



UNIVERSITÉ
DE GENÈVE

FACULTÉ DES SCIENCES
Département de physique
nucléaire et corpusculaire



The Calibration Pipeline of the Data Processing and Preservation System of CTAO

Mykhailo Dalchenko

Swiss CTA Day, Zurich, 14/12/2022

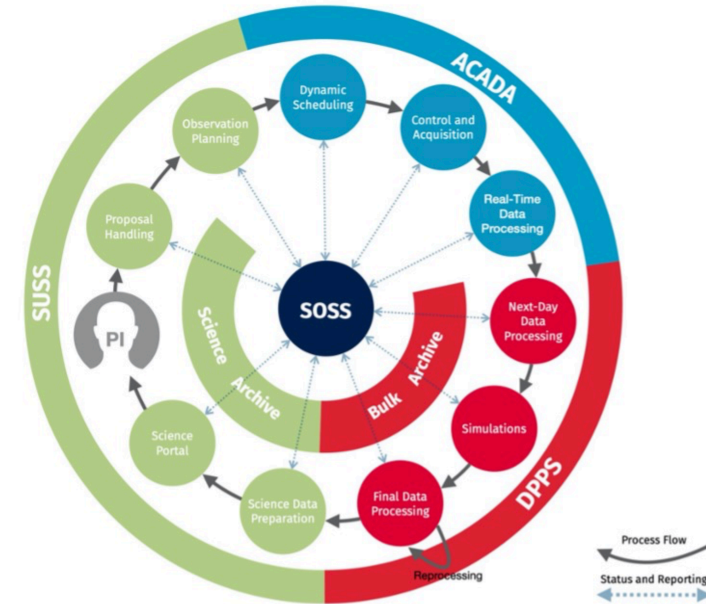
DPSS in a nutshell



DPSS objectives

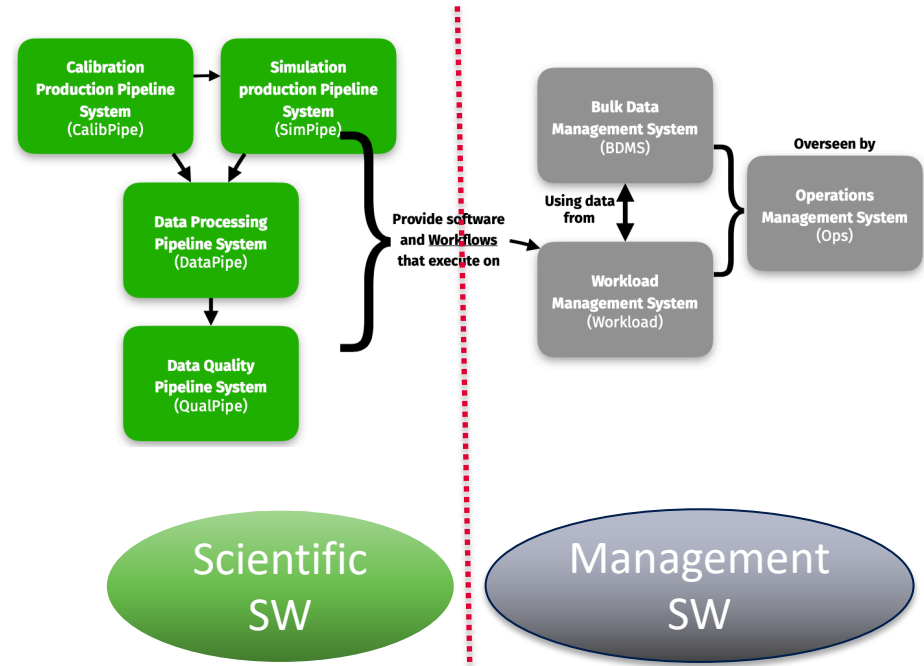
- Ensure preservation, processing and delivery to scientific user of low-level data
 - applies to both simulated and observed data
 - data products must be traceable and reproducible
- Provide monitoring and quality of the data products
 - with periodic reprocessing to ensure highest data quality
- Provide user interface to the DPSS sub-systems to the SOSS team
 - Including quality metrics and reports on the provided services

