LST Workflow for Observation Data Processing

Daniel Morcuende (IAA-CSIC) for the LST analysis team
DataPipe f2f meeting | October 2023

Packages

- LST EventSource plugin: **ctapipe_io_lst**
- LST-1 analysis library: **lstchain**
- Wrappers:
    - Pipeline for on-site & off-line processing of **observed data**: lstosa
    - Pipeline for **MC processing, RF models**: lstMCpipe

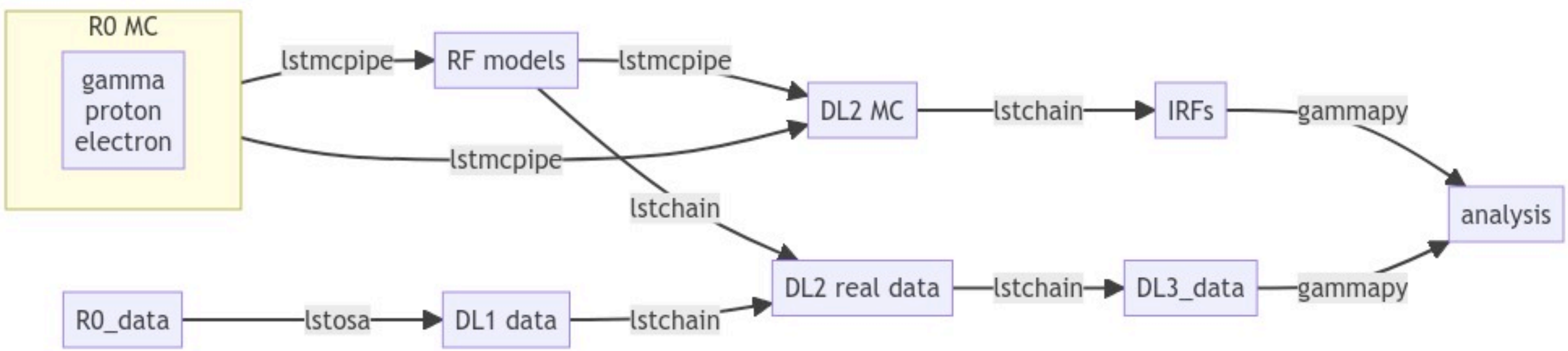
ctapipe_io_lst

LST EventSource plugin (https://github.com/cta-observatory/ctapipe_io_lst)

- Read in LST zfits files
  - Currently compatible with EventBuilder (EVB) v5 + ACADA Data Handler
  - Almost ready to digest data taken with EVB v6
- Apply low-level DRS4 corrections
- Add pointing information to the data stream
- Timing and trigger information correction

lstchain

Low-level analysis library to DL3 (https://github.com/cta-observatory/cta-lstchain)



