Rubin
Observatory

\<Double Click Here\>

Control Box

Event Name | Location | Month DD - DD, YYY	→ DMLT VF2F | Virtual | February 23, 2021

Please fill out the table above, it will automatically populate the footers and titles.

NSF

U.S. DEPARTMENT OF ENERGY | Office of Science
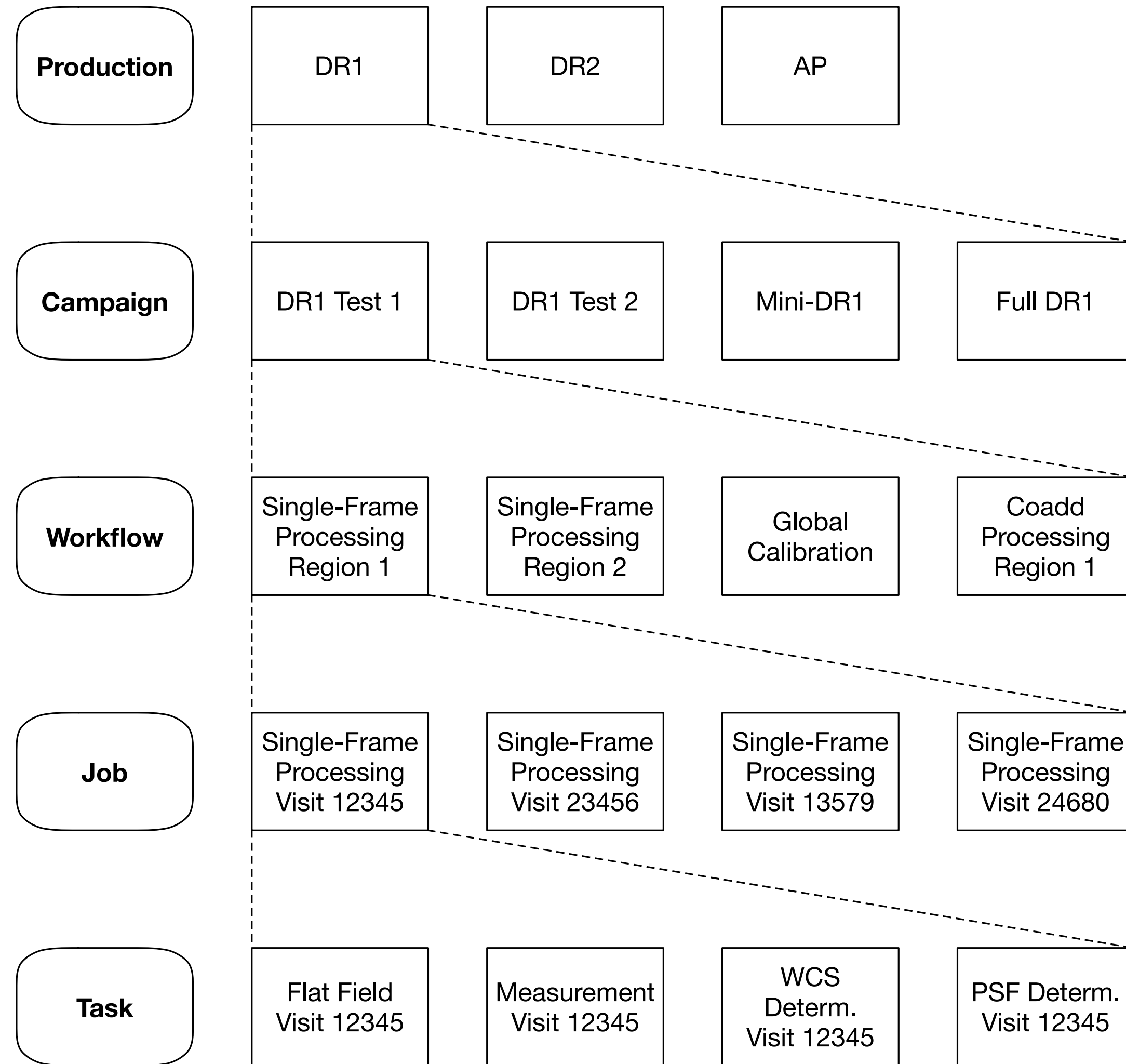
Goal

- Define the <u>minimum</u> 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.

U.S. DEPARTMENT OF **ENERGY** | Office of Science

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

U.S. DEPARTMENT OF **ENERGY** | Office of Science

Definitions

Production	DR1	DR2	AP	
Campaign	DR1 Test 1	DR1 Test 2	Mini-DR1	Full DR1
Workflow	Single-Frame Processing Region 1	Single-Frame Processing Region 2	Global Calibration	Coadd Processing Region 1
Job	Single-Frame Processing Visit 12345	Single-Frame Processing Visit 23456	Single-Frame Processing Visit 13579	Single-Frame Processing Visit 24680
Task	Flat Field Visit 12345	Measurement Visit 12345	WCS Determ. Visit 12345	PSF Determ. Visit 12345

U.S. DEPARTMENT OF ENERGY | Office of Science

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 <u>records</u> campaign definitions and <u>links</u> 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

U.S. DEPARTMENT OF **ENERGY** | Office of Science