CTAO Science Data LifeCycle



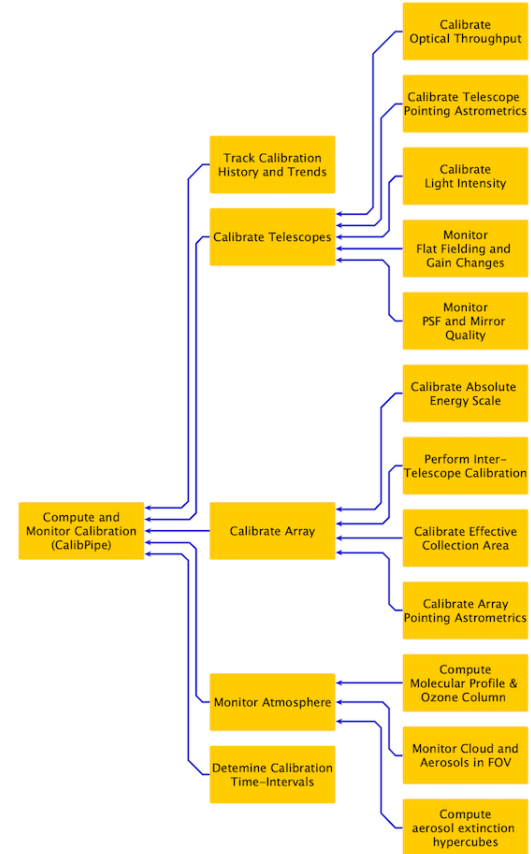
General DPPS architecture

- DPPS is a data-centered system
 - The approach to its SW architecture must be data-driven
 - The main task of DPPS is to perform data transformation
 - With a large side task of long-term data storage
 - The data transformations (and reduction/augmentation) are done step-by-step and referred as pipelines
 - Execution of a (combination of) pipeline is a workflow

