# Validating YesWorkflow annotations using declarations of fine-grained dependencies of script outputs on script inputs

Timothy McPhillips
University of Illinois (UIUC)
tmcphillips@absoluteflow.org

Qian Zhang
University of Illinois (UIUC)
zhangqian06@gmail.com

Bertram Ludäscher
University of Illinois (UIUC)
ludaesch@illinois.edu

## Abstract

YesWorkflow is an annotation language for declaring the dataflow structures otherwise hidden in scripts implementing scientific workflows. An accompanying software toolkit extracts these annotations, constructs workflow models corresponding to annotated scripts, and renders the resulting models graphically. YesWorkflow visualizations reveal the computation steps involved in producing each script output nd the paths taken by data through those steps during a script run. By exposing these models as Datalog facts, YesWorkflow enables logic programs both to query the workflow graph and to produce additional visualizations that highlight how specific outputs are computed by cript. These capabilities require not just that the individual YesWorkflow annotations within a script be correct, but also that the collection of annotations in a script be both internally consistent and sufficiently complete to support these queries. Here we describe a way for researchers to confirm the completeness and internal consistency of the YesWorkflow models implied by the annotations in their scripts. The approach employs new YesWorkflow annotations that the researcher applies to the script as a whole to declare the fine-grained dependencies of each script output on specific script inputs. Because annotations on the individual steps comprising a script are applied independently of these new, script-level data-dependency declarations, the latter can be used to validate the former.

## 1. Introduction

## 2. Conclusions

***Acknowledgments.*** Work supported in part by the National Science Foundation under awards DBI-1356751 (Kurator), SMA-1439603 (SKOPE), ACI-0830944 (DataONE).

## References