LST-1 analysis workflow

lstchain

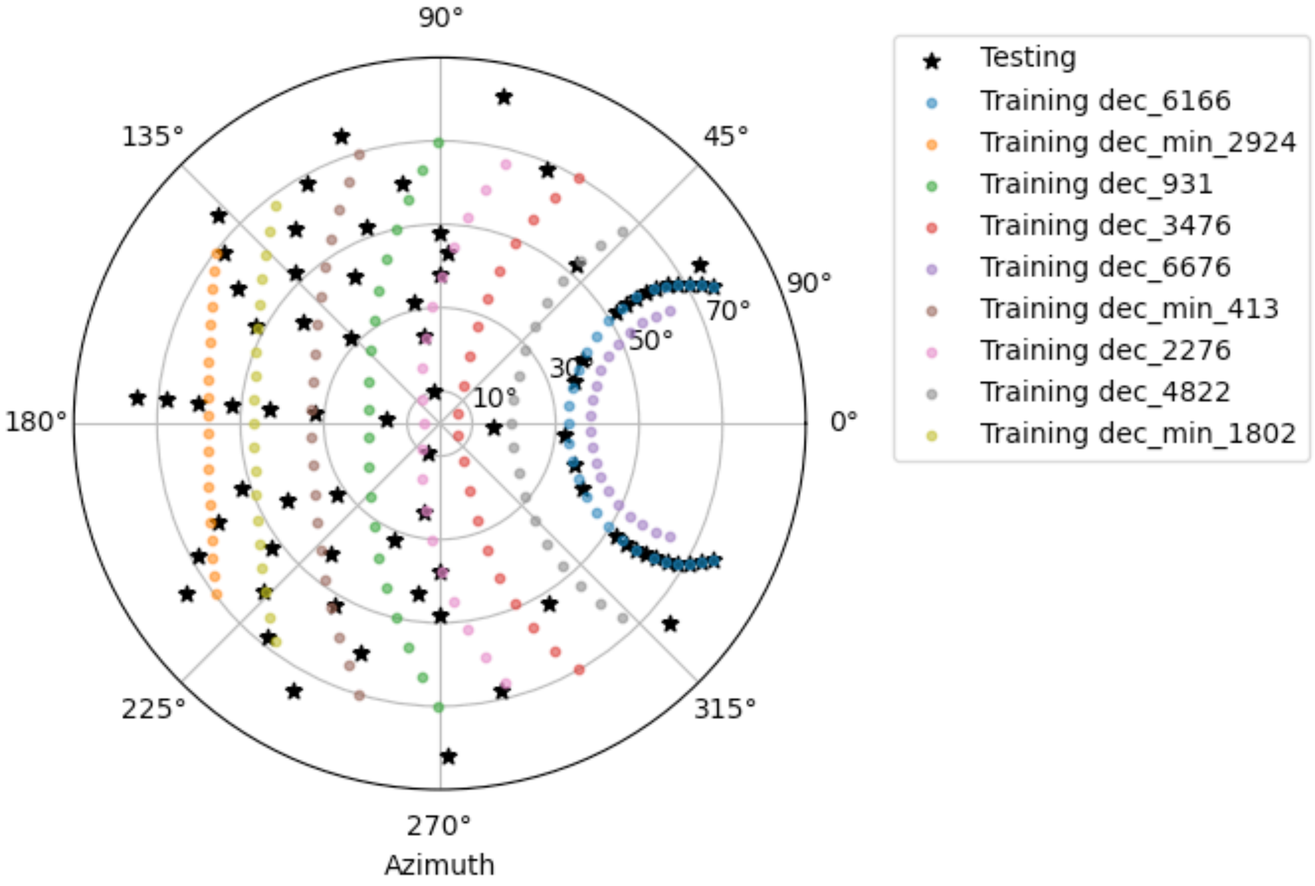
General aspects

- Does not use *ctapipe-process* nor any other ctapipe tool directly yet
    - Custom tools and scripts to produce calibration coefficients and analysis up to DL3 (including IRF creation)
- Utilities:
    - Tune NSB in DL1 MC to produce more realistic RF models
    - Create observations summary
    - Data quality selection

lstchain

MC general aspects

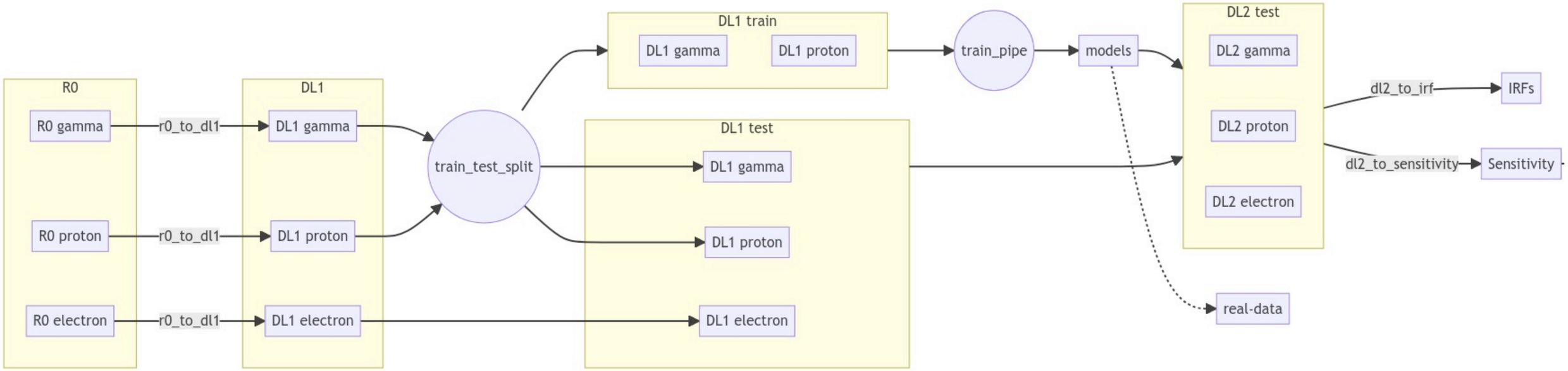
- RF models trained along declination lines of the sources
- IRF calculated in MC pointing nodes:
  - For each data run, choose the IRF calculated in the closest pointing node of the test MC grid
  - IRF interpolation



