

Authors: Jay Jay Billings, Matteo Turilli, Shantenu Jha

All Existing Workflow systems are megaliths that should be rebuilt

A discussion of future online data analysis platforms will inevitably turn to questions around workflows and workflow management systems that support and ostensibly streamline analysis for users. Of all the questions that are asked, the question that is too often ignored is “What do users actually want *to do with* their workflows and workflow management systems in this online data analysis platform?”

There are many reasons why the answer to the last question is extremely different than the answers to the other, “cart before the horse” questions. First, those of us in the workflows community generally present ourselves as “here to help” by offering a solution that, at first glance, looks as powerful, production ready, and usable as the computer from Star Trek, when, in reality, “practical issues” really mean that we are offering something akin to Windows 3.1.

For that matter, what the is a workflow or workflow management system? Surveying five papers will often give five different definitions of both as opposed to adopting historical definitions from the Workflow Management Coalition (or others) from the 90s. This has led to what the Oak Ridge Leadership Computing Facility’s 2016 report refers to as “the current proliferation of workflow systems in response to perceived domain-specific needs of scientific workflows.” Essentially what we have is a set of collectively useless point solutions with absolutely zero hope of ever being interoperable, whereas the development of a more general theory of workflows and workflow management systems would have created a collection of well-formed solutions based on common “building blocks.” What’s worse is that even with perfect provenance, there is no hope of ever having truly reproducible workflows across workflow systems because the workflow descriptions are not interchangeable!

Perhaps a final nail in the coffin is that deploying or extending workflow management systems is often impossible without human sacrifice or career suicide: one must essentially give up everything and become a developer for the workflow project! In too many cases “using X to manage my workflows” means grossly rewriting or even contributing large parts of X!

Thus, we have come to the conclusion that all existing workflow management systems are monolithic, even megalithic, systems that should be broken apart and rebuilt to better meet future needs of the community.