

DATA FACILITY NEAR-TERM PLANS

M. Gelman



Large Synoptic Survey Telescope

Phase 0: HSC reprocessing with legacy framework (ctrl_pool)

- Instantiates service “shell” for operational process development, at-scale production of large datasets for pipeline development support, also gather metrics (e.g., execution time).
- Running RC to verify readiness; expect full processing in next cycle.

Phase 1: Integrate job handling elements of production framework.

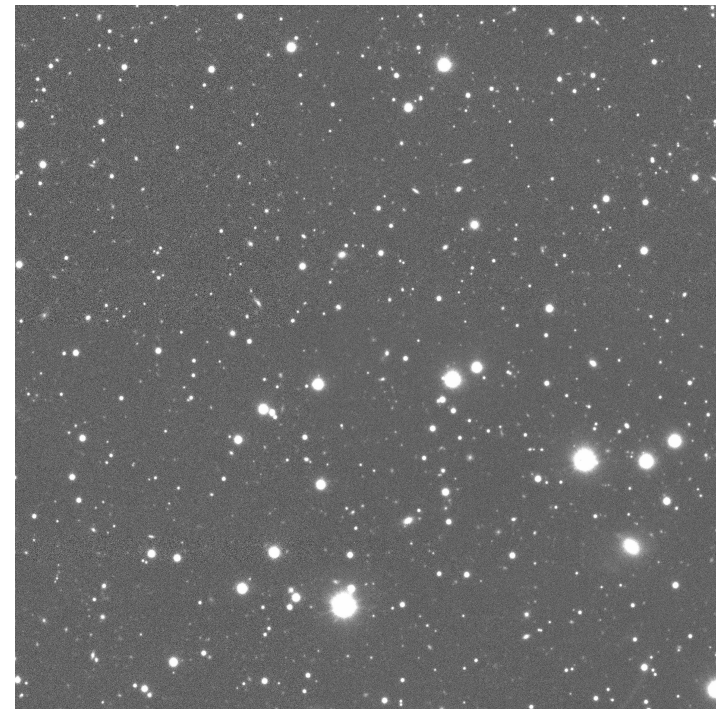
- Continued development of Pegasus+HTCondor+SuperTask framework.
- TBDs re: Butler and SuperTask interfaces, schedule of features needed by production.

Phase 2: Integrate input/output data handling schemes.

- Stage-in/-out from “standalone” data backbone.
- Will need file catalog, metadata catalog, provenance.

Have prototyped HSC pipeline processing using DESDM production framework.

- Plausible at this scale.
- Further work includes ability to submit to Slurm, testing other pipelines (e.g., multifit), extending metadata collected, performing scaling tests, working out Butler interface (e.g., require unique filenames, ability to query for files on metadata).



DATA BACKBONE SERVICES



Priority is to develop minimal Data Backbone for archiving Aux Telescope test stand in Tucson (~Feb 2018) and supporting Batch Processing Service development.

- Soon after (~Aug 2018) – support data archiving and serving for Aux Telescope commissioning; EFD ingestion and ETL.

File Services

- Evaluating Rucio (used by ATLAS project) for distributed data management within the Data Backbone custodial file store.
 - Has triggering mechanisms needed for raw file ingestion, replicates over WAN, supports multiple protocols, possibly has built-in tape semantics.
 - TBD use by workflow system to (pre-)stage files.
- Need file catalog, metadata catalog, provenance.

Administered Database Services

- Working towards consolidated database for production.
 - New DBA will be setting up Oracle and SQL Server prototype systems. Begin to evaluate operational performance (e.g., scalability),
- Support operational schema development, EFD ingestion.

Focus is on provisioning and resource management for containerized application management system.

- Investigations proceeding assuming Kubernetes.
- Ability to draw from and yield to common resource pool.
 - Interaction of Kubernetes with Slurm batch management system.
 - Redirecting sessions to consolidate/repack nodes.

Phase 1 hosts interactive analysis environment sufficient for project-internal use (developers, commissioning, early ops) – i.e., “trusted” users over which there are operational controls; lower elastic demand.

Phase 2 for community – i.e., “untrusted” users, no operational controls.

- Current pre-operations plans for serving commissioning data releases (2020-2022) are data access, minimal-to-no interactive environment.

