

Rubin Observatory

Vera C. Rubin Observatory
Data Management

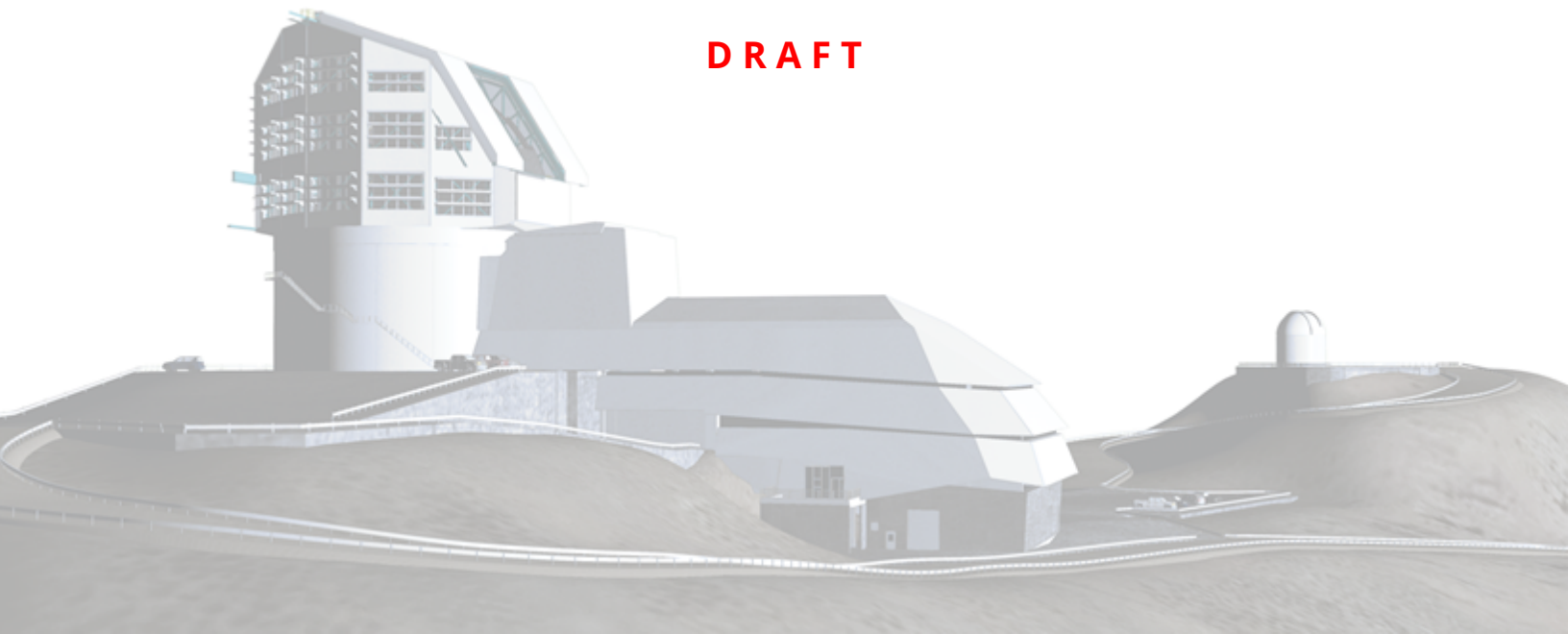
Guidelines for Community Participation in Data Preview 0

The Community Engagement Team and the Operations Executive
Team

RTN-004

Latest Revision: 2020-12-18

DRAFT



Abstract

This document provides an overview of the goals and policies for community participation in the Rubin Observatory pre-operations Data Preview 0 (DPO). Starting in mid-2021, simulated LSST-like data sets will be made available in the Rubin Science Platform (RSP), and up to 300 RSP accounts will be allocated to community participants. RSP access is limited for DPO due to the Rubin pre-operations team's limited scope in providing support for data software and services that are still in development. The term "DPO delegate" has been adopted to reflect the fact that this limited number of participants will be representing the science community as early users of the RSP. This document describes the diversity-driven application and selection process for DPO delegates, and briefly describes some of the resources, support, and activities that will be available to them.