lstMCpipe

On-site MC processing (https://github.com/cta-observatory/lstmcpipe)

Orchestrate the processing of simtel files to train RF models, produce IRFs, dealing with all-sky MC grid



lstosa

On-site observed data processing (https://github.com/cta-observatory/lstosa)

Orchestrates the LST-1 analysis stages by using the LST-1 analysis library, **lstchain**

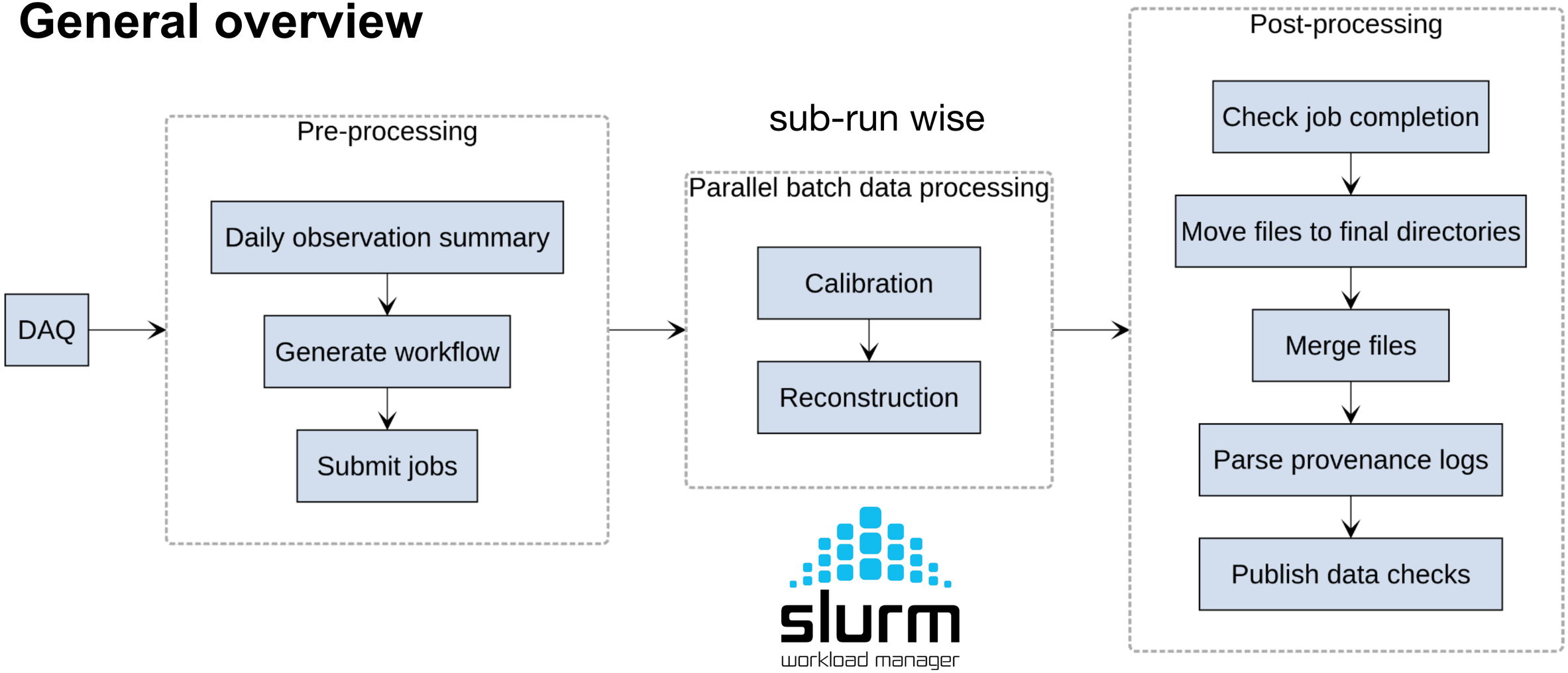
Aim:

