Converting Scripts into Reproducible Workflow Research Objects

L. A. M. C. Carvalho*, K. Belhajjame**, C. B. Medeiros*
* University of Campinas (UNICAMP), ** University of Dauphine
lucas.carvalho@ic.unicamp.br

Problem
Understand the processing units and data dependencies of script-based experiments, reproduce these experiments and reuse their model and data.

Objectives of this Research
To support the process of conversion of script-based experiments into a reusable and reproducible workflow-based representation.

Requirements
The methodology was based on requirements elicited given our experience and collaboration with scientists who use script-based experiments.

1. Produce workflow-like view of the script.
2. Create executable workflow and compare execution of workflow and script.
3. Modify of the workflow resources.
4. Record provenance data.
5. Aggregate all resources to support Reproducibility and Reuse.

Results
Methodology to guide curators through (a) conversion from script code to an executable workflow, and (b) construction of a Workflow Research Object that bundles the workflow plus resources needed to reproduce the experiment.

Running Example: Molecular Dynamics

Step 1
Manually annotate the script code using YesWorkflow tags to generate the abstract workflow.

Step 2
Manually implement the executable workflow based on the abstract workflow and reusing the corresponding script block for each activity.

Step 3
Change some resources (e.g. data set, algorithms) from the initial executable workflow to reproduce the experiment.

Step 4
In steps 1, 2 and 3 check the quality of the conversion and provide annotations to describe the workflow.
- Use provenance data
- Check the quality of the conversion process.
- Check the Reproducibility of the conversion.
- Check to verify the soundness of the workflow.

Step 5
Aggregate all resources required to reproduce the experiment into a bundle called Workflow Research Objects.

Acknowledgments:

WRO available at http://w3id.org/w2share/s2wro/

Figure 1. Script code + YesWorkflow tags.
Figure 2. Abstract workflow derived from the YesWorkflow tags.
Figure 3. Executable workflow implemented using Taverna system.
Figure 4. New version of the executable workflow.
Figure 5. Provenance data describing the conversion process of the script.
Figure 6. Workflow Research Objects.