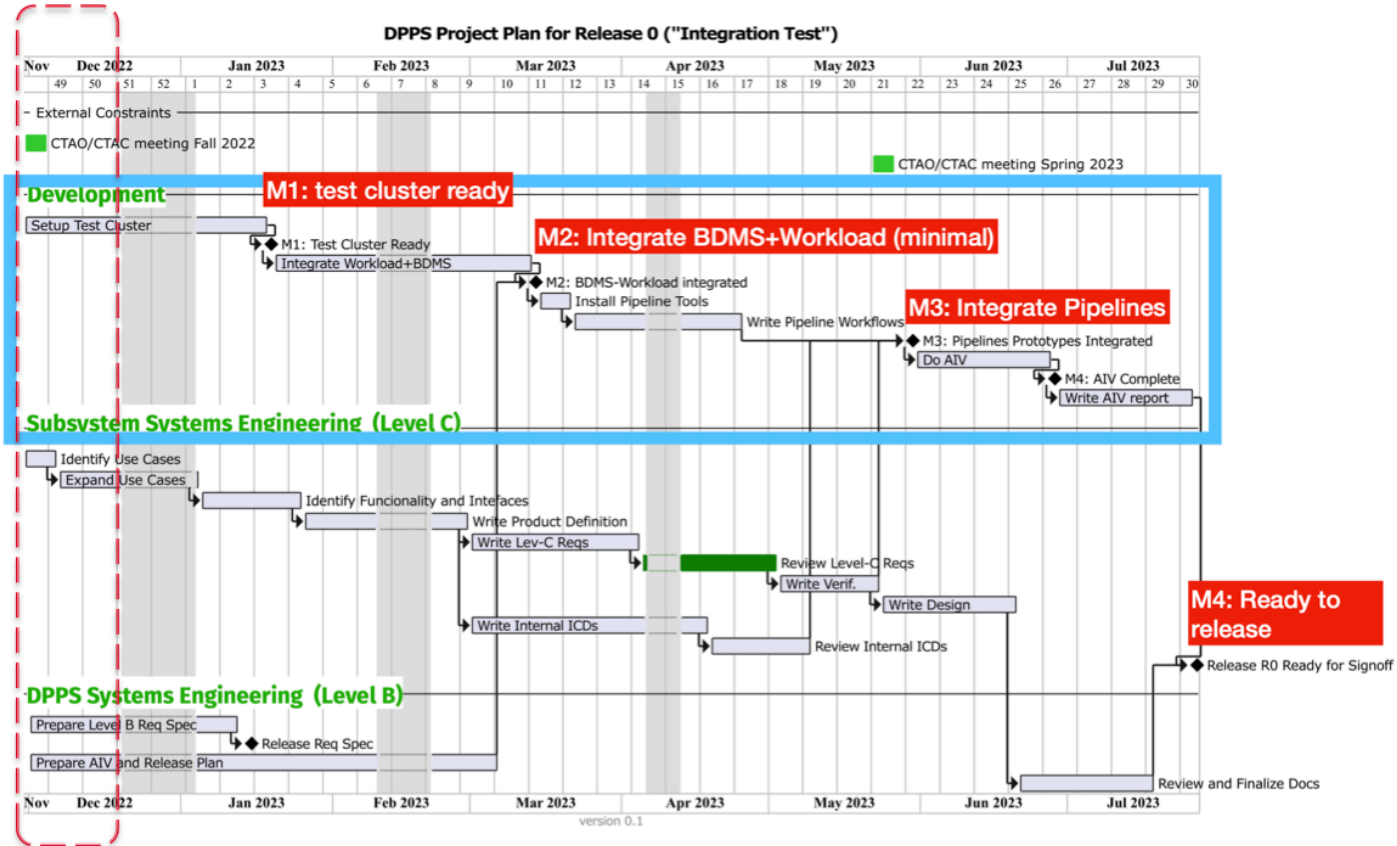


Calibration Pipeline Product Definition

- CalibPipe is a scientific subsystem of the DPPS, responsible for the computation of the calibration coefficients
- CalibPipe aims to achieve the best possible instrument accuracy and precision under all kinds of observation conditions
- To perform its functions, CalibPipe analyzes
 - internal to CTAO
 - physics data
 - monitoring data
 - service data
 - external to CTAO
 - meteorological data



DPPS Schedule towards Rel. 0





DPPS Rel.0 deliverables: Use Cases

- Generate a small set of Simulated R0 data
- Run a data processing workflow (R0-DL2) on the data produced above.
 - No (major) new functionality for DataPipe, SimPipe, Workload
 - BDMS needs minimal functionality
 - just store and retrieve, query
 - could run on a single DC
 - preservation (replication from one to another DC)
 - CalibPipe
 - molecular atmosphere calibration
 - molecular profile exchange with SimPipe (interface development)
 - DataPipe
 - Run basic data quality checks on the final output
 - minimal check that data is ok, not a full science metrics

Calibration Pipeline Status: organization



- An early start helps to advance the project
 - The first round of CTAO DPPS product requirements (Level-B) revision is done for the molecular atmosphere
 - Second round is planned for Jan. 2023
 - The first revision of the Use Case Registry is ready
 - The Calibration Pipeline Concept Document is drafted
 - Work in progress, to be completed by the end of the year
 - The interfaces description is ready
 - Still need to implement a formal document, work in progress
 - The testing and integration environment is ready
 - Initial CI, QA, and documentation pipelines in place
 - The configuration will be updated upon the availability of the DPPS-wide one

Calibration Pipeline Status: code

- Molecular profiles code is implemented
 - The main functionality is ready, small updates required on packaging and distribution/deployment
 - External SW (GDAS routines) expected to be re-packaged on our request
- The interface between the Calibration Pipeline and Simulation Pipeline is agreed upon (through the WMS)
 - Data Model is ready
 - implementation and integration testing are to be done
- The interface between WMS and CalibPipe is being investigated
 - Common to all pipelines task, lead by CalibPipe developers
 - Different workflow description languages are investigated
 - CWL, snakemake, in-house YAML dialects are being evaluated
- Muon Rings analysis implemented
 - Algorithms might require revision
 - Discussion with ACE experts ongoing

