

The background of the slide is a dark blue technical drawing. It features various geometric shapes, including circles, arcs, and lines, rendered in a light blue and yellow color. The drawing appears to be a schematic or blueprint of a complex mechanical or structural system, possibly related to the Large Synoptic Survey Telescope (LSST).

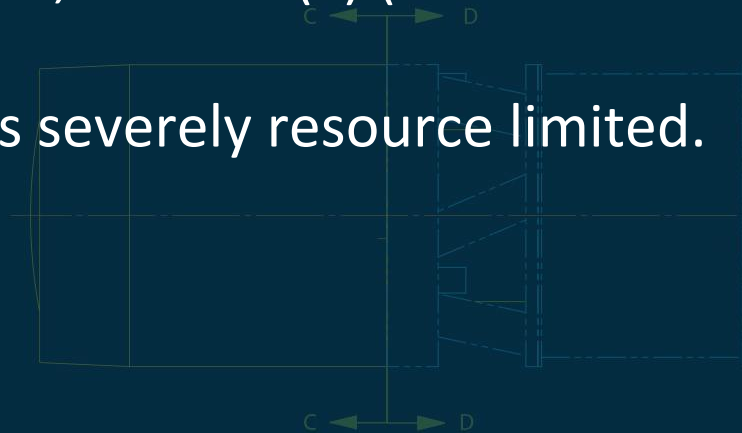
# DM Middleware Updates, 2020-02



*Large Synoptic Survey Telescope*

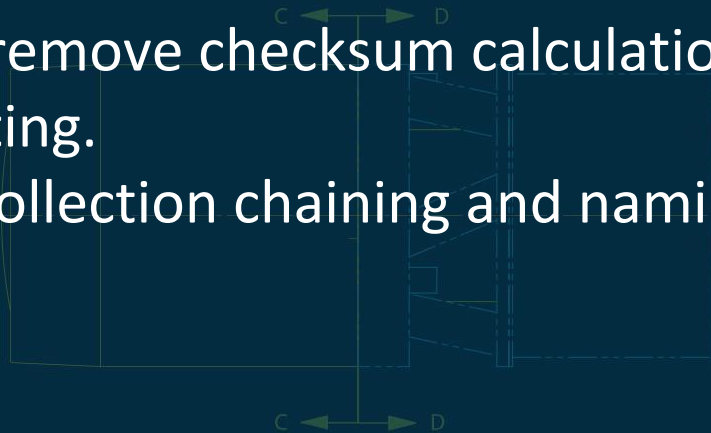
# Management Changes

- During January Fritz Mueller transitioned management of Middleware from DAX to Architecture (Tim Jenness)
- Robert Gruendl is now the product owner.
- Middleware continues to be a virtual team responsible for `daf_butler`, `pipe_base`, `ctrl_mpexec`, and BPS(?) (definition of “what is middleware” is fluid).
- Core team for middleware team is severely resource limited.



# Recent Progress

- Gen 3 features now enables cp\_pipe to be completed.
- Support added for ingest of DECam data files that have multiple datasets per file.
- Refactoring of database interface has simplified access to Oracle and Postgres.
- Registry refactoring and ability to remove checksum calculations has improved ingest speed for testing.
- Finalizing policies for handling of collection chaining and naming conventions.

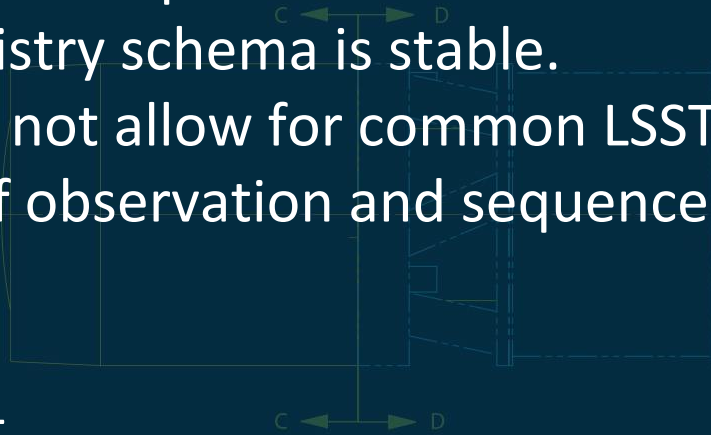


# Gen 2 Deprecation

- Currently 40 open tickets labeled as gen2 deprecation blockers.
- Biggest tickets:
  - Port pipe\_analysis, jointcal, and alert production to Gen3 (not blocked on middleware)
  - ctrl\_pool quantum graph executor.
  - Rewrite daf\_butler ingest to separate ingestion from visit determination and also to support incremental ingest of an exposure.
  - Finish implementing Gen3 support of supported obs packages (drop obs\_sdss for now).
  - Write ingest and butler management command line tools.
  - Optimize running at large scale (cf RC2 in gen2 vs 3)

