

# Use Case Proposal for HEP

Philip Bechtle

13.12.21

# TA4 Connections to Use-cases

*An old slide from August 20th 2020 which I found very interesting while preparing these slides:*

- ▶ Our usecases are almost all about
  - ▶ FAIR use of the data
  - ▶ Interaction with the data
  - ▶ Making the data “live”: Use the data and provide knowledge (derived data, code, ...) back into the community
  - ▶ Gaining new knowledge though combining data
- ▶ TA4 is about facilitating all of that through the science portal

# Useable Use Case Classes

For us, I think

- 1.1 Open data workflows (HEP)
- 1.9 analysis software preservation
- 2.1 HEP/HuK: Data processing
- 3.3 HEP: Processing of simulated data
- 4.1 Make sensitivity plots for new theoretical ideas or experiments directly on the PUNCH-SDP
- 4.2 Global analysis across our consortium

would be good starting points. It is important to consider all use case classes.

# General Considerations

- ▶ I think it is dangerous to **only** start with non-dynamic RPs as use cases, simply because we might miss a lot of details necessary for the dynamic part. And I think the dynamic part is essential for creating true progress.
- ▶ I think we have enough simple workflows readily available to achieve the above
- ▶ I think we need a well balanced set of about 3 starting use cases from Astro/HEP/..., where for ... I think lattice would be a good choice
- ▶ I think it is **very** important to include the other TAs, especially TA2, very early in the design process. It is easy (and good) for TA2 (and TA3) to use the TA2 (TA3) resources to improve the computing infrastructure for HEP/Astro/HuK/... in useful steps independently from the platform. It is **our** task to keep that all together.

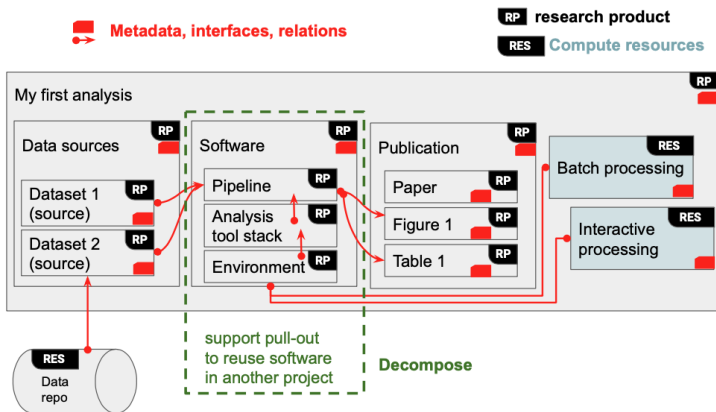
# A Staged Proposal

- ▶ Idea: Start with e.g. HEP/Astro/Lattice Use Case 1
- ▶ Define first prototype to tackle Use Case 1
- ▶ Discuss needs with other WP/TA's, start implementing first prototype for use case 1
- ▶ As soon as Use Case 1 is in the stage of implementation, start defining use case 2
- ▶ As soon as Use Case 1 enters the testing phase, use first experiences refining the planning for the prototype implementation of Use Case 2
- ▶ And so on

# (D)RP Components

Try to find use cases where “simple” examples cover all of the foreseen components:

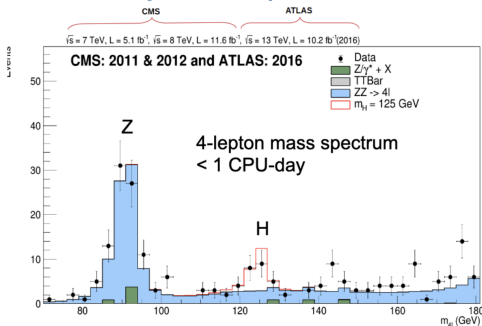
## Composite RPs: illustration



# Starting Point from WP3 for the HEP Use Case Proposal

- Convert available CMS, ATLAS and ZEUS data (others to follow) to common analysis format (WP3-3) (many CPU-months, make output available to users)
  - > analyze them with the same Root or Python script on your laptop
- (remote data access; Jupyter notebook possible, (not yet))

## Analysis from Open Data sets



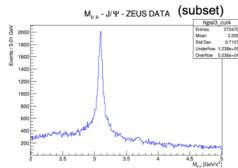
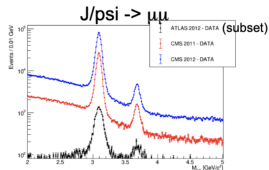
DESY Summer Student Program 2021

13

21.9.21

A. Geiser, DPOA meeting

DESY summer student projects 2021:  
A. Bal, A. Geiser, L. Olivi,  
R. Schwenzer, M. Velasco, Y. Yang



6

# HEP Proposal Stage 1

- ▶ operate Achims workflow as a “centrally run” or “background” **DRP** and save the outcome with its metadata as a set of **RP**s. That would be interesting for:
  - ▶ data sources (mostly CERN open data, but also DESY)
  - ▶ Software:
    - ▶ “simple” workflow
    - ▶ run common analysis format code from WP3
    - ▶ metadata management: both relevant for the content (definitions of units etc) and data provenance, etc from WP2
    - ▶ dependence on environment: can depend on tools from CVMFS, like root
- ▶ relevant as an example for “publishing” data as an RP
- ▶ already here we need TA2 for processing and storage

*This of course means that the work to **define** the (D)RP, set up the catalog, etc., still has to be done together/before/with examples like that. What is written above is not a replacement for defining the (D)RP.*



## HEP Proposal Stage 2

- ▶ implement a “simple” analysis based on the RPs from above. One example would be  $J/\psi \rightarrow \mu\mu$ , which probably can be found at all of these 4 experiments. That could be important for:
  - ▶ workflow now including additionally tools from TA3: fit tools
  - ▶ workflow can already include “automatic” detection of which tool runs on which input:
    - ▶ generic analysis framework for common analysis format for CERN Open data
    - ▶ output of this framework should comply with inputs to fit tools from TA3
    - ▶ all of the above should be captured by metadata describing interface between tools and data/simulation file formats (WP2,WP3,?)
- ▶ environment now includes TA3 products
- ▶ transition towards a real analysis
- ▶ relevant example for publishing a physics result from a single type of analysis from a common abstraction level, but starting from several *different* RPs (Open Data from different experiments)

*If you think that's challenging, then please don't forget that Achim's summer students found the Higgs boson in exactly these datasets.*

# HEP Proposal More Stages

- ▶ I think it is important to think ahead and not only design for the most elementary cases, exactly because we are aiming for functionality which goes beyond the existing
- ▶ In an email exchange, there are many more stages listed, but it will take some time to get so far, so I hope it is enough to start thinking about the first steps. I won't list them here in addition, but the general idea is:
- ▶ If we have such “escalating” use cases where the next step builds on the successful implementation of the step before and then adds one more element, it is both easy to
  - ▶ Prepare for more functionality to be added later on while designing the prototype for the most basic example
  - ▶ Actually proceed quickly from step to step because we don't have to start from scratch with each added complexity.

# More/Other Proposals?

- ▶ Of course we need such (or better) proposals also from other areas of our science
- ▶ We still need to go much more in detail through the selected Use Cases to really start the design work. This discussion is only about the starting point for the Use Cases, it does not replace looking at the starting points in detail and then get to work defining the (D)RP, the catalog, the user interface, the portal, the interfaces to TA2/TA3/, etc