Change Record

Version	Date	Description	Owner name
1	YYYY-MM-DD	Unreleased.	Melissa Graham

Document curator: Melissa Graham

Document source location: <https://github.com/rubin-observatory/rtn-004>

Draft

Contents

1 Introduction to Data Preview 0	1
1.1 The Goals of DP0	1
1.2 Two Stages: DP0.1 and DP0.2	2
2 Delegate Benefits and Responsibilities	2
2.1 Benefits	3
2.2 Responsibilities	3
3 Delegate Selection Process	4
3.1 Delegate Groups	5
3.2 Selection Committee	5
3.3 Selection Criteria	6
3.4 Equitable Access	7
3.5 Application Form Contents	7
3.6 Selection Timeline	9
4 Resources for DPO Delegates	9
4.1 Policies for RSP Accounts	10
4.2 Support for Delegates	11
4.3 Example Delegate Actions	11
A References	12
B Acronyms	12

Guidelines for Community Participation in Data Preview 0

1 Introduction to Data Preview 0

There are three planned Data Previews (DPs) during the period leading up to the start of Rubin Operations, termed pre-operations. DP0 will be based on simulated data. DP1 and DP2 will be based on data taken during Rubin Observatory commissioning; DP1 based on data taken with the Rubin Observatory Commissioning Camera (ComCam), and DP2 based on data taken with the LSST Camera (LSSTCam) – the science instrument for Rubin Observatory. DP0 is planned for 2021 with DP1 and DP2 currently foreseen for 2022 and 2023 respectively. The Data Previews will enable the community to access and analyze LSST-like data in the Rubin Science Platform (RSP) on a shared-risk basis.

1.1 The Goals of DP0

As described in RTN-001, the two goals of DP0 are:

1. To serve as an early integration test of the Data Management System (DMS) and inform further development of Rubin's LSST Science Pipelines (hereafter LSST Science Pipelines) and the Rubin Science Platform (RSP).
2. To prepare the community to use the LSST data products and services, and especially to enable early science with commissioning and year 1 data.

In order to achieve these goals, community participation in DP0 has two main components: (1) performing scientific exercises with the data products and services in the RSP, and (2) taking action to inform or enhance the performance of the DMS, the RSP, or the science community.

Participants from the science community, "DP0 delegates", will be data rights holders¹ who represent a diversity of backgrounds. The term "delegates" reflects the fact that these individuals will be representing the science community as learners and as contributors to the achievement of the DP0 goals, as an essential step along the path towards the success of the LSST.

¹Data rights holders are US or Chilean scientists, or individuals associated with an International Contribution (in-kind) proposal in progress, as defined in RDO-013.

The nature of the two goal-driven components of DP0 participation listed above will vary according to individual delegates' interests and experience.

1.2 Two Stages: DP0.1 and DP0.2

The simulated data set for DP0 was generated by the Dark Energy Science Collaboration (DESC) and is being used under an agreement between the DESC and Rubin. This data set is the second in a series of data challenges, and is thus named DC2. More information about DC2 can be found in the DESC DC2 paper (LSST Dark Energy Science Collaboration et al., 2020) and on the DESC website².

The DP0.1 data set will contain the DC2 images and catalogs processed by the DESC. The DP0.1 data products will be available in the RSP in the format as processed by the DESC, which is similar to that of planned LSST data products. The DP0.1 data products that are made available to DP0 delegates will all be *science-ready*, which means that they have passed full-scale validation and quality analysis. The projected date by which these DP0.1 science-ready data products will be made available to DP0 delegates in the RSP is June 30 2021. Therefore, in Section 3.6 the first stage of DP0 delegates are scheduled to have their accounts activated by June 30 2021.

The DP0.2 data set will contain the DC2 images and catalogs reprocessed by the LSST Science Pipelines. The DP0.2 data products will be available in the RSP in a format that is consistent with the planned LSST data products. The projected date by which the DP0.2 science-ready data products will be made available to DP0 delegates in the RSP is March 31 2022. Therefore, in Section 3.6 the second stage of DP0 delegates are scheduled to have their accounts activated by March 31 2022.

Further details about the DP0.1 and DP0.2 data sets will be presented elsewhere.

2 Delegate Benefits and Responsibilities

As DP0 access will be a limited resource for the science community, it comes with both benefits and responsibilities.

²<https://lsstdesc.org/DC2-production/Documents/>

2.1 Benefits

Anticipated benefits for DPO delegates include:

- an accelerated learning experience in the Rubin Science Platform
- becoming more competitive for grants and early science
- opportunities to take a leadership role in their field and/or Science Collaboration
- having options to publish or publicize the results of their DPO work (software, analysis)
- the ability to advocate for RSP developments beneficial to their scientific field

On-boarding resources and organizational infrastructure provided by Rubin Observatory will be necessary for DPO delegates to obtain the benefits. These are described in Section 4.2.