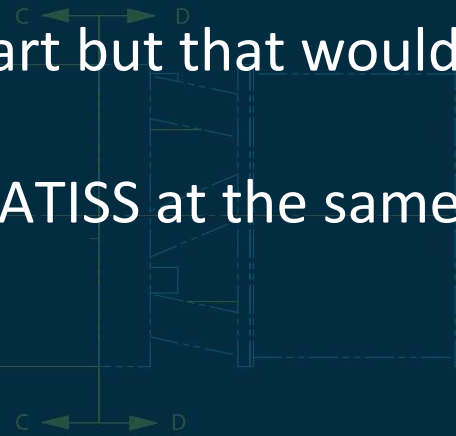
# Other Issues

- afw Filter singleton is beginning to cause real trouble now that Gen3 can support multiple instruments in a single registry.
- Chained registries still not in the near-term plan.
- Registry relationship with ObsCore still unclear. May need to make additions to data model to ease export to VO. We are not yet able to tell people that the registry schema is stable.
- Unified exposure data model does not allow for common LSST instrument practice of using day of observation and sequence number tuples as a dataId.
- Need a gen3 variant of ci\_lsst
- No provenance being recorded yet.
- Versioned camera Geom.



# New Requests

- DMTN-133 describes the OCS Controlled Pipeline System (OCPS) for doing immediate processing at the summit under direct control of observing scripts.
- Would like to use LATISS as a pathfinder for experimenting with this.
- Would prefer to base it on Gen3 from the start but that would require a Gen3 OODS.
  - Can we run a Gen2 and Gen3 OODS for LATISS at the same time?
- How does OCPS relate to display systems?
- Who will work on this? What is its priority?



# Where does BPS fit into this?

- Was not planned for general release this year.
- `ctrl_mpexec`'s `pipetask` runs Gen3 pipeline on single node
- What are the minimum requirements to be usable for gen3 pipeline testing?
  - What provenance is needed?
  - No pipeline restart (other than beginning)?
  - Shared install or code and dependencies part of LSST stack?
  - Where are output logs handled?
- Can we use it instead of porting `ctrl_pool`?
  - Requires HTCondor (either directly or via glidein)

# Upcoming Milestones

- DM-DAX-8: Scatter/Gather support in PipelineTask [May 2020]
  - Jointcal and pipe\_analysis now unblocked.
  - Can claim this in February.
- DM-DAX-12: Gen3 Feature Parity [May 2020]
  - At least 3 months away for parity? (if we have resources)
  - Is feature parity all instruments? Summit using Gen3?
- DM-DAX-13: Gen2 Butler is retired [Jun 2020]
  - How long should we run in parallel?



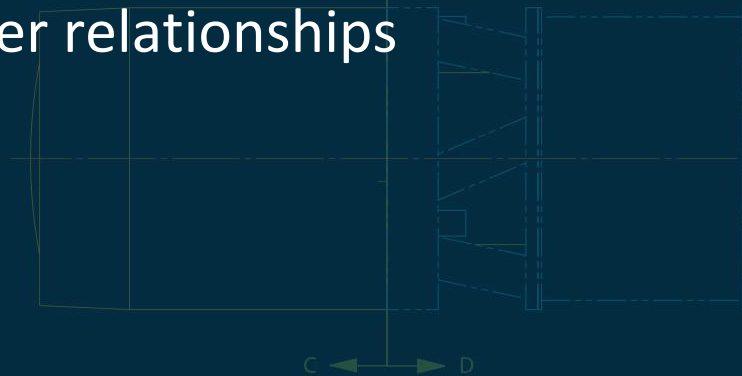
# Ongoing Registry Refactoring



*Large Synoptic Survey Telescope*

# The Registry TODO list, last time

- normalize the dataset table(s)
- denormalize collections
- fix unique conflict resolution
- abstract/encapsulate dataset storage
- split autoincrement IDs (add site/origin to primary keys)
- generalize visit/exposure and filter relationships
- improvements to DatasetRef
- multi-collection butlers
- use ranges for skypix IDs
- subpackage reorganization
- merge SqlRegistry and Registry



