Large Synoptic Survey Telescope

The Science Data Quality Assessment System for LSST

D. Levine (IPAC/Caltech), V. Mannings (IPAC/Caltech), R. Cutri (IPAC/Caltech), T. Axelrod (LSST), C. Claver (NOAO),*
S. Dodd (IPAC/Caltech), G. Helou (IPAC/Caltech), K. Gilmore (SLAC), Z. Ivezic (U. Washington), D. Monet (USNO), T. Schalk (SCIPP),
J. Kantor (LSSTC), T. Tyson (UC Davis) and the LSST Team

We describe the Science Data Quality Assessment (SDQA) system for LSST. SDQA will be an analysis system that examines and reports on the quality of LSST data and derived products from a scientific perspective. The key driver of SDQA planning and practices is enabling LSST to answer the question "can the data obtained support the science requirements of the LSST?"

An SDQA System is:

❖ A comprehensive system, including both automated and human-intensive components, which collects, analyzes and records information about the quality of raw and derived science data and makes that information available to observatory scientists, downstream processing, the science planning/scheduling process and science users of the data.

Why does LSST need SDQA?

❖ LSST will produce large volumes of science data at an unprecedented rate, so rapid & efficient data QA will be critical
  ❖ The data collected at any point may not be of sufficient quality to satisfy the driving LSST science requirements
    ❖ This Information must be fed back to the camera/telescope support scientists for trouble-shooting.
    ❖ It may also affect LSST observation scheduling
    ❖ Any released data must be of known, well-characterized quality standards so scientists are able to correctly evaluate and interpret the data.
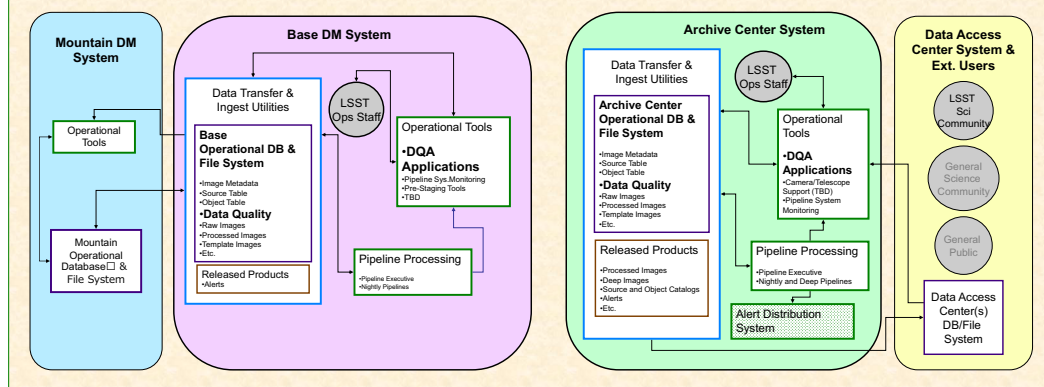
Initial LSST SDQA Concept:

❖ Formulated taking into account experience from data quality assessment on SDSS, 2MASS and Spitzer Space Telescope
❖ Collect summary outputs from all data processing elements and will compile them into concise reports that can be reviewed by a responsible scientist, e.g.
  ❖ software completion status reports
  ❖ statistical summaries
  ❖ listings of out-of-bound physical parameters that can be related to achieved science data quality
  ❖ Facilitate correlation of "engineering data" with with issues raised by SDQA
    ❖ e.g., weather data, calibration data and system health and safety data
  ❖ Characterization of image artifacts
    ❖ e.g. ghosts and glints
  ❖ Monitoring of the quality of the removal (where possible) of the signature of the sky/camera/telescope system
❖ Collate parameters in the individual reports into a series of concise metrics
  ❖ present them, e.g in a web-based form.
  ❖ compare to specific benchmark values
  ❖ enable assignment of a quantitative quality index or score
❖ Automate pre-determined searches for outliers in statistical distributions
❖ Facilitate "drill down" and spot check
  ❖ Identify and analyze specific questionable data products (including raw images)
  ❖ Randomly review apparently healthy data.
❖ A continuous automated search for new or subtle artifacts/effects would be desirable
  ❖ This would primarily supply an extra layer to system characterization, but it might also occasionally identify new science

SDQA in context in the LSST System

Strategy for formulating SDQA Diagnostics:

❖ Initial quality metrics have been identified by analyzing the Science Requirements and formulating a metric for each requirement.

❖ Metrics and science-based diagnostics used by SDSS, 2MASS and Spitzer will be reviewed and incorporated where relevant to LSST

❖ Additional self-diagnostic metrics will be identified for pipeline stages as part of Data Challenge 3 and future Data Challenges.

Examples of DQA diagnostics from SDSS, 2MASS & Spitzer
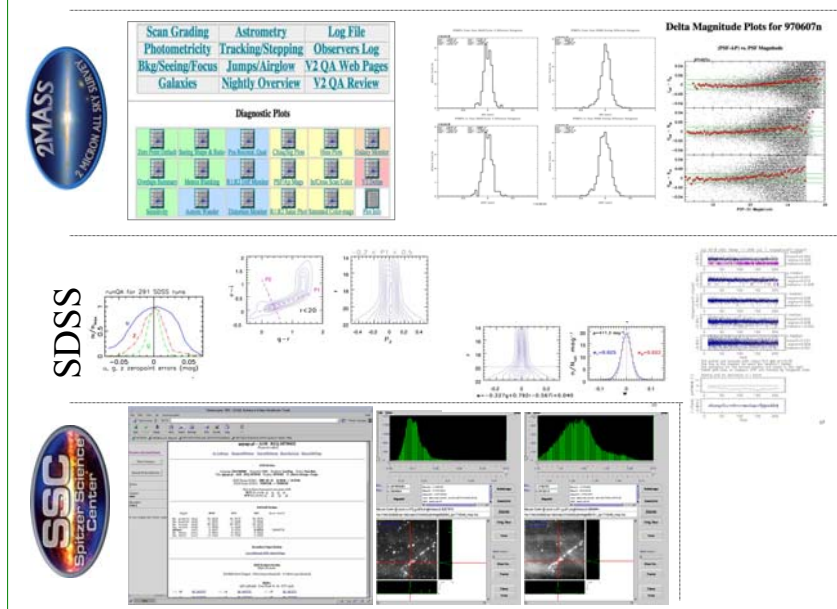
* The Infrared Processing and Analysis Center at Caltech is leading the DQA effort for LSST during the R&D phase