These potential benefits of DPO participation motivate the focus on diversity and equity in the DPO delegate selection process in Section 3.4.

2.2 Responsibilities

The most important responsibility is that DPO delegates learn how to use the RSP via the on-boarding process well enough to perform some basic scientific exercises, and then take action that informs or enhances the DMS, the RSP, or a Science Collaboration.

Such actions are envisioned to be simple and optional (they will not be tracked), and it will be left to the discretion of individual delegates to decide what kind of actions best match their interests and expertise. Examples of delegate actions are provided in Section 4.3.

The time spent by DPO delegates will not be tracked and there is no minimum (or maximum), but the estimated time commitment for DPO delegates to gain at least some of benefits of participation is approximately 12-36 hours (e.g., 1-3 hours per week over three months). This time can be distributed in any manner that is convenient to the delegate, and allocated accounts do not expire³.

³In the case of oversubscription for the 300 RSP accounts, delegates might be asked to relinquish their accounts if they are not using them, in order to make way for others (Section 4.1).

In addition to the above, all DP0 delegates will be expected to:

- abide by the data rights policy in RDO-013
- adhere to security policies regarding their RSP accounts at the IDF
- abide by usage policies of the RSP and the IDF
- use the on-boarding resources to learn about the RSP
- follow the off-boarding steps if/when appropriate
- engage in DP0-related discussions via the Community Forum when possible
- help other DP0 delegates resolve issues when possible

Policies regarding account usage, security, and off-boarding are described in Section 4.1, and on-boarding resources for DP0 delegates in Section 4.2.

3 Delegate Selection Process

There will be 300 RSP user accounts at the IDF available for science community DP0 delegates, in addition to the 150 accounts for Rubin staff members. This marks an almost order of magnitude increase in the number of community accounts in the current NCSA instance of the RSP (via the Stack Club).

The number of RSP accounts remains constrained during the Rubin Observatory construction and pre-operations phase – and especially for DP0 – due to the Rubin pre-operations team’s limited scope in providing support for data software and services that are still in development. Time is needed to scale up in a sustainable manner. At least by the start of Rubin Observatory Operations, all data rights holders will be able to have RSP accounts. It is a primary goal of the Rubin pre-operations team to provide RSP access as soon as possible to all data rights holders.

All DP0 delegates must be data rights holders, e.g., US or Chilean scientists, or individuals associated with an International Contribution (in-kind) proposal in progress RDO-013.

3.1 Delegate Groups

Achieving the goals of DP0 requires a broad diversity in the DP0 delegates, each with their own experience, expertise, or novice perspective, and each representing and reporting back to different parts of the broader community.

The primary goal of this application process is to ensure that the final set of DP0 delegates includes the full range of diversity in the science community, and provides the beneficial opportunities of DP0 participation equitably to them.

To guide the selection process, the groups of DP0 delegates are defined in Table 1, and the *approximate* number of delegates in each group estimated.

TABLE 1: Approximate targeted distribution for diversity-based DP0 delegate groups.

Group	~%	Description
A	10	experienced users/builders of the DESC DC2 simulated data set
B	10	experienced users of science platforms (e.g., RSP, SciServer, DataLab)
C	10	representatives of teams building iDACs (independent Data Access Centers)
D	10	scientific expertise in Rubin science pillars, deep imaging, big data, etc.
E	10	representatives of small and/or underserved US institutions and colleges
F	10	self-identified novice-level learners and students (and students' advisors)
G	10	representatives of the Chilean astronomy community
H	30	representatives of the Science Collaborations

Representatives in all groups have the same expectations and responsibilities, as in Section 2.2. The selection committee described in Section 3.2 will consider all applicants equally for all groups with which applicants self-identify, using the diversity-based selection criteria in Section 3.3. Applicants will be able self-identify with multiple groups on the application form (Section 3.5). Specifically, please note that Chilean astronomers and Science Collaboration members are *not* restricted to gaining a delegate position via Groups G and H.

3.2 Selection Committee

The delegate selection committee will be facilitated by the Rubin Observatory Community Engagement Team (CET), and will contain representatives from (or identified by) the following sources. Individual committee members will be identified by the CET and the Rubin directorate.

TABLE 2: Approximate targeted distribution of selection committee representation.

