Towards an Infrastructure for Enabling Systematic Development and Research of Scientific Workflow Management Systems
workflows are becoming more complex and require more sophisticated workflow management capabilities

workflows now...

can analyze terabyte-scale data sets

be composed of millions of individual tasks

require coordination between heterogeneous tasks

manage tasks that execute for milliseconds to hours

can process data streams, files, and data placed in object stores
WMS development is ad hoc!

*there has been a profusion of different, yet similar in functionality, systems*

https://s.apache.org/existing-workflow-systems

A community list of WMSs includes **280+** systems

https://github.com/pditommaso/awesome-pipeline

A curated list of pipeline toolkits includes **140+** systems
“there is a pressing need to bring the workflows community together (both from academia and industry), with the goal of defining a rigorous methodology for characterizing current and upcoming WMS capabilities, and supporting the innovations necessary to address the range of imminent workflows challenges”
workflows
researchers
developers

science and engineering
communities

cyberinfrastructure
research computing centers
national CI providers

surveys & workshops

- common knowledge taxonomy
- experimental methodology
- WMS inventory
- blueprint for an R&D infrastructure

https://workflowsri.org
surveys
workflow research and developers, science and engineering domain, and cyberinfrastructure practitioners

community RI workshop
define the goals of the planning project, participation of key stakeholder groups (WMS developers and researchers, users, and CI)

CI and science workshop
provide an opportunity for these communities to describe their challenges and propose ways they could contribute to this initiative

developers workshop
discuss challenges, identify stakeholders, and define a taxonomy of WMSs capabilities and associated benchmarks

CCRI planning workshop
review and consolidate outcomes and create a blueprint document for a community research infrastructure

CCRI Proposal
develop an open, community infrastructure for scientific workflows research and development

Fall 2020
The roadmap includes milestones for:
- Community RI Workshop: Define the goals of the planning project, participation of key stakeholder groups (WMS developers and researchers, users, and CI)
- CI and Science Workshop: Provide an opportunity for these communities to describe their challenges and propose ways they could contribute to this initiative
- Developers Workshop: Discuss challenges, identify stakeholders, and define a taxonomy of WMSs capabilities and associated benchmarks
- CCRI Planning Workshop: Review and consolidate outcomes and create a blueprint document for a community research infrastructure
- CCRI Proposal: Develop an open, community infrastructure for scientific workflows research and development
Towards an Infrastructure for Enabling Systematic Development and Research of Scientific Workflow Management Systems

getinvolved@workflowsri.org

https://workflowsri.org

Rafael Ferreira da Silva
Principal Investigator

Henri Casanova
Co-Principal Investigator

Kyle Chard
Co-Principal Investigator

Tainã Coleman
Graduate Research Assistant