Swiss Contribution to the Data Processing and Preservation System (DPPS)

Mykhailo Dalchenko

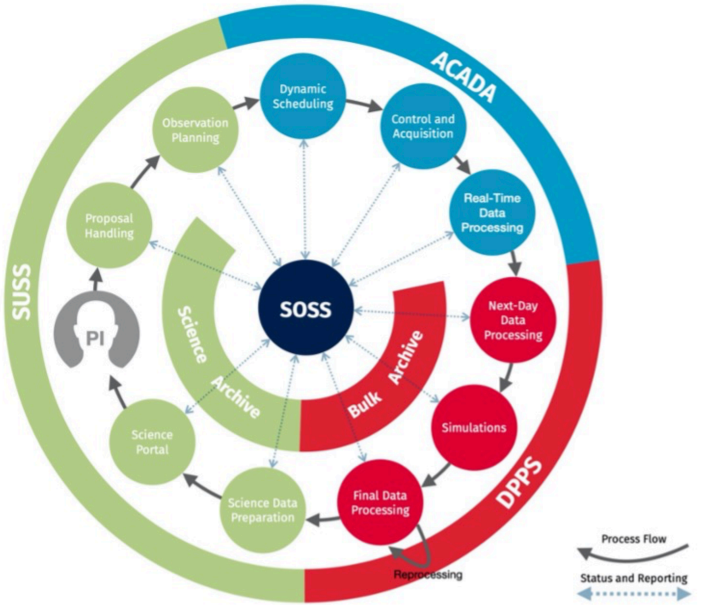
Swiss CTA Day, EPFL, 12/01/2022

DPPS in a nutshell

DPPS objectives

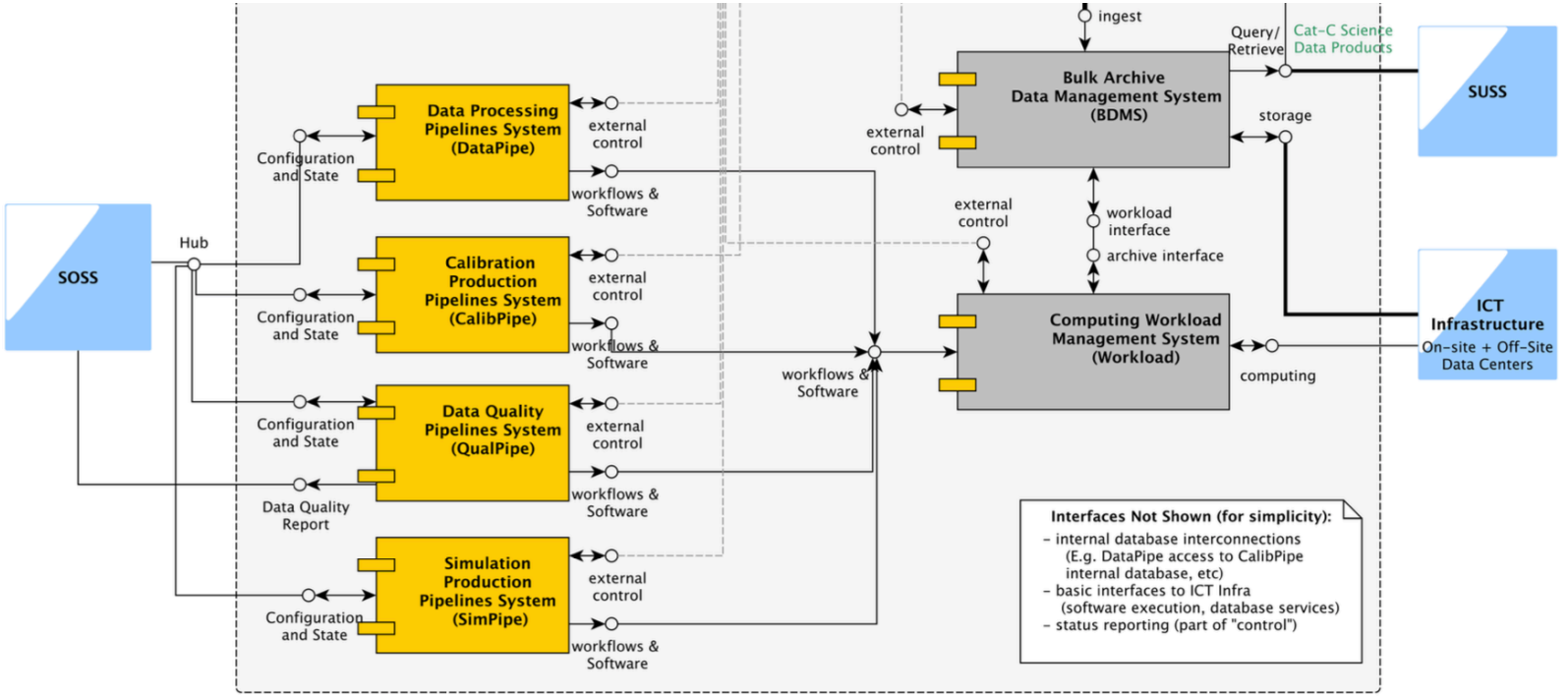
- Ensure preservation, processing and delivery to scientific user of low-level data
  - applies to both simulated and observed data
  - data products must be traceable and reproducible
- Provide monitoring and quality of the data products
  - with periodic reprocessing to ensure highest data quality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