Number	Source
4	Science Advisory Council (SAC) and the LSST Science Collaborations
3	Community Engagement Team and Rubin Observatory Director's Office
3	Rubin Operations (e.g., Data Production, SPaRE, RSP, In-Kind Program Coordinators)
2	NOIR Lab (e.g., Community Science Data Center), DOE, or other institutions
2	Chilean astronomical community

To keep committee participation manageable, committee members will be primarily tasked with providing shortlists of applicants for one or more groups. For example, committee members from the SAC and Science Collaborations would be primarily tasked with generating a shortlist for Group H, but would also provide input on the other groups. The CET and the Rubin directorate will merge these shortlists into a final list of DPO delegates.

All selection committee members will use the diversity-based selection criteria in Section 3.3 when forming shortlists for the various groups. Prior to meeting, the CET will facilitate a session on unconscious bias for the selection committee members. The capacity to view form results without names will be provided to the committee.

3.3 Selection Criteria

Diversity in representation is the top priority. The groups described in Section 3.1 are designed to achieve a diverse population of DPO delegates in several facets: experience, expertise, career level, institution type, geographical location, and scientific field. The selection committee will also consider these aspects of diversity within the groups.

Short-form write-in responses will be solicited from applicants so that they may explain their relationship to the groups with which they self-identify. However, applicants will not be asked to, e.g., provide a robust scientific justification for their participation in DPO, or a detailed plan for their DPO delegate action (Section 4.3). While it might at first seem reasonable to consider aspects such as "science impact" or "contributed value" in a selection process like this, doing so could run counter to diversity initiatives (i.e., not all applicants will be equally able to express such aspects in language familiar to the committee members). In a supportive environment with adequate resources (Section 4.2), all DPO delegates should have an equivalent potential to succeed in their endeavors.

3.4 Equitable Access

There are 3 main ways in which equitable access to DP0 is being supported: broad advertising, a staged application process, and on-boarding resources.

Broad advertising is necessary for delegates to self-identify as potential participants, and for reaching potential delegates who are not yet actively participating in Rubin science. The CET will work together with the Rubin Observatory communications team to ensure the call for applications is broadly advertised. The CET will also work with the Rubin directorate to identify individuals in groups A through E, and reach out directly to encourage applications.

A staged application process is implemented to maximize the inclusion of delegates who would not be able to make a time commitment at the start of DP0.1, and because it is recognized that it might take time to achieve a diverse set of DP0 delegates and to provide equitable access to DP0 (Section 3.6).

On-boarding resources and a supportive community will be generated and maintained by the CET (Section 4.2). Such resources are recognized as necessary for delegates to acquire the benefits of DP0 participation, regardless of their past experience or expertise.

3.5 Application Form Contents

A diversity-based selection process requires that adequate appropriate information be collected from applicants. The information on the application form will also be used by the CET to help design the delegate resources described in Sections 4.2 and 4.3.

The application form should be self-explanatory – applicants should not *have to* refer to this document (or any other) in order to complete the form – but the form should also link to additional information about DP0. The application form should include a preamble with the deadline for consideration, and information for a point-of-contact in case of questions.

Below is a proposed list of the data that should be collected. All of the drop-down menus should have multiple selections enabled.

Personal Information

- first name, surname (2 write-in fields)
- pronouns (drop-down menu with write-in option)
- affiliation(s) (drop-down menu with write-in option)
- city, country (2 write-in fields)
- email address (1 write-in field)
- first and second languages, preferred correspondence language (3 drop-down menus)
- career stage (drop-down menu with write-in option)

DP0-Related Information

- science keywords (1 short-form write-in field)
- Science Collaboration (drop-down menu that includes 'none')
- delegate group identification (drop-down menu that includes 'none')
- delegate group relation (one short form write-in field per group, plus one for 'none')
- group F plus-one request (a yes/no option for applicants to indicate if they are a student/advisor in need of an additional RSP account for their advisor/student).

Applicant Confirmations

- time commitment (yes/no to confirm recognition of minimum expected time commitment)
- delegate responsibilities (yes/no to confirm recognition of expectations and responsibilities)
- data rights (yes/no/idk to confirm data rights)

3.6 Selection Timeline

The application form will remain open, accumulating submissions, for the duration of DP0. Applications will be reviewed by the selection committee in two stages which coincide with the projected dates when the DP0.1 and DP0.2 science-ready data products will be made available to DP0 delegates in the RSP (Section 1.2). However, as described in Section 2.2, delegates need not begin their DP0.1 participation at the time of account activation and may distribute their time over DP0 at their convenience. The DP0.1 data will be available to delegates who join with the second stage.