# The Registry TODO list, last time

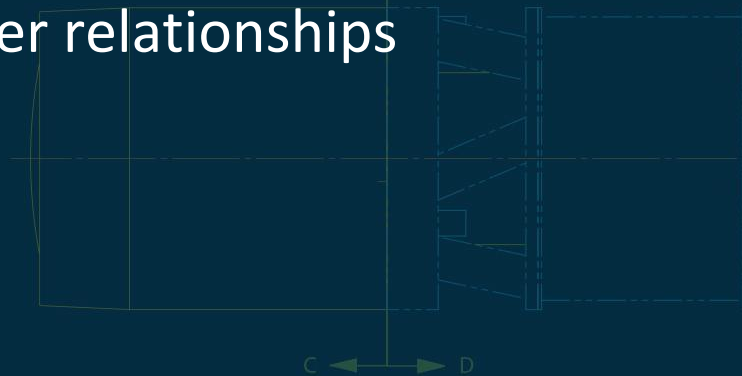
- normalize the dataset table(s)
- denormalize collections
- fix unique conflict resolution
- abstract/encapsulate
- split autoincrement IDs
- generalize visit/exposure
- improvements to Data
- multi-collection butler
- use ranges for skypix
- subpackage reorganization
- merge SqlRegistry and Registry

## Goals:

- stable, future-proof data repository schemas
- performant ingest
- QuantumGraph-generation queries that scale

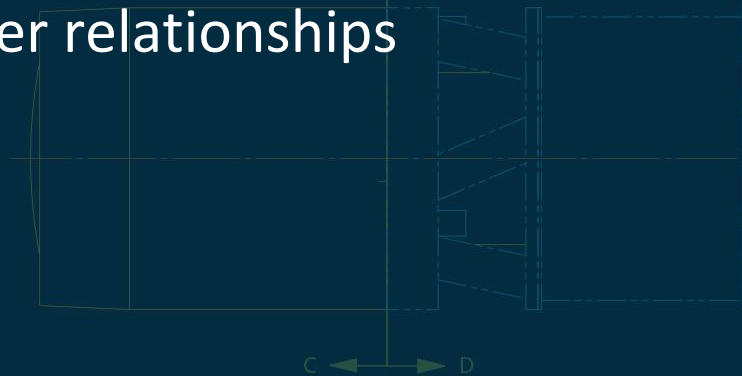
# What's totally done

- normalize the dataset table(s)
- denormalize collections
- fix unique conflict resolution + vectorization
- abstract/encapsulate dataset storage
- split autoincrement IDs (add site/origin to primary keys)
- generalize visit/exposure and filter relationships
- improvements to DatasetRef
- multi-collection butlers
- use ranges for skypix IDs
- subpackage reorganization
- merge SqlRegistry and Registry



# What's almost done

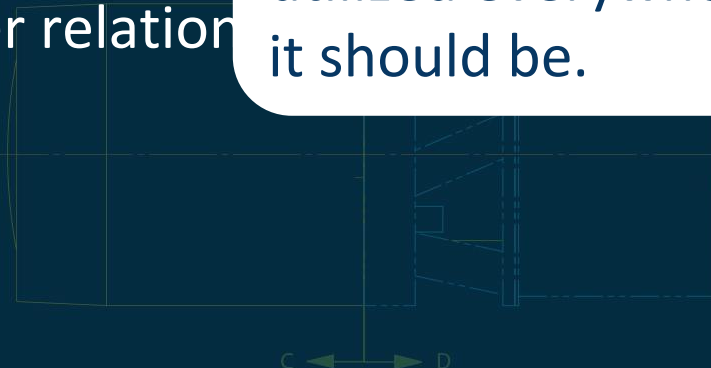
- normalize the dataset table(s)
- denormalize collections
- fix unique conflict resolution + vectorization
- abstract/encapsulate dataset storage
- split autoincrement IDs (add site/origin to primary keys)
- generalize visit/exposure and filter relationships
- improvements to DatasetRef
- multi-collection butlers
- use ranges for skypix IDs
- subpackage reorganization
- merge SqlRegistry and Registry



# What's almost done

- normalize the dataset table(s)
- denormalize collections
- fix unique conflict resolution + vectorization
- abstract/encapsulate dataset storage
- split autoincrement IDs (add site/origin to)
- generalize visit/exposure and filter relations
- improvements to DatasetRef
- multi-collection butlers
- use ranges for skypix IDs
- subpackage reorganization
- merge SqlRegistry and Registry

Low-level code exists, but is not utilized everywhere it should be.