CTAO Science Data LifeCycle



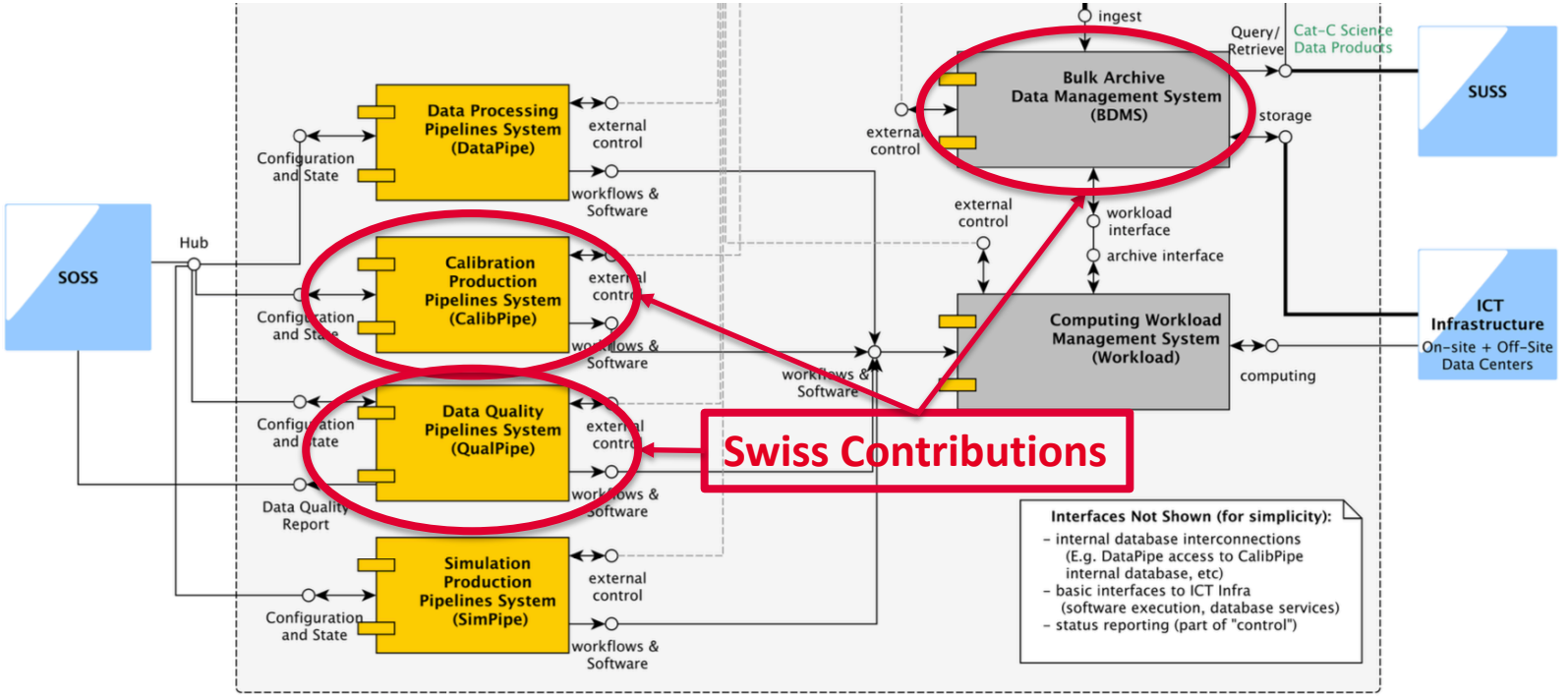
DPPS decomposition

Entire project is estimated to 105.5 FTEy



DPPS decomposition

Entire project is estimated to 105.5 FTEy



Swiss contribution to DPPS: BDMS

- Manpower share
  - 15 FTEy total estimate, 33% of FTEy planned to come from CH
- Past activities in CH
  - INTEGRAL archive (first OAIS archive)
  - GAMAS: prototype OAIS distributed archive for CTA (90% of its functionalities are covered by CERN's Rucio)
- Potential activities
  - If Rucio is chosen and its installation managed by the Data Centers
    - we could produce the remaining bulk archive software
  - If CTAO selects the w3browse or eXamine interfaces
    - we could customize it for CTA (then it can also interface to the science archive)

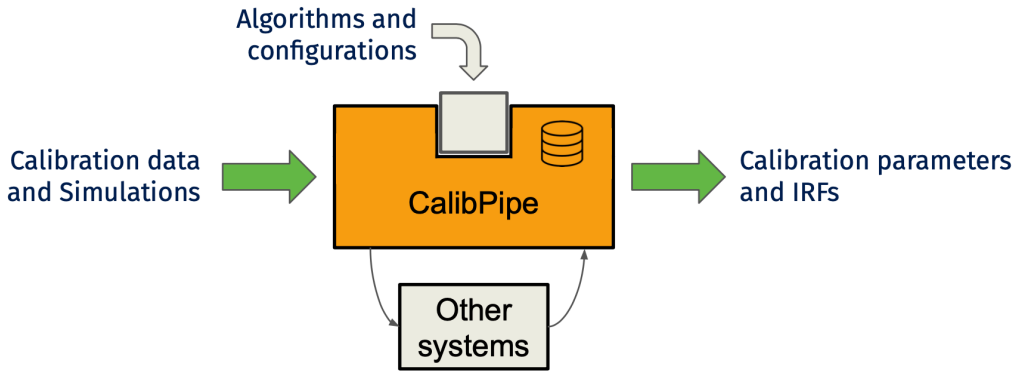
Swiss contribution to DPPS: Calib and Qual Pipes

- Manpower share
  - Calibration Pipeline
    - 20 FTEy total estimate, at least 67% of FTEy planned to come from CH
  - Quality Pipeline
    - 14 FTEy total estimate, 43% of FTEy planned to come from CH
- Past activities in CH
  - Extensive experience with the SST-1M telescope prototype RnD
  - Operational experience with the LST-1 telescope
- Potential activities
  - Calibration Pipeline: project has started, details in the next slides
  - Quality Pipeline: under discussion with CTAC/DPPS management

Calibration Pipeline deliverables

- Different types of calibration:
  - Event-level
    - Camera, telescope and array calibration
    - Depends on real calibration data (from ACADA)
  - IRF-level
    - PSF, BG model, energy dispersion matrix, effective area
    - Depends on simulated data (Simulation Pipeline)
- Different DL3 data productions in DPPS:
  - Category B / Cat-B (next day)
  - Category C / Cat-C (final, up to a week)

Calibration Pipeline deliverables



- Input data is provided
- Algorithms are (mostly) provided
- We have to design (and write) a package which
  - Can use different predefined algorithms in different configurations
  - Interacts appropriately with the other (sub-)systems (not only DPPS)
  - Obtain calibration products and store them in the CalibPipe DB

