



Arch S22A Status & Plans

Kian-Tat Lim and Tim Jenness



U.S. DEPARTMENT OF
ENERGY

SLAC

CHARLES AND LISA SIMONYI FUND
••• FOR ARTS AND SCIENCES •••



Status for S22A

Requirements and Design

- Image service and regeneration — helped define target for JimB and RussA
- Prompt Processing — delivered prototype and design document
- Campaign management — discussions with BrianY group
- Consolidated Database — started work
- Updated DMTN-092 for pipeline feedback to Summit systems
- Reissued DMTN-135 sizing and costs; major correction to storage sizing

Oversight and Release

- v23 release — many RCs, 23.0.1, 23.0.2, instability of infrastructure
- Build engineering
 - `lsstinstall` — beta test going well
 - `rubin-env` 1.0 — even 2.0.0 now released
 - Default branch change — successful
 - Work with build engineer — defining tasks for MatthiasW
 - `mamba` in `newinstall` and `lsstsw/deploy`
 - Publish containers to GAR
 - `eups` versioning fix (x2), disk space monitoring, certificate authorities, indirect dependencies vanished, excessive disk usage, doc build fixes

Investigation and Coding — Middleware

- Retire Gen2 — official start to removals
- pipetask Python API — not yet
- Provenance — not yet
- Prompt Processing and Campaign Management needs — not yet
- Significant speedups
- Reorganization of code to simplify usage by other projects
- Loggers are always prefixed with "lsst"
- New commands for `butler` and `bps` to improve usability
- Pipelines can now be run in a “mock” mode for testing middleware.
- Many small improvements were made to middleware to support DP0.2.

Investigation and Coding — Other

- Prototype "Consolidated Database" — not yet
- OCPS return values and error handling — not yet
- DMTN-143 image capture — defined initial implementation, working on later ones

Pre-Ops

- Integration and standing up USDF — waiting on equipment, helping define next purchases, workshop in Jan
- Campaign Management — occasional consulting and reviews
- Security — ongoing design updates and discussions about OGA
- Hybrid model — design and spreadsheet updated
- DMTN-213 and expansion of DMTN-198 for Rucio/Butler/PanDA integration

Plans for S22B

Requirements and Design

- Consolidated Database design document
- Produce drafts for DMTN-204 (Dataset annotations) and DMTN-205 (provenance)

Oversight and Release

- Build engineering
 - lsstinstall release and incorporation into lsstsw
 - ~~rubin-env: Python 3.10 and rubin-env-extra~~
 - Keep working with MatthiasW (changelog and stack containers are first tasks)

Investigation and Coding — Middleware

- Release documentation on how to write pipeline tasks
- Possible schema migration with new abilities to support summit observing
- Provenance
- Dataset annotations (if part of butler or butler adjacent — DMTN-204)
- Resolve handling of metrics (is butler involved once file is written?)
- Graph-backed registry for batch production
- Support Rucio work
- Support prompt processing
- More flexible “runner” for PanDA
- pipetask Python API
- ~~Butler client/server~~

Investigation and Coding – Other

- Productionize (but not optimize) Prompt Processing
 - Complete middleware interface
 - Build Kafka and MinIO into prototype
 - Move to SLAC K8s if available
- Consolidated Database prototype
 - Per-visit EFD summary
- ~~OCPS~~
 - ~~Running SFM at the Summit~~
 - ~~Deploy at Base Commissioning Cluster~~
 - ~~Use BPS instead of pipetask?~~
- DMTN-143
 - Raw image upload service

Pre-Ops

- Integration and standing up USDF
- Campaign Management
- Security
- Hybrid model
- Rucio/Butler/PanDA integration