Calibration Pipeline Status: extras

- Database studies for Calibration DB
 - Comparison studies done
 - DB prototype is implemented and documented
- A new, auxiliary pointing monitoring method based on physics data stream is developed
 - The code is ready, the publication on new method is in review
 - Tests on the LST-1 data show good results
 - Ready to be integrated with the CalibPipe

Calibration Pipeline Status: to do

- On a project documentation side:
 - Finalize the Calibration Pipeline Concept Document
 - Detailed Level-C requirements decomposition
 - Formal interface documentation
- On a SW development side
 - Calibration Data Model specification and implementation
 - In relation to the use cases specified above
 - Database core is ready with corresponding serializers/deserializers
 - Adding new tables is straightforward
 - Interface to WMS implementation

CH planned and allocated contribution



Total CTAO-CH planned contributions:

Primary Interest task	2021	2022	2023	2024	2025	2026	2027	2028	Tot SERI	Tot done + IKC from Inst	Tot in table	Institute	CTAO CB
FTE (SERI)											SERI+Inst	SERI+Inst	
P07.2 LST System Engineer	0.4	1	1	1	1	0.6		0	5.00	2.25	5.25	12.50 DPNC	5.44
P07 SiPM R&D	1.0	1	1	1	1	0		0	5.00			5.00 DPNC - off CB	0
P07.2.1 LST + P07.4.1 SST TCS	1.2	1.2	1.2	1.2	1.2	0.26		0	6.26	2.1	0	8.36 Astro	5.89
P06.5.8 DPPS CalibPipe	1.0	4	4	3	1	0.5			13.50			13.50 DPNC	13.5
P06.5.2 DPPS Bulk Archive	0.0	0.5	0.5	0.5	0.5	0.5		0	2.50			2.50 Astro	5
P06.5.2 DPPS Bulk Archive	0.0	0.5	0.5	0.5	0.5	0.5		0	2.50			2.50 ETHZ	
P06.5.9 DPPS Data Quality	0.3	1	1.5	1.5	1	0.7			6.00			6.00 DPNC	6
P06.4.4 ACADA ADH	1.8	1.8	1.8	1.8	1.4	0			8.60	2.3		10.90 Astro	16.50
	1.0	1	1	1	0.9	0			4.90	0.7		5.60 ETHZ	
P06.11.4 Off Site Data Centre	0.3	1	1	1	1	0.7			5.00			5.00 EPFL	6
	0.0	0.2	0.2	0.2	0.2	0.2			1.00			1.00 CSCS/EPFL	

DPPS CalibPipe allocated:

2021: 1 FTE

2022: 3 FTE

2023:

2024:

2025:

2026:

Current team members: Mykhailo Dalchenko (team lead) Georgios Voutsinas (lead dev), Antonio Di Pilato (dev)

Retired team members: Gregoire Uhlrich (dev) Gabriel Emery (dev)



Summary and Outlook

- DPPS Project is getting up to speed
 - Decision on the content of the DPPS Rel.0 has been made
 - Detailed planning is provided
- Calibration Pipeline Project has successfully ramped up
 - Basis of the PM documentation is available
 - Calibration Pipeline code is in a good shape
 - Have all the main blocks for Rel.0 already and some extras
- FTE accounting is in accordance to the CH contribution planning



BACKUP

Time spent report details, CalibPipe

- 2021:
 - Project coordination: 0.25 FTE
 - Project design: 1.25 FTE
- 2022:
 - Project coordination: 0.25 FTE
 - DB studies: 0.75 FTE
 - Project management documentation: 1.0 FTE
 - SW development: 1.0 FTE

Team composition:
Mykhailo Dalchenko
(team leader)
Georgios Voutsinas
(lead developer)
Gregoire Uhlich
(developer)
Gabriel Emery
(developer)
Antonio Di Pilato
(developer)