Karlsruher Institut für Technologie

Scientific Computing Center

LMU MeerKLASS workflow

TA2 WP2 meeting

Benoit Roland, Manuel Giffels | 26. May 2025

Dockerfile is needed

- Necessary and sufficient description of the software
- Fundation for a FAIR approach
- Analysis setup available on any C4P resource via CVMFS
  - Workflow independent of the resource's architecture and OS
  - No need for manual deployment of the software
  - Possibility to run the analysis anywhere on C4P
- File transfer to be optimised to make full use of this approach
- **Software and setup used today to be known unambiguously**

Dockerfile is suited

- Software extensively tested during several months of dedicated campaigns
- Required software packages frozen to 30th May 2023 version
- **Components where developments still needed: independent of the container**
  - Retrieved from parallel filesystem by dedicated script at run time

Tracking back the original software - caveat 1

- Git repository used to retrieve main software - DDFacet - does not exist anymore
- Merged afterwards in a global project - SAOPICC
- **On the long term - which SAOPICC branch to be used?**

Tracking back the original software - caveat 2

- Dockerfile does not exist
- Singularity image has been build iteratively running a Debian Bullseye image in interactive session
- **Dockerfile to be build using the imperfect knowledge retrievable from the image**

Tracking back the original software - caveat 3

- Some packages installed but removed afterwards from the image
- Only needed during build phase
- For some of them, master branch was used
- **Difficult to track back the original master branch - as package not present in the image**

Tracking back the original software - caveat 4

- Main software packages depend on Git submodules
- Files containing information about submodules present but empty
- **Not possible to track back submodules versions**

Conclusion

- Ongoing work on LMU MeerKLASS workflow
- Difficult to obtain an unambiguous knowledge of the original software
  - Dockerfile missing
  - Versioning missing
  - Submodules provenance missing
- **Software Management highly needed for MeerKLASS**