Have central AA system (group management, account creation, Kerberos, LDAP), central network-based IT security services (IDS, scanning, etc.) operations at NCSA.

Deploying network-based IT security and AA endpoint at the NOAO facility in La Serena.

- ~September 2017

Integrate Windows AD with AA system for Tucson/HQ endpoint.

Deploy AA endpoint at Summit; IT security and AA endpoint at Base Facility.

Integration with applications, business processes.

Major planned deployments and investigations:

- Level 1 Test Stand (ITC for Camera DAQ test stand + Base-side orchestration components).
- AAA and IT security systems at NOAO La Serena facility (configure @ NCSA - July, install in Chile - Sept.).
- Small test system for consolidated database investigations.
- Testing of remote deployment/configuration capabilities for administration of Base Center ITC.
- Refresh Isst-db (EOL).
- Evaluating organization of file system and policies to support production and development use cases and to assess near-term additional capacity needs on development resources.

Analyzing pre-operations capacity needs.

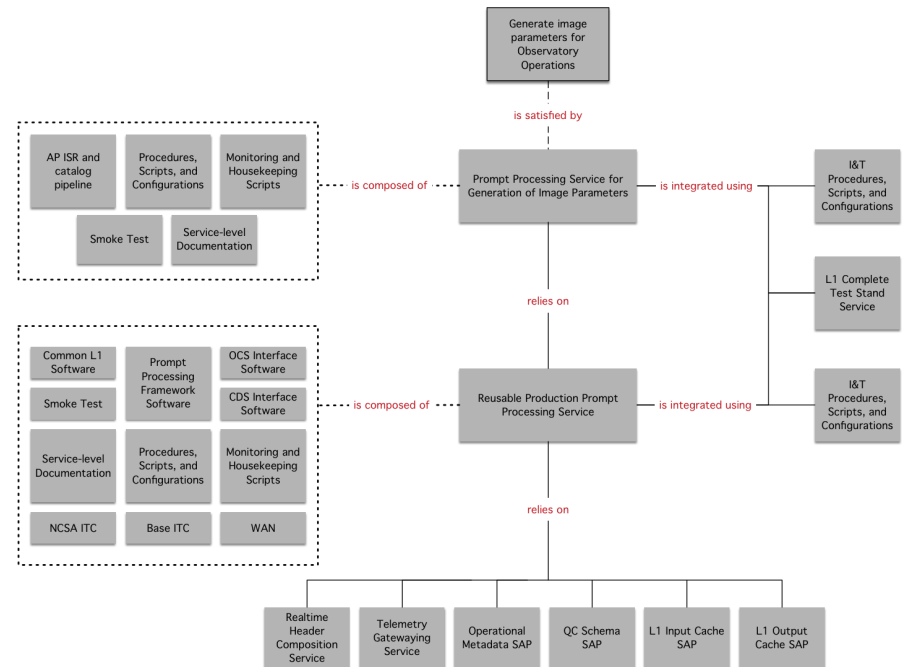
- Model is based on use cases:
 - Raw data archiving for pre-assembly tests of camera at SLAC, aux telescope test stand in Tucson, Summit reverification tests, commissioning observing programs, etc.
 - Assumes high utilization of instruments once they are available.
 - Formal data processing for commissioning, pre-ops releases, etc.
 - Partial - DRs as defined by early ops, TBD additional commissioning.
 - Integration of production services, data challenges, dress rehearsals.
 - TBD.
 - Development support and commissioning (e.g., John's HSC reprocessing doc).
 - Partial - only through FY17.
 - Input from the project is essential.
- So far focusing on retained storage only, need to work on computing capacity.
- Basis for provisioning schedule and costing.
 - Further: allocation of costs between construction, commissioning, (pre-)operations.

