln/>

<https://github.com/nlesc/scriptcwl>

 Apache-2.0



Demo

- **Tool creation: lowercasing text**
- **Pipeline: anonymize text**



Development and Maintenance

- **CWL: active community**
- `scriptcwl`
- `nlppln`
 - **Developed and maintained by me**
 - **More projects/people interested**



Development Goals

- **Combining more generic and project specific functionality**
- **Gather feedback from (potential) users**

