doi:10.20396/revpibic262018805

XXVI Congresso de Iniciação Científica Unicamp
17 a 19 de outubro Campinas | Brasil

Evaluating Quality Dimensions in Scientific Workflows

Leila P. Zwanziger, Claudia M. Bauzer Medeiros

Abstract
Open Science aims to make research products more reusable and accessible. One of these products are workflows - step-by-step descriptions of scientific computational processes, usually expressed in terms of inputs, outputs and activities, or processes. This work explores the notion of quality of scientific workflows, thereby allowing scientists to record quality properties of their experiments.

Key words:
Scientific Workflows, Data Management, Quality of scientific experiments

Introduction

Scientific data management is important to promote reuse, reproducibility and interoperability of experiments. The term “data” encompasses many research artifacts, such as workflows. Workflows are models of computer processes, expressed through activities and data flows.
To better support reproducibility, we propose to link quality data to workflows, thereby helping scientists assess the quality of experiments.

Results and Discussion

Our methodology is the following:

1. Analyze quality information stored in scientific platforms.
2. Interview researchers in Biostatistics, Agronomical Engineering and Ecology.
3. From 1. and 2, define the main quality dimensions for workflows. A quality dimension is a set of attributes that represent an aspect of quality. The overall quality of a workflow is given by the combination of its dimensions.
4. Implement a prototype to support specification of workflow quality dimensions.

Our dimensions are:

- Performance: Describes the speed with which the workflow completes its tasks.
- Correctness: Describes the correctness of the results generated by the workflow as a whole and by its processes.
- Usability: Describes the ease to use that workflow.
- Reliability: Describes external reasons to believe that the workflow generate correct results.
- Utility: Describes which tasks the workflow solves.

We also determined a mandatory list of quality evidence data for quality assessment. These were selected to be the most general, creating a common ground for all research areas and facilitating documentation. Each quality data type describes one or more quality dimensions.

Chart 2. Quality data types and their quality dimensions

Workflow runs	Performance
Intermediate files	Correctness, utility
Input/Output files	Correctness, utility
Libraries/web services used	Correctness, usability
Use/Installation Guide	Usability
Contact information	Usability, reliability
Academic reference	Utility, reliability
Team/authors	Reliability
Financing	Reliability
Citations	Reliability
Hypothesis	Utility, reliability, correctness
Test Cases	Correctness
Analysis’s description	Utility, correctness

An extension of the Quality Flow Ontology is proposed to allow storage. Elements in light grey where added to the original ontology (dark grey)

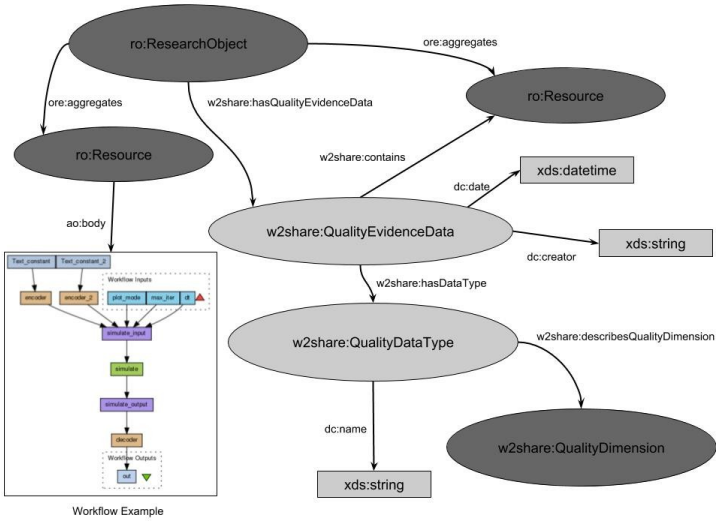

Image 1. The extension of the Quality Flow Ontology

Conclusions

Our work will help scientists assess the quality of each other’s experiments. We are now implementing it.

SOUSA, R. B. (2015). *Quality Flow: a collaborative quality-aware platform for experiments in eScience.* UNICAMP.MSC Thesis.