Summary of UFS Workflows Workshop

Benjamin Cash (GMU/COLA)

UFS Workflows Workshop Organizing Committee

Ligia Bernardet (GSL, DTC)
Jennifer Bolton (NCAR)
Benjamin Cash (GMU)
Arun Chawla (EMC)
Mike Ek (NCAR, DTC)
Evan Kalina (CIRES@GSL, DTC)
Louisa Nance (NCAR, DTC)
Mariana Vertenstein (NCAR)

UFS Workflows Workshop Background

- Motivation:
  - Diverse set of workflows for models at NCEP; inflexible, lack documentation & portability.
  - UFS calls for broader community alignment, e.g. in workflows.
- Initial high-level/small workflows meeting
  - Summer 2019 in Boulder:
  - Concerns expressed about need for greater coordination wrt workflows.
- Follow on: Advanced Workflows (WG) telecon, 22 November 2019:
  - MRWeather App includes model & CIME workflow.
  - HSUP workflow activity – use CIME/CROW.
  - UFS Workflow Workshop planned for Fall 2019; postponed to Spring 2020.

- Workshop planning (January – April 2020):
  - Organizing committee established January 2020
  - CROW Review and Workflows Workshop combined
    - Topics/audiences similar.
  - Telecons to discuss topics, invitee list, etc., target ~30 attendees.
- Workflows Workshop, 29-30 April 2020 (online)
  - DTC provided logistical support.
  - Many pages of slides, notes, and chat logs
  - Workshop report pending.

UFS Workflows Workshop: Overview

- Very well attended
  - Broad swath of the UFS community and other partners represented
  - Initial target was for roughly 30 attendees
  - Actual attendance was over 60!
  - Most people attended both CROW Review and Workflows Workshop
  - Identifying a platform that was permitted for all attendees was not trivial

- Virtual format was surprisingly engaging!
  - Each presentation generally represented a group effort
  - Use of chat to ask questions led to unexpected result of other group members answering in real time
  - Created lively discussion and generally addressed all questions before speaker had finished
  - All chats captured and will heavily inform workshop report and coordination efforts

- Structure
  - Day 1 spent on familiarizing ourselves with workflows from across the UFS
    - Including Operational Requirements and JEDI
  - Day 2 spent on breakout sessions
    - What are the problems? What are possible solutions?

UFS Workflows Workshop: Goals

- Listening and Learning
  - Workshop was an opportunity to bring together a large cross-section of the community and learn from and about each other
  - Day 1 consisted of presentations on the different UFS applications
  - Some of the questions asked of presenters included:
    - What does your workflow do? (Does it include DA, pre- and post-processing, etc)
    - Who is the target audience? (Who uses the workflow and for what?)
    - How portable is it?
    - Does it include the build system?
    - How do users configure their experiment?
    - What workflow manager does it use? (Does it use one at all?)
    - Where do you see opportunities for collaboration?
  - Day 2 consisted of breakout groups
  - Focus on identifying issues in the existing workflows and how they could be improved
    - What are barriers to use? (Particularly to moving between applications?)
    - Where are there areas of duplication? (Are they necessary?)
  - NOT trying to find The Right Answer
    - Broad agreement that One Workflow to Rule Them All is probably not feasible
  - Focus instead on finding A Better Answer
    - How can we improve on what we have?

What are the Problems?

- **Too many workflows!**
  - Current state of affairs dubbed the 'workflow explosion' (h/t Kate Friedman)
    - Different workflows have arisen for many different reasons
    - As many different workflows as there are applications
  - Barrier to moving from one application to another
    - Users must often learn an entirely new workflow each time
      - Includes new scripts, configuration managers, workflow engines, directory structure, etc.
  - Significant duplication of developer effort
    - The same problems are getting solved multiple times in different ways
    - Innovations in one workflow often do not find their way into others

- **Recognized as an issue that should be solved**
  - Possibly an obvious point but an important one
  - If the multiplicity of workflows is not causing an issue, then nothing needs to be done
  - If there was no buy-in that this was a problem in need of solving, then nothing is likely to be done
  - General recognition of the need to work together, especially across organizations
    - Take operational requirements into account

Why Has This Happened? And What Can We Do?

- **Application needs**
  - Each application has specialized elements
  - Some diversity in workflows is necessary and appropriate
  - Creating a common set of tools that could then be used to construct multiple, diverse workflows had significant support
- **Lack of Communication**
  - Users and developers are scattered across multiple labs, computational platforms, software stacks, etc.
  - There has been no established channel for the various workflow development groups to talk to each other
  - Many key terms are overloaded, and mean entirely different things to different people
  - More direct involvement by NCO – no official presence due to pressure of existing duties
    - Implementation Standards are fairly opaque and not generally familiar to the broader community
  - Major point of emphasis in UFS R2O Workflow project
    - Initial outreach efforts include use of UFS Confluence, Slack channel invitation to participants, and ongoing webinar series
    - We need a common, agreed upon set of definitions
      - UFS glossary is a good start but not complete and should be updated
    - NCO needs to be involved early on and in all steps of the process
      - A comprehensive checklist of requirements from NCO would be a tremendous help

Why Has This Happened? And What Can We Do?

- **Lack of Mandate**
  - Coordination with other workflow developers is **not** a metric of success
    - Success for workflow development is defined as a workflow that works for your app
    - Everyone is busy and trying to succeed at the job in front of them
  - Workflow standards for UFS currently an ‘unresolved issue’.
    - Currently no official development standards
    - No governing body has been tasked with creating them
  - Role for the System Architecture and Infrastructure CCT? UFS R2O Cross-Cutting Infrastructure?
    - SA CCT is still spinning up, but seems to be the logical place to address this
      - CCI does not include funding
    - Writing workflow requirements from scratch and recreating all workflows around them is probably too heavy a lift
      - Can we identify applications and workflows with similar elements/needs and work first to reduce barriers and duplications there?
      - Multiple applications make use of similar tools (chgres, UPP) that could allow for common use

What Is Happening Now?

- UFS Workflows Slack channel has gotten significant traction already
  - https://ufsworkflows.slack.com
    - Any NOAA email address should allow you to join
  - FedRAMP approved
  - Developers from multiple applications regularly participating in discussions
- Definition of Terms
  - Workflow discussions include a number of overloaded terms
    - Including Workflow!
    - Discussions have led to updates to UFS Glossary
- Workflow Comparison
  - It is currently impossible to directly compare workflows across the UFS applications
    - Detailed information on specific workflows tends to be limited to a single individual within an application
    - Documentation varies wildly in both level of detail and format
    - Cannot know how workflows overlap without knowing what each one contains
  - Workflow Diagrams
    - Detail each workflow down to the level of the executable
    - Use a common notation (High-Level Petri Nets)

Lots of enthusiasm - now we need to keep it going!

Questions?