



Vera C. Rubin Observatory  
Rubin Observatory Project Office

# Rubin Observability Project

Cristian Silva, Joshua Hoblitt

ITTN-070

Latest Revision: 2023-02-24

**DRAFT**



## Abstract

Scope and requirements of the observability project.

Draft

## Change Record

Version	Date	Description	Owner name
1	2023-01-24	First Version	Cristian Silva, Joshua Hoblitt

*Document source location:* <https://github.com/lsst-it/ittn-070>

Draft

## Contents

<b>1 Introduction</b>	<b>1</b>
<b>2 Scope</b>	<b>1</b>
<b>3 Infrastructure</b>	<b>1</b>
<b>4 Requirements</b>	<b>3</b>
<b>A References</b>	<b>4</b>
<b>B Acronyms</b>	<b>4</b>

Draft

# Rubin Observability Project

## 1 Introduction

The Rubin Observatory is a state-of-the-art astronomical facility designed to study the universe through the Legacy Survey of Space and Time (LSST). With its advanced telescopes and cutting-edge technology, the Rubin Observatory is expected to generate an enormous amount of astronomical data that will offer unprecedented insights into the cosmos. However, managing and processing such a massive amount of data presents a significant challenge, and it is essential to ensure the observability of the system to detect, diagnose, and resolve issues that may arise.

The goal of this document is to set the requirements for an observability project for the Rubin Observatory

## 2 Scope

The observability project for the Rubin Observatory will cover the following system components:

- Hardware and software systems supporting the telescope operations.
- The data processing pipelines used to transform raw telescope data into scientific results
- The storage and retrieval systems used to store and manage astronomical data

## 3 Infrastructure

The following is the summary of the Rubin's Infrastructure

- Foreman instances: 4
- Puppetserver/puppetdb instances: 4

- Hosts known to foreman: 328
- Additional servers coming online this year: 100
- Network devices (switches, PDUs, etc.): 100
- Additional network devices coming this year: 40
- Physical locations: 3
- Logical sites: 4
- K8s clusters: 13
- Rook-ceph deployments: 11
- Graylog instances: 2
- Graylog burst input (per second): 9000
- Graylog log volume day (GiB): 205.1

#### Current software status

- K8s logs shipped via fluentbit
- Host logs via rsyslog
- K8s metrics prometheus
- Host metrics telegraf
- Influx on k8s (per site)
- Grafana on k8s (per site)
- User authentication on IPA

## 4 Requirements

The following requirements are considered essential:

- Provide performance monitoring of hosts and network devices
- Collect log data from hosts and network devices
- Collect audit logs from hosts
- Trigger notifications based on metrics
- Trigger notifications on syslog data
- Scrape k8s data. E.g. collect kverno audits
- Collect switch counters (snmp, etc)
- Collect ceph metrics... e.g. volume data and trigger notifications
- Push notifications via squadcast
- Push notifications to a slack channel
- Push notifications to a kafka topic
- The ability to easily define monitoring checks via config/code in a git repo
- Per site observability/monitoring stack
- Dashboards per k8s ns/pod
- Dashboards per host and group of hosts
- Dashboards for network devices which include per and error counters
- Log view which shows sudo usage. E.g. messages to a slack channel when user sudo's to root
- Useful workflows for jr. and helpdesk
- Dashboard per group of users on ipa
- Simplified dashboard for non-technical staff

- Collect firewall logs
- Provide security related alerts.

The following requirements are considered desirable:

- Single dashboard integration of all sites
- "Per tentant" k8s workspace metric
- "Per tentant" views of hosts logs
- "Per tentant" views of pod logs
- Basic logs/metrics/monitoring of new hosts to happen automatically (e.g. via foreman or puppetdb)

## A References

## B Acronyms

Acronym	Description
DM	Data Management