To ensure that at least some accounts are available for new people to join for DP0.2, 25% of the 300 total RSP accounts will not be allocated until the second stage (DP0.2). The following paragraph provides additional motivation for this multi-stage approach to accounts allocation.

The entire duration of DP0 is long, at least 1 year and potentially longer. It is anticipated that many delegates would not be able or willing to make time commitment for DP0.1 (minimal as it may be; Section 2.2). Furthermore, some scientists might be more interested in the DP0.2 data set than DP0.1. Additionally, achieving diverse and equitable access to DP0 might also take some time: the advertisements must percolate, people must free up time in their schedules or obtain grants to fund their work, and the DP0-related resources and support network will become enriched over time (and thus more attractive to a wider variety of participants). To avoid the perception that access to any of DP0 is restricted to individuals who can commit to participating at the time of DP0.1, 25% of the 300 RSP accounts will be reserved for DP0.2. Any accounts relinquished by DP0.1 participants will be added to this reservation and reallocated for DP0.2.

Applications might also be reviewed on intermediate timescales in order to more fully include members of the groups identified in Table 1 and achieve the diversity-related goals of DP0 participation. On such intermediate timescales the applications might be reviewed only by the Rubin Operations executive team and/or the Community Engagement Team instead of reconvening a full selection committee.

4 Resources for DP0 Delegates

TABLE 3: Projected DP0 delegate application process timeline.

Stage	% (of 300)	Applications Due	Delegates Notified	Data Available
DP0.1	75%	Apr 30 2021	May 31 2021	Jun 30 2021
DP0.2	100%	Jan 31 2022	Feb 28 2022	Mar 31 2022

4.1 Policies for RSP Accounts

For the duration of DP0, the RSP will be installed in the Interim Data Facility (IDF) deployed on the Google Cloud Platform. More information about the IDF and its installation will be available elsewhere. Access to the RSP and the DP0 data products and services are limited to data rights holders, and all DP0 delegates must abide by the data rights policies in RDO-013.

All DP0 delegates will be asked to abide by specific policies related to account authentication, security, use, and deactivation. The full details for these policies will be provided to delegates prior to account creation.

Delegates may not share accounts or passwords, but will be free (and encouraged!) to provide tours of the RSP or live demonstrations of their work to individuals who are not DP0 delegates. Account usage policies might include, e.g., best practices for starting and stopping servers, understanding that computational processes which overburden the system might be curtailed, abiding by storage capacity quotas, and following on- and off-boarding steps.

As the total number of accounts will be limited to 300, if applications are oversubscribed then delegates who don't appear to be using their accounts might be periodically contacted and asked to consider voluntarily deactivating to make way for others. In the case of a high over-subscription for DP0 participation, account holders who do not respond to a sustained number of such inquiries might have their access suspended to make way for others.

However, delegates who are given accounts at the start of DP0.1 but who do not have the work-hours to start using them for a few months (due to prior commitments, teaching loads, etc.) will be accommodated. Active, responsive delegates who are given accounts for DP0.1 may keep them, and will not need to reapply for DP0.2.

Policies related to account suspension will be clarified for DP0 delegates during on-boarding. Note that this is a temporary measure for the Data Previews, and not how accounts will work during Operations.

4.2 Support for Delegates

It is recognized that on-boarding resources and organizational infrastructure provided by Rubin Observatory will be necessary to enable equitable access and benefits for all delegates, including novice science platform users (Section 3.4).

The full set of resources will be presented elsewhere, but would include, e.g., virtual live sessions and recordings of RSP demonstrations, user guides and documentation, and jupyter notebook tutorials. The organizational infrastructure will include online open forums for question-and-answer, issue resolution, and general discussion (e.g., Community.lsst.org), as well as regularly scheduled working groups and “office hours” run by Rubin staff. Opportunities will be provided for DPO delegates to communicate with the CET about needed resources and support (e.g., periodic surveys, drop-in discussion sessions).

DPO delegates are responsible for engaging with the resources in order to learn about the RSP and the DPO data set (Section 2.2). Creating or enhancing DPO resources and support infrastructure – e.g., writing tutorial notebooks, leading a collaborative working group, helping other delegates resolve