HIGH-LEVEL PLANNING & MGMT (2)

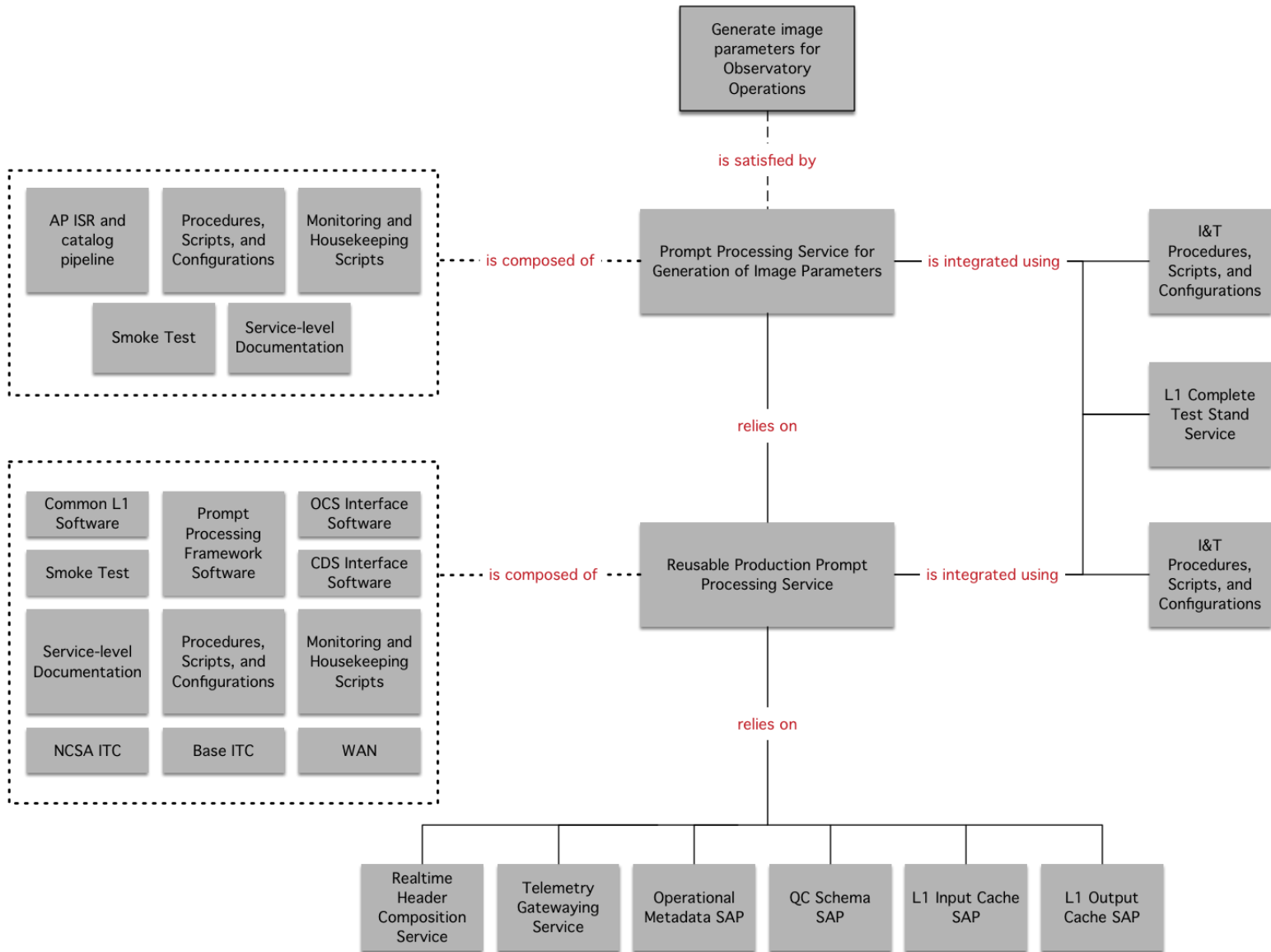


Mapping service dependencies as data (tuples), based on our WBS.

- Vertical cut through the service layers, down to ITC.
- For each service:
 - What are the service components? (software to inflate, hardware to inflated on, procedures, configurations, documentation, verification/smoke tests)
 - What are the service reliances? (reusable production services the high-level service depends on)
 - What is used to integrate the service?
 - Complete model will also include: What does the service comply with? (cross-cutting elements like security)
- Basis for:
 - Scheduling integration of services and components.
 - Sizing ITC.
 - Managing services.
 - Being complete with verification tests, documentation, etc.



SERVICE MAP EXAMPLE



Documentation:

- Concept of operations for each service.
 - Describes satisfaction of needs.
 - Clean up and port conops documents into new DM doc template framework.
 - Check content for consistency with matured operations plan.
- High-level implementation design for each service.
 - Extensive design prose in our detailed WBS.
 - Supplemented by service dependency map for tracing services to main components.
- ICDs