- **Standardize** processing of observed data (minimize user-related errors)
- **Workload manager system** integration (optimize usage of the on-site data center)
- **Automatic** next-day low-level analysis and DL1 data check within hours
- **Large-scale data reprocessings**
- **Provenance tracking**

Based on MAGIC onsite analysis pipeline

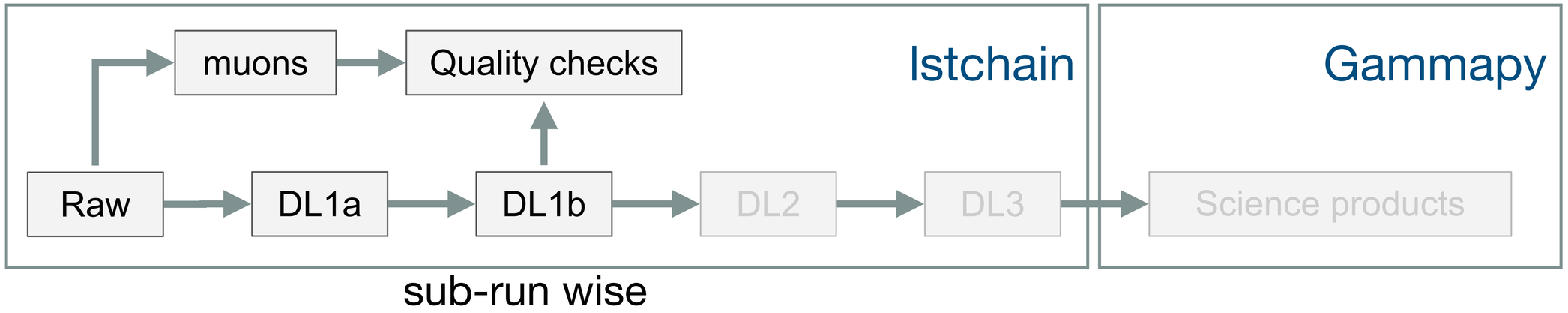
Istosa

General overview



Istosa

- Producing up to DL1b + Quality checks (see Abelardo's presentation)