Calibration Pipeline project ramp up

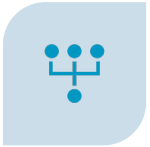
- CH (represented by UniGE/DPNC group) is a major contributor
  - Provides at least 13.5 FTEy
- CTA management appointed UniGE/DPNC representative as a coordinator of the DPPS Calibration Pipeline subsystem
- Decided on initial phase of the project
  - Preparation of the management documentation
    - Use Case Registry
    - Detailed Requirements Specification
    - Calibration Pipeline Concept Document
    - Calibration Pipeline Work Package Management Plan
- Negotiating the contributions of other IKC partners (FR, ESP)

Calibration Pipeline project progress

- Management documentation:
  - Use Case Registry
    - basis for project technical description
  - Detailed Requirements Specification
  - Calibration Pipeline Concept Document
  - Calibration Pipeline Work Package Management Plan

STUDY AVAILABLE DOCUMENTATION ON DPPS-WIDE USE CASES

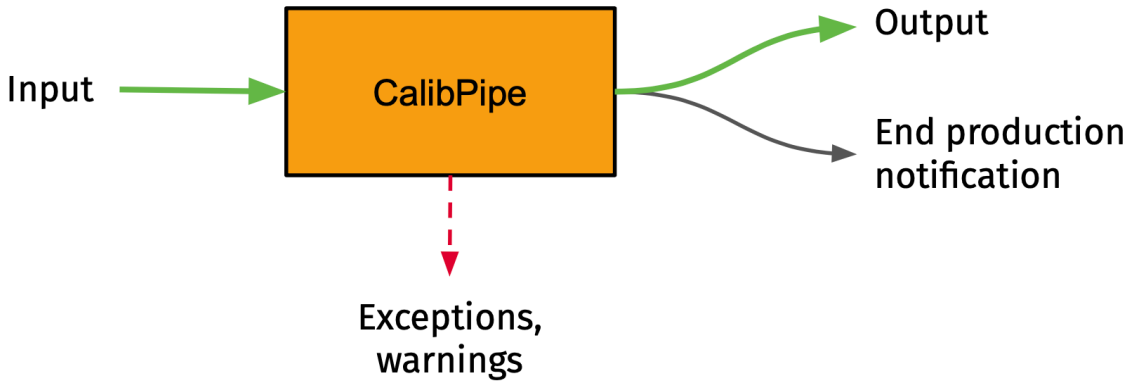
IDENTIFY THE ACTORS LIST



PREPARE EXAMPLES OF USE CASES AND SEEK FEEDBACK FROM DPPS COORDINATORS

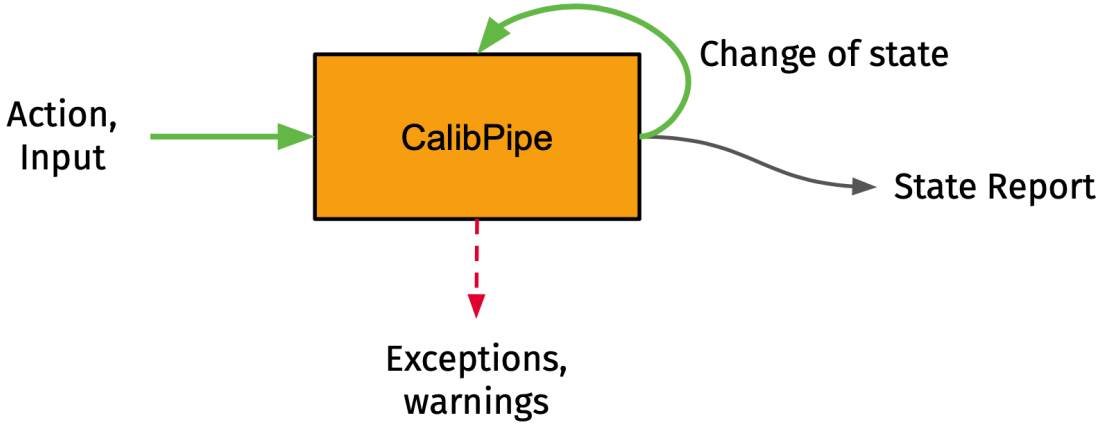
PREPARE REQUESTS FOR DPPS SUPPORTING DOCUMENTATION

Generic Use Case types: production



- Must determine:
  - Input / Output pairs
  - Context of use (frequency, pre- and post-conditions)
  - Actors (input, output **and** reports)
  - The architecture of procedure in Calibration Pipeline

