

Summary

An aspect of open science is to be able to make workflows available. This helps in the reproducibility of scientific computation processes. However the problem was that there was a lack of a standard way to publish and link these workflows.

Therefore, a system based on Linked Data principles and publishing workflows and associated resources on the Web was developed. All the resources are URI accessible and WINGS workflow system is used. The ontological relationships are represented in RDF and are stored in triple stores.

But, along with the workflow which is both human readable and machine readable, what would help to make the explanation clearer is human readable text that details the workflow - the flow of inputs to outputs through intermediate steps and processing.

Of course, the researchers do themselves provide such explanation, but it is often incomplete or ambiguous. Also, computational processes are hard to explain.

Here, DANA is trying to automatically generate such text description. It should be accurate and inspectable. Furthermore, it is customizable to give different views based on the focus - such as data, execution, software etc. In short, the text is generated by filling in the blanks in a few predefined explanation patterns by issuing SPARQL queries to a knowledge base that holds experiment related information.

The data narratives so generated are intended to be publishable by the authors. For this, they must be able to provide feedback on any deficiencies that they spot in the generated narrative account. This could have to do with incorrect information, or dissatisfaction with the level of abstraction/specification provided in the view. Such an exercise might also allow researchers to find mistakes if any in their own work, or might help them simplify their workflow. The feedback so got must be incorporated into the narrative account. And perhaps allow a facility for researchers to modify the workflow or rerun modified workflow if a change has to be done.

As more feedback data becomes available, machine learning techniques might be used to change the generated text to be more suitable.