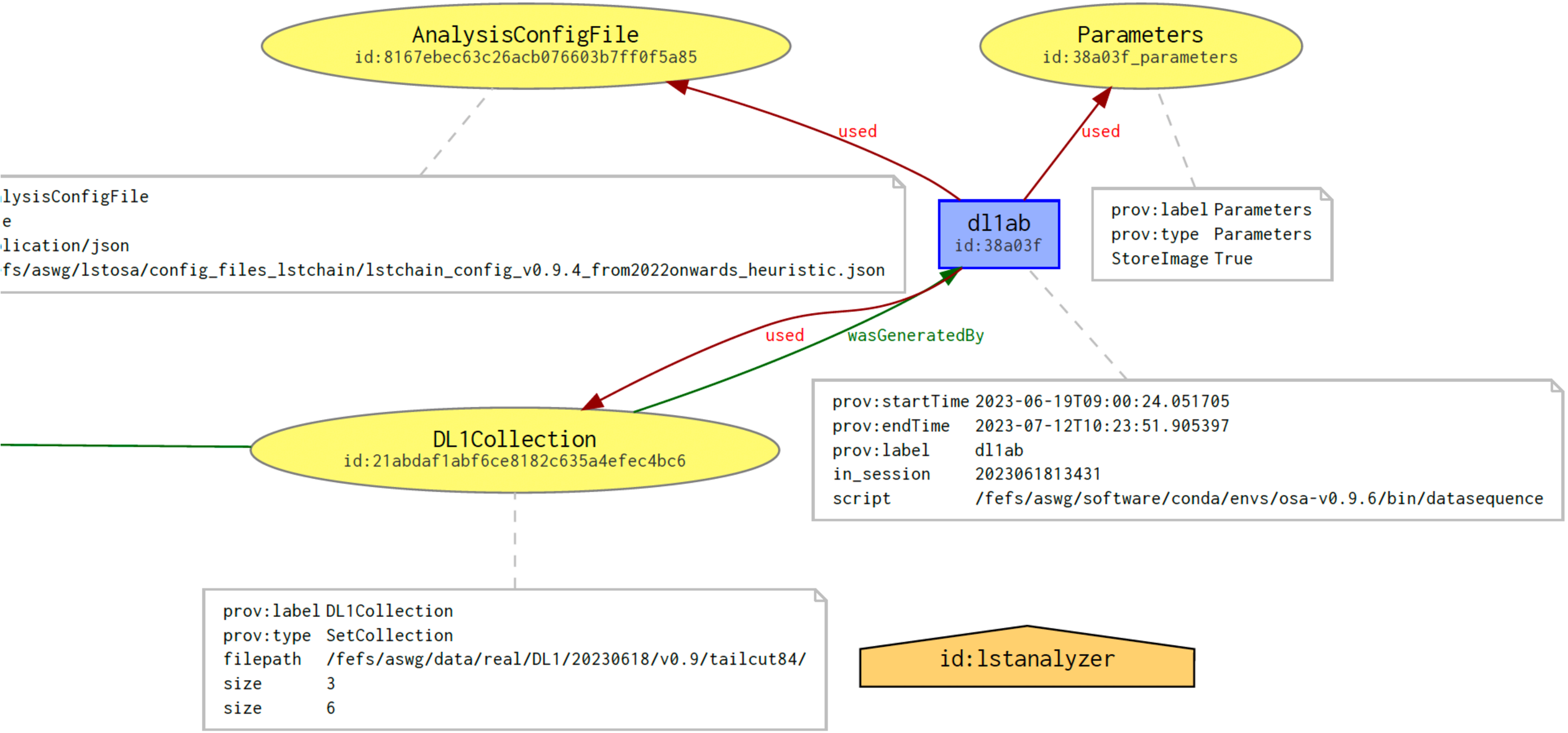
- DL1b to DL2 step was being done at the beginning. However, centralized DL2 production has been stalled since the all-sky grid MC scheme was adopted
- Skeleton for DL2 to DL3 stage present (preliminary fast analyses)
- Higher-level next-day analysis is also planned

Post DL1b
done by users

lstosa

- **Gain selection** and **data volume reduction** steps to **free disk space**.
  - Based on an external tool
  - Currently performed off-line
  - ~15% of the original raw files
  - Eventually, it will be done by EVB so that it provides directly DL0

