Some news on computing for LSST

dominique boutigny, fabio hernandez

Journées Rubin LSST-France, Marseille, June 10th, 2024

Centre de Calcul
de l'Institut National de Physique Nucléaire
et de Physique des Particules

doc.lsst.eu

○ [RTN-011](#) Rubin Observatory Plans for an Early Science Program

Rubin Early Data Release Scenario	Jun 2021	Jun 2022	Aug 2023	Apr 2025 - May 2025	Dec 2025 - Apr 2026	Jun 2026 - Jan 2027	Jun 2027 - Jan 2028	Jun 2028 - Nov 2028	Jun 2029 - Nov 2029
	DP0.1	DP0.2	DP0.3	DP1	DP2	DR1	DR2	DR3	DR4
Data Product	DC2 Simulated Sky Survey	Reprocessed DC2 Survey	Solar System PPDB Simulation	ComCam/LSST Cam Data	LSSTCam Science Validation Data	LSST First 6 Months Data	LSST Year 1 Data	LSST Year 2 Data	LSST Year 3 Data
Raw images	☑	☑	☐	☑	☑	☑	☑	☑	☑
DRP Processed Visit Images and Visit Catalogs	☑	☑	☐	☑	☑	☑	☑	☑	☑
DRP Coadded Images	☑	☑	☐	☐	☑	☑	☑	☑	☑
DRP Object and ForcedSource Catalogs	☑	☑	☐	☐	☑	☑	☑	☑	☑
DRP Difference Images and DIASources	☐	☑	☐	☐	☑	☑	☑	☑	☑
DRP ForcedSource Catalogs including DIA outputs	☐	☑	☐	☐	☑	☑	☑	☑	☑
PP Processed Visit Images	☐	☐	☐	☐	☐	☑	☑	☑	☑
PP Difference Images	☐	☐	☐	☐	☐	☑	☑	☑	☑
PP Catalogs (DIASources, DIAObjects, DIAForcedSources)	☐	☐	☐	☐	☑	☑	☑	☑	☑
PP SSP Catalogs	☐	☐	☑	☐	☑	☑	☑	☑	☑
DRP SSP Catalogs	☐	☐	☐	☐	☐	☑	☑	☑	☑

DP: Data Preview
DR: Data Release

DP1: ComCam/LSSTCam Data, 6-7 months after System First Light

LSST Survey Start, 8-12 months after System First Light

DR1: LSST First 6 Months Data, 20-26 months after System First Light
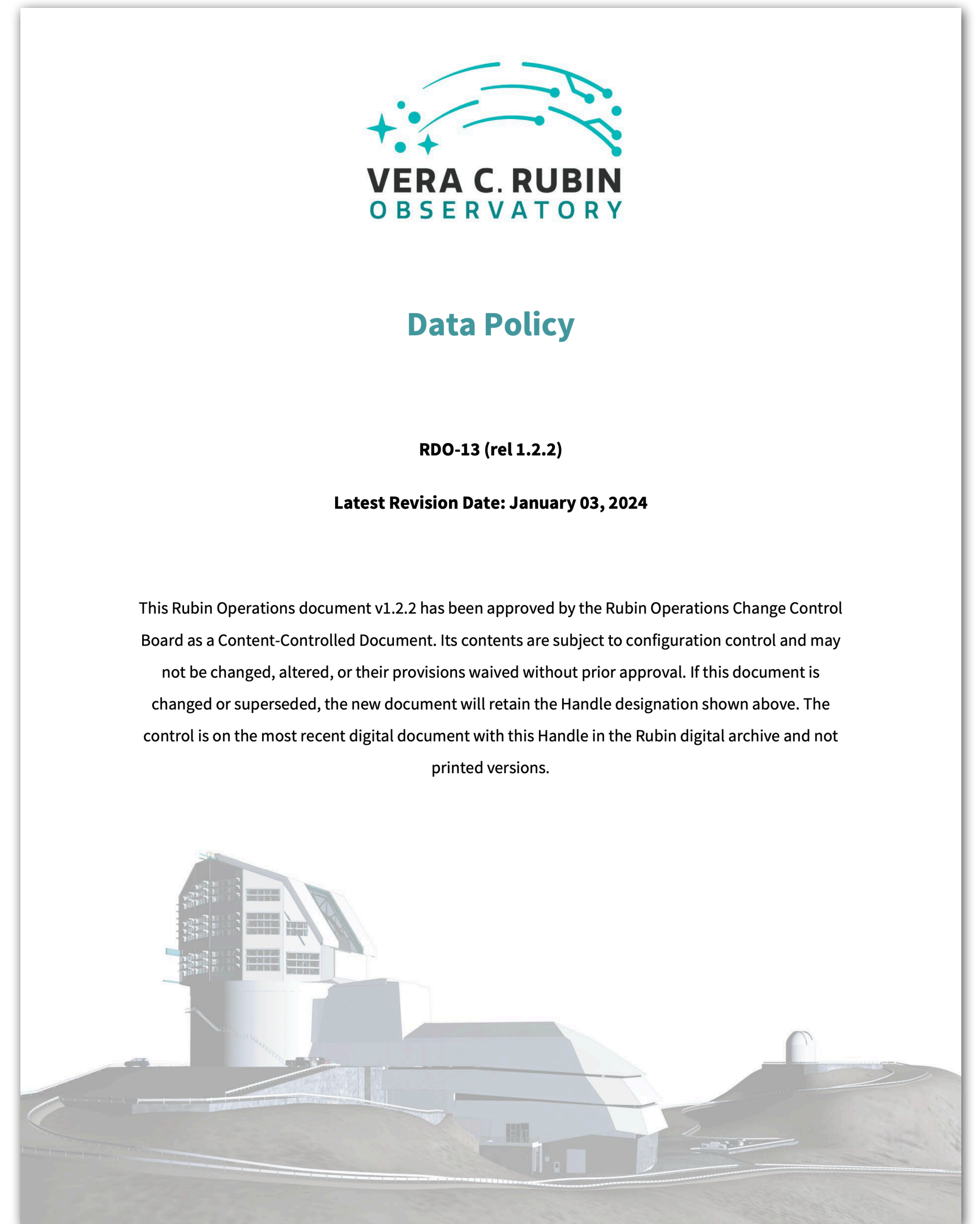
next milestone