Generic Use Case types: configuration



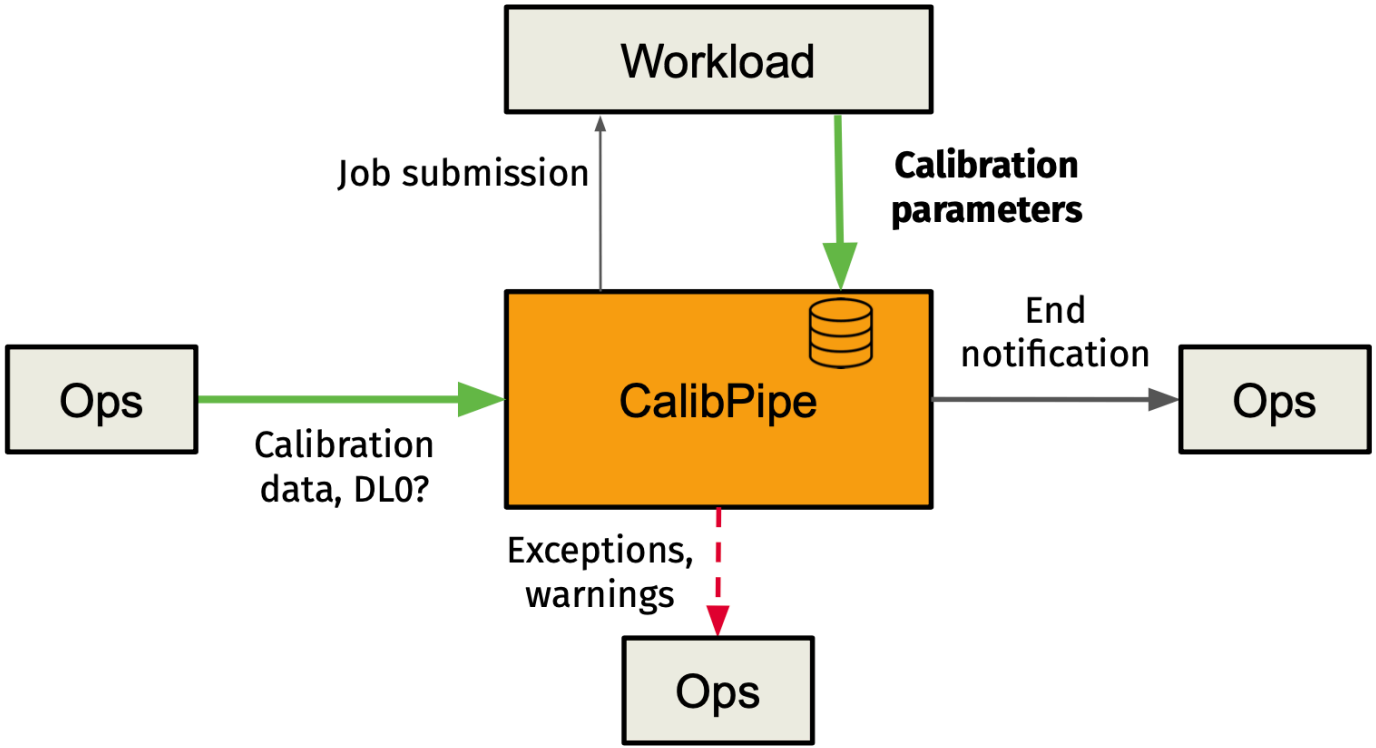
- Must determine:
  - Action / State pairs
  - Context of use (frequency, pre- and post-conditions)
  - Actors (input, output **and** reports)
  - The architecture of procedure in Calibration Pipeline