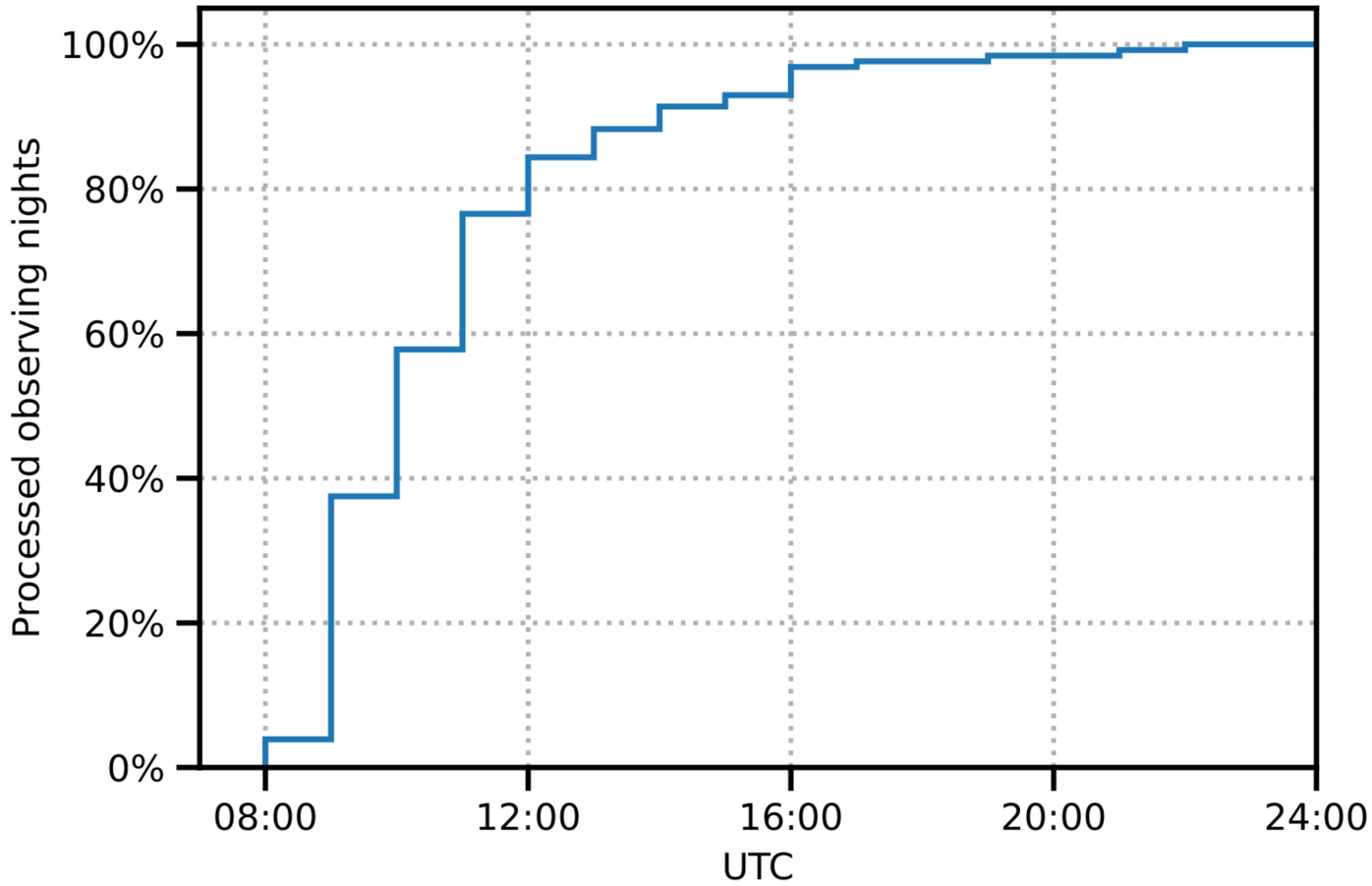
Graph provenance example



Istosa

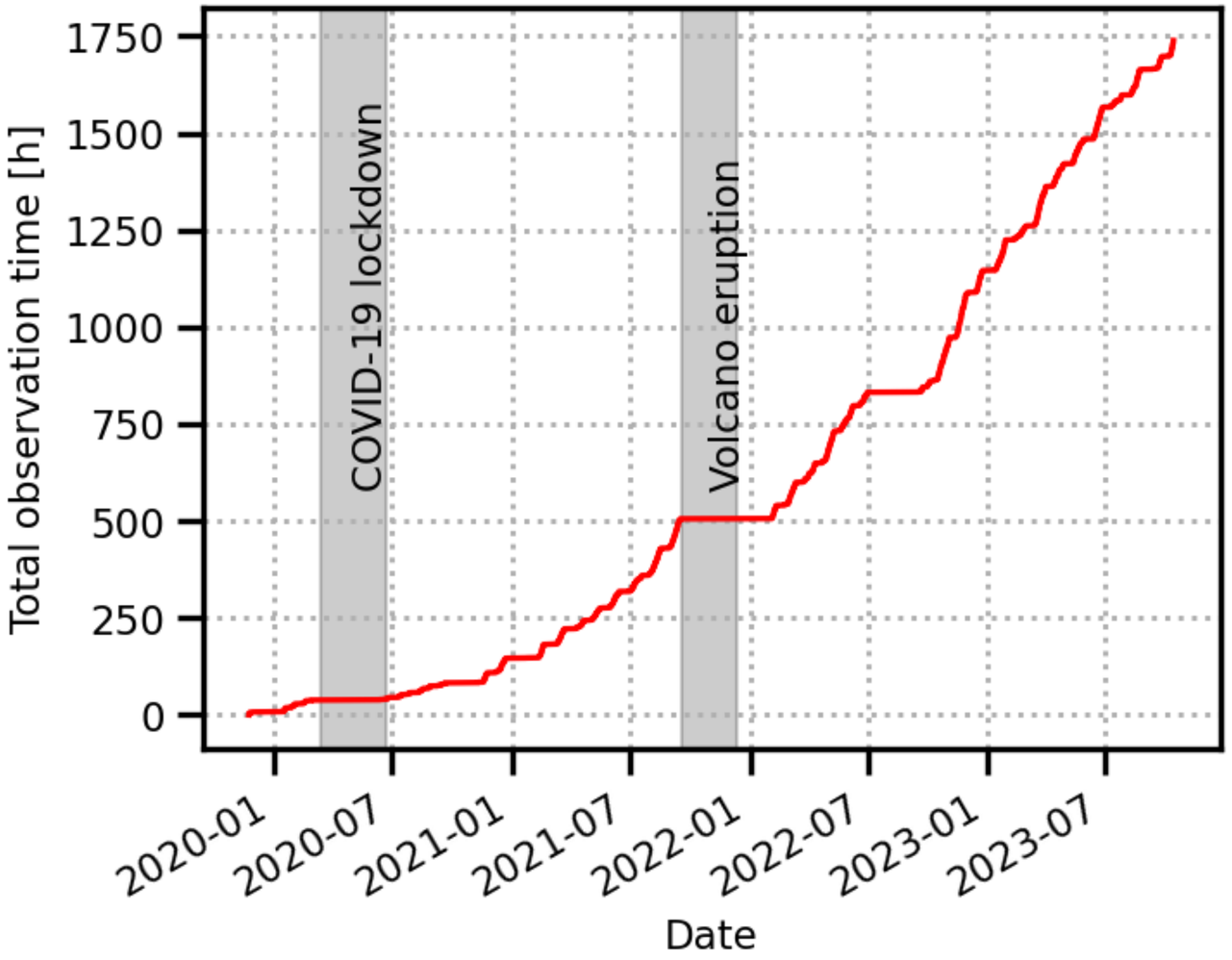
Next-day data processing

- Runs automatically the next day after data-taking
- Typically no human intervention is needed
- Data processing within a few hours: 80% of nights are processed the next day by 12 UTC



Istosa

Large scale reprocessing



- Done in the on-site data center
  - ~ with **each major lstchain release** (3-4 times so far)
- Currently 1700 h of observed data **~3 weeks** of CPU time (assuming 1000 simultaneous jobs)
  - Additional time when considering human supervision/intervention

lstosa

Important aspects in our experience

- Proper workflow management system (Snakemake, CWL, etc.)
    - Custom implementation for workflow orchestration and batch system integration
- Metadata: Integration with Telescope Control Unit (TCU) database
- Error handling (retry on known errors)
- Database for analysis book-keeping, traceability, provenance
- Job status monitoring