○ [RDO-013](RDO-013) Rubin Data Policy

excerpts from the Executive Summary*:

"This document defines policies controlling Rubin Observatory Legacy Survey of Space and Time (LSST) **data rights and access** *for U.S. and Chilean scientists, international members, the worldwide scientific community, and the public.*

An **individual's rights to access, analyze, publish, and/or share** *the full range of LSST data products and services are described. The difference between a* **proprietary data product** *and a* **derived data product** *is described and clarified, with examples."*

* emphases ours

VERA C. RUBIN OBSERVATORY

Data Policy

RDO-13 (rel 1.2.2)

Latest Revision Date: January 03, 2024

This Rubin Operations document v1.2.2 has been approved by the Rubin Operations Change Control Board as a Content-Controlled Document. Its contents are subject to configuration control and may not be changed, altered, or their provisions waived without prior approval. If this document is changed or superseded, the new document will retain the Handle designation shown above. The control is on the most recent digital document with this Handle in the Rubin digital archive and not printed versions.

- As [discussed](#) last December in Lyon, we need to estimate the computing equipment required for **science analysis**
  as opposed to the requirements for data release processing for which we have a reasonable estimation

- Dominique is conducting a study taking inputs from the science coordinators
  [initial estimation](#) of the computing (mostly CPU) and disk storage capacity needed for analysing Rubin data at CC-IN2P3

  *important for determining the **budget**, making **contribution statements** to the science collaborations (e.g. DESC) and ultimately **purchasing** and **provisioning** the equipment*

- Science use cases considered so far
  3x2pt + cluster analysis, simulations, synthetic source injection, supernovae studies

- Your input could help improve this estimation, please give feedback to Dominique
  e.g. our understanding of the needs of GPU-based processing capacity is limited (how much capacity would be needed? for development vs. large-scale production? etc.)

- The availability of Rubin data for science analysis is not that far in the future
  *we need to get prepared for the science analysis **<u>you</u>** plan to conduct with these data*

EXERCISES ON DISTRIBUTED PROCESSING

- A processing exercise involving the 3 Rubin data facilities (FR, UK, US) is ongoing

  similar to what we will be doing for data release processing, although at a smaller scale

  input datasets: [public data release 2 of Hyper Suprime-Cam Subaru strategic program](#)

- This processing involves executing at each facility the LSST science pipelines for processing its assigned fraction of the sky

  ultimately collecting and combining the generated data products at USDF

- This exercise is an opportunity to develop, test and refine the tools for extracting datasets from each facility's local butler repo

  and transporting them to another facility for ingestion there

  we use CERN's [Rucio](#) and [FTS](#) as the Rubin-agnostic tools to transport the data, in addition to tools specifically developed by Rubin for extracting the relevant datasets from butler repos and ingesting them at destination

CC-IN2P3: OPERATING SYSTEM UPGRADE

○ The operating system used in production for both the login and batch farm is CentOS 7

  it will reach its end of life end of June 2024 ([details](#))

○ Taking into account the supported experiments' needs, CC-IN2P3 decided to migrate to RedHat Enterprise Linux v9

  you likely received an [announcement](#) on May 24th

○ There are already hosts running RHEL v9 in both the login and batch farms that you can use to get familiar with the system and test your tools

  see the [documentation](#) for details on how to get started

○ Schedule

  1. *June 2024: 50% to 100% of the* `lsst` *partition of the Slurm farm will run RHEL v9*

  2. *September 2024: 50% of the batch farm will run RHEL v9*

  3. *December 2024: 100% of the batch farm will run RHEL v9*

RUBIN SCIENCE PLATFORM

- Are you contributing to commission the observatory?

  you will likely need to get familiar with both the [working environment at USDF](#) and the Rubin Science Platform (RSP)

- A specific slot is devoted later on today to demonstrate how the RSP can be used for your analysis

  for reference, a [presentation](#) about this was made last December