Summary

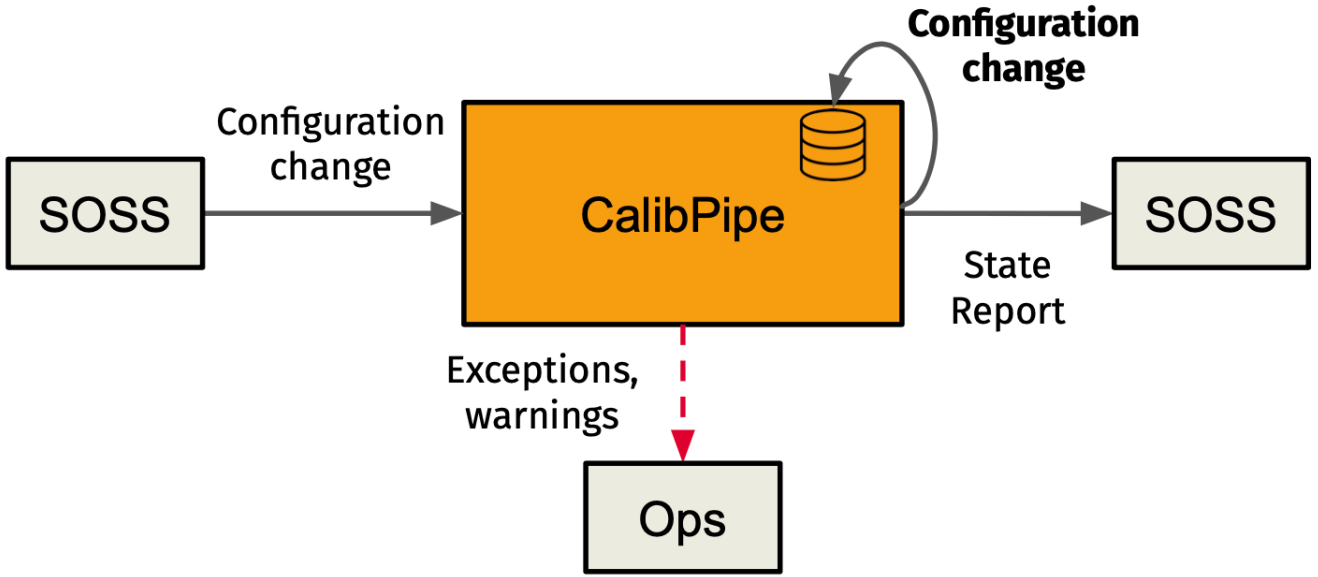
- Calibration Pipeline project is steadily ramping up
  - Manpower share is still under question
    - For the moment we are well stuffed given the delays in other DPPS subsystems
  - Tentative planning:
    - Produce first version of Use Case Registry by the end of February
    - Produce (and present) first drafts of all management documentation by next CTA General Meeting (May 2022)
- BDMS and Quality Pipeline
  - Ready to start, IKC team formation ongoing

BACKUP

Example: Event-level calibration



Example: Configuration change



Implemented Use Case drafts

- Specific calibration use cases:
  - FRAM down (missing data)
  - Relative optical throughput from CTC
  - Atmospheric calibration from CTC
- Generic calibration use cases:
  - Real data calibration for Cat-B data
  - IRF production for Cat-B data
  - Reprocessing of calibration data (Cat-C)
- Configuration change request

Scenarios for FRAM down

Main Scenario

Actor Action	System Response
1. SOSS is informed that FRAM data are missing.	
2. SOSS calls the alternative procedure for atmospheric extinction / absorption calculation.	3. The corresponding algorithm determines the atmospheric extinction / absorption and the associated uncertainties.
4. SOSS calls qualpipe to assess quality of data.	5. Qualpipe decides that the quality of the data satisfies the requirements.
6. DL1 data are produced. Information is included in report to TOSS. SOSS continues business as usual to produce DL2 data. Cat-B data will be produced.	

Alternative Scenarios:

A1

Actor Action	System Response
1. SOSS is informed that FRAM data are missing.	
2. SS calls the alternative procedure for atmospheric extinction / absorption calculation.	3. The corresponding algorithm determines the atmospheric extinction / absorption and the associated uncertainties.