Analyses: "functions" that map data / hypotheses to results.

$$\text{result} = f_{\text{analysis}}(\text{data}, \text{model})$$

For Reproducibility, and Re-use need to preserve analysis **separately.**

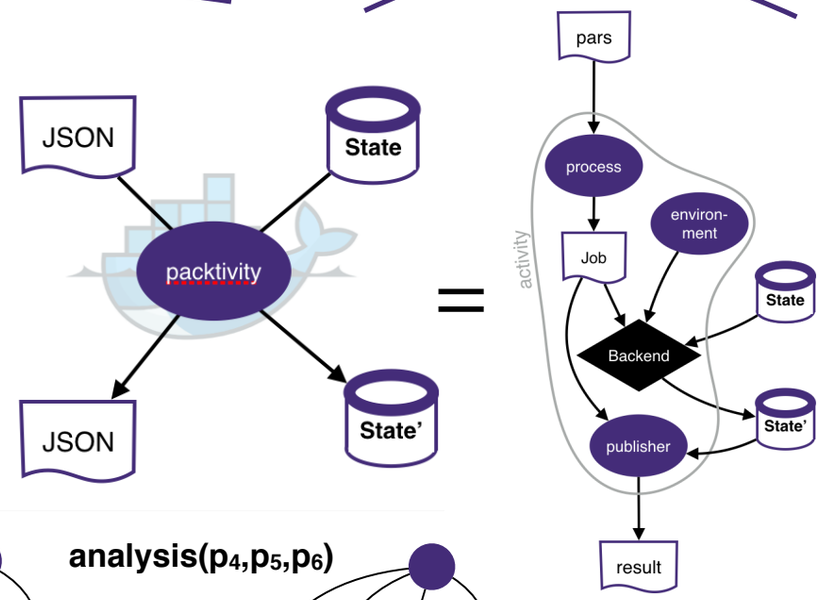
$$f_{\text{analysis}}(\,\cdot\,) \qquad f_{\text{analysis}}(\text{data}|\text{model})$$

1. Preserve Individual Processing Steps (packtivities)

Process:
parametrized task description

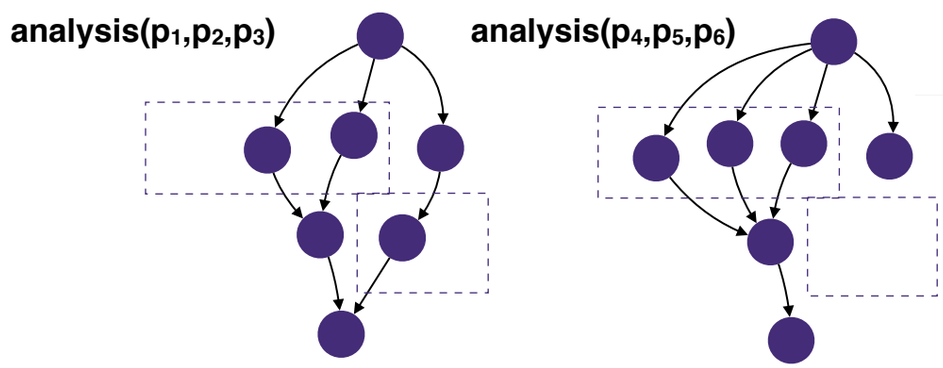
Result publishing:
What data fragments produced by the task are relevant?

Environment:
Where to run task. Containers (**Docker**) very promising. for max. host-independence.



2. Preserve Workflow logic between steps (stages)

Directed Acyclic Graphs (DAGs) suitable model, but may depend on parameters of analysis:



preserve logic how to build DAG instead of DAG,
i.e. series of "stages" adding nodes / edges

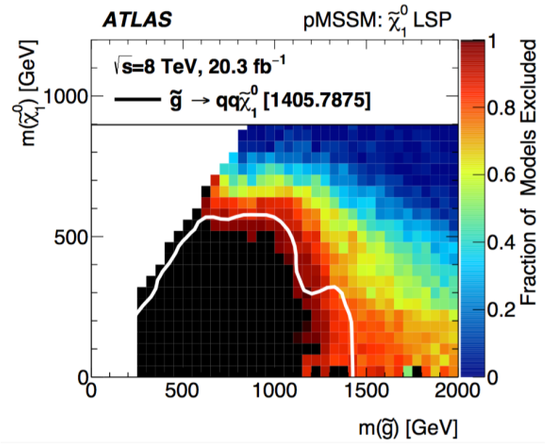
Implementation

definition of packtivities and stages as JSON docs according to schemas and storable in CERN Analysis Preservation (CAP)

CERN Analysis Preservation

yadage: engine for local or distributed parallelized execution of workflows

Integration with CAP and RECAST projects, for systematic reinterpretation.
Used in a number of ATLAS reinterpretation campaigns



NEW YORK UNIVERSITY