

Archive Ingest and SDQA

Ingest

Catalog products

- Ingested into temporary database
 - For use in other DRP steps or by SDQA
 - "Patch" updates happen by execution of new Tasks
 - Must be possible to reproduce entire DRP, including any "patches"
- Level 1 products ingested into internal Level 1 database
- Level 2 products ingested into Qserv database
 - Can remove a batch (with queries disabled), then reingest after "patch" update Task
 - Need to track batches and their status

EFD

- Transformed within 24 hours into Level 1 Science Data Archive EFD
 - Transformation includes removal of sensitive data (e.g. personnel-relevant log entries)
 - Transformation includes restructuring schema to be more science-query-friendly
 - Adding join keys
 - Denormalizing
 - Creating views
 - Note: it's possible that the Science Data Archive EFD will not actually be in relational form. Something like a NoSQL document database or BigTable/Hypertable might be more appropriate.
- Cleansed and transformed for Level 2 EFD as part of annual CPP
 - Cleansing includes flagging of invalid data

Image products

- Ingested into archive, including provenance
- Metadata ingested into archive metadata database
- Accessible by internal image services

SDQA

Catalog Products

- Zeroth pass is metrics produced by Tasks
 - Automatic flagging of metrics outside threshold
- First pass is in temporary database
 - Automatic metric generation (for metrics using data from multiple Task executions)
- Second pass is in Qserv database for Level 2 products, internal Level 1 database for Level 1 products
 - Verification of completeness and consistency
 - Large-scale analyses across entire dataset
 - Verification of performance for end users

Image products

- Most SDQA on metadata
- Inspect outliers using internal cutout service