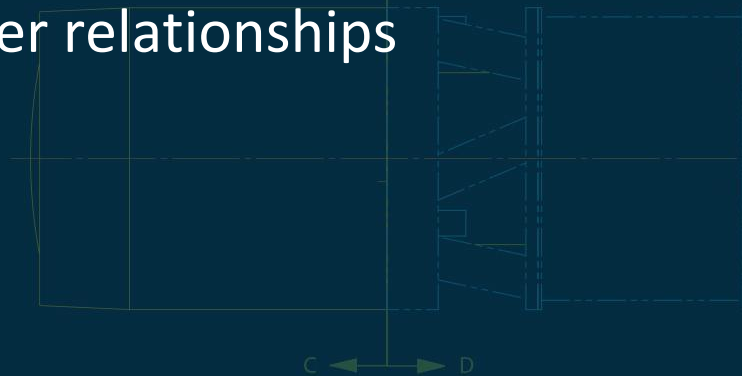
# What's almost done

- normalize the dataset table(s)
- denormalize collections
- fix unique conflict resolution + vectorization
- abstract/encapsulate dataset storage
- split autoincrement IDs (add site/origin to primary keys)
- generalize visit/exposure and filter relationships
- improvements to DatasetRef
- multi-collection butlers
- use ranges for skypix IDs
- subpackage reorganization
- merge SqlRegistry and Registry

RFC-663/DM-21849,  
almost ready for review

# What's fully prototyped

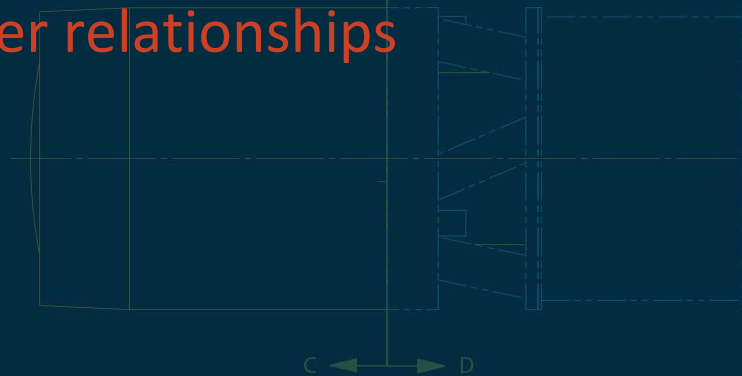
- normalize the dataset table(s)
- ~~denormalize~~ better-normalize collections
- fix unique conflict resolution + vectorization
- abstract/encapsulate dataset storage
- split autoincrement IDs (add site/origin to primary keys)
- generalize visit/exposure and filter relationships
- improvements to DatasetRef
- multi-collection butlers
- use ranges for skypix IDs
- subpackage reorganization
- merge SqlRegistry and Registry





# What's **mostly not done**

- normalize the dataset table(s)
- ~~denormalize~~ better-normalize collections
- fixing unique conflict resolution + vectorization
- abstract/encapsulate dataset storage
- split autoincrement IDs (add site/origin to primary keys)
- generalize visit/exposure and filter relationships
- improvements to DatasetRef
- multi-collection butlers
- use ranges for skypix IDs
- subpackage reorganization
- merge SqlRegistry and Registry



# What's **mostly not done**

- normalize the dataset table(s)
- ~~denormalize~~ better-normalize collections
- fixing unique conflict resolution + vectorization
- abstract/encapsulate dataset storage
- split autoincrement IDs (add site/origin to primary keys)
- **generalize visit/exposure and filter relationships**
- improvements to DatasetRef
- multi-collection butlers
- use ranges for skypix IDs
- subpackage reorganization
- merge SqlRegistry and Registry



probably not needed

# Registry Big Picture

- List of bullets has not grown since before PCW.
- ~June completion of Registry work is still definitely doable (that *is* what we said last time, right?).
- Less clear whether Jim will finish Registry work enough in advance to help finish other Gen3 deliverables by then.
- Still assuming friendly users for multi-user Registry, but I'm now confident we can evolve to multi-schema DB gradually later.
- Not going to be much progress in March (AlgoWS prep).