VERA C. RUBIN OBSERVATORY

Campaign Management

Kian-Tat Lim

NSF AURA U.S. DEPARTMENT OF ENERGY SLAC CHARLES AND LISA SIMONYI FUND FOR ARTS AND SCIENCES LSST CORPORATION

Goal

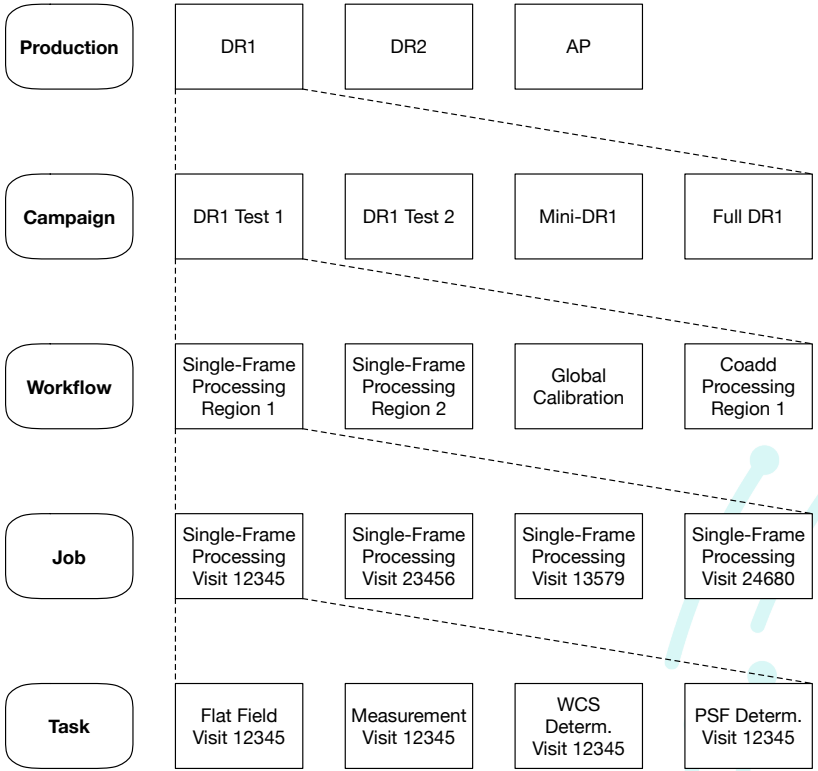
- Define the minimum technical requirements for a tool (or tools) to facilitate Campaign Management.

- Not trying to prescribe or even describe all the human processes that will be used.
- Not trying to lay out a schedule or assign tasks or define roles and responsibilities.
- But must have sufficient flexibility to encompass all of those processes.

Definitions

- Production: a large-scale class of processing
- Campaign: processing for a specific purpose; composed of multiple Workflows
- Workflow: a submission to a workflow management tool (e.g. PanDA, Pegasus, DAGman, or the Prompt Processing system) that results in the generation of Jobs; typically pre-processed by BPS from a QuantumGraph
- Job: a sequence of Tasks that runs on a single worker machine; typically a PipelineTask defined by YAML
- Task: an instance of the Python Task class, configured by a pex_config Config

Definitions



Key Points

- Humans start workflows within a campaign.
  - There is no need to “kill” a campaign; humans merely stop executing it.
  - Whatever degree of parallelism, ordering, prioritization, etc. is needed is implemented by humans.
- Humans define the inputs to workflows.
  - The inputs may be raw data or intermediate data products.
  - The default is generally to use all available relevant data.
  - But inclusion lists and exclusion lists may be applied.
  - These lists may be generated from any available information by any available method.
  - As long as the appropriate DataIds are recorded, the processing is reproducible. (The reasoning behind inclusion/exclusion should be recorded as well.)

DMTN-181

- Describes some use cases
- Defines the scope for a tool like an electronic lab notebook that records campaign definitions and links to other tools that provide access to execution status, outputs, logs, performance metrics, analysis capabilities, etc.
- Suggests that Jira be used as a basis for that tool
- Specifies a format and tools to handle inclusion/exclusion lists