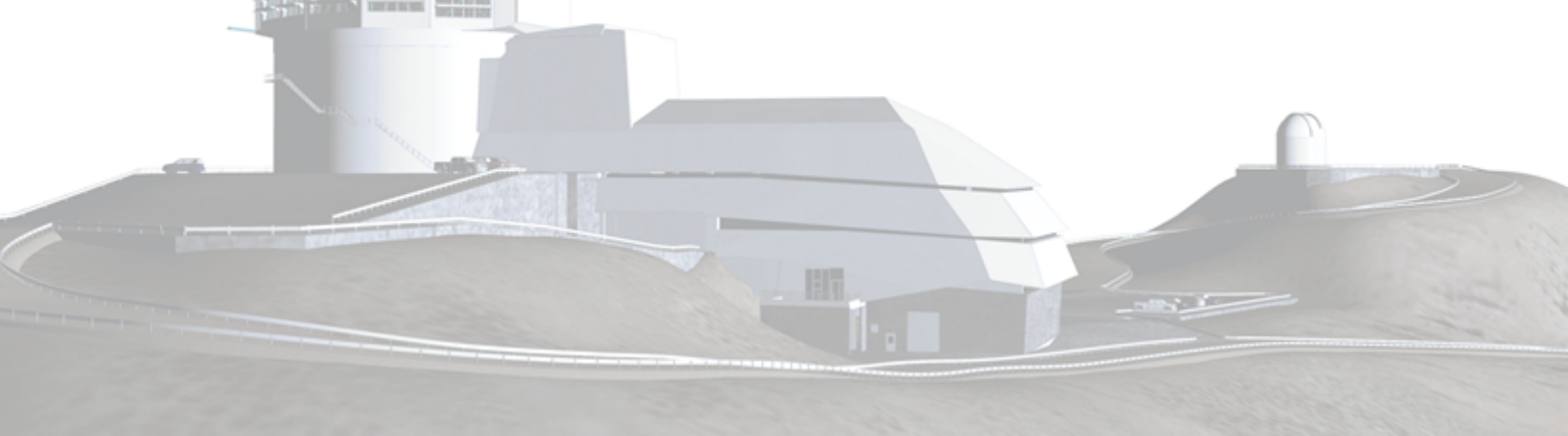
Vera C. Rubin Observatory
Data Management

DM provenance review WG

William O'Mullane

LDM-722

Latest Revision: 2020-10-12



Rubin Observatory

Abstract

As actioned at DMLT F2F this note is to set up a review of provenance.

Rubin Observatory

Change Record

Version	Date	Description	Owner name
1	2019-10-30	First version in accordance with RFC-637	William O'Mullane

Rubin Observatory

Contents

Rubin Observatory

DM provenance review WG

1 Scope

This Working Group is to review the Data Management (DM) provenance based on what is now available in butler gen 3 and the DPPDized data model.

Ideally this working group would convene and finish before the end of 2019.

2 Responsibilities

The Working Group (Working Group (WG)) has the following responsibilities:

- Assess available provenance information in the current Gen3/Butler metadata registry
- Identify provenance gaps in the data model and how they should be filled.
- Classify the resulting provenance needs as critical (needed for commissioning), important (need for full ops) or nice to have.
- Assess whether there are observatory provenance needs that are beyond the scope of Gen3/Butler registry and propose a plan for addressing those.

This group is not intended to solve how missing provenance information should be added - this should fall the middleware group. This group may of course suggest ways to gather/store such provenance information.

3 Specific Tasks

3.1 Assess Available Information

This may have to be provided by the Middleware group based on he gen3 runs of HSC processing and the resulting data model.

3.2 Identify Gaps and needs

The group should then identify use cases e.g' " can I find which PVIs used this calibration file " and determine if they can be answered with the existing metadata. This will obviously be best facilitated by someone very familiar with gen 3 butler.

In general the list of provenance requirement or "use cases" is the most important output of this group.

3.3 Classify use cases

The identified use cases need to be classified and that classification has to be agreed within DM - hence it may need to be in a formal document.

4 Membership

The proposed membership is:

- Frossie Economou (Chair)
- Simon Krughoff
- Robert Gruendl
- Gregory Dubois Felsmann
- Amanda Bauer (EPO Tucson)
- Brian Stalder
- Keith Bechtol

5 Reporting

The WG Chair shall report directly to the DM Project manager weekly.

References

A Glossary

Butler A middleware component for persisting and retrieving image datasets (raw or processed), calibration reference data, and catalogs.

calibration The process of translating signals produced by a measuring instrument such as a telescope and camera into physical units such as flux, which are used for scientific analysis. Calibration removes most of the contributions to the signal from environmental and instrumental factors, such that only the astronomical component remains.

Data Management The LSST Subsystem responsible for the Data Management System (DMS), which will capture, store, catalog, and serve the LSST dataset to the scientific community and public. The DM team is responsible for the DMS architecture, applications, middleware, infrastructure, algorithms, and Observatory Network Design. DM is a distributed team working at LSST and partner institutions, with the DM Subsystem Manager located at LSST headquarters in Tucson.

DM Data Management.

DMLT DM Leadership Team.

EPO Education and Public Outreach.

HSC Hyper Suprime-Cam.

metadata General term for data about data, e.g., attributes of astronomical objects (e.g. images, sources, astroObjects, etc.) that are characteristics of the objects themselves, and facilitate the organization, preservation, and query of data sets. (E.g., a FITS header contains metadata).

provenance Information about how LSST images, Sources, and Objects were created (e.g., versions of pipelines, algorithmic components, or templates) and how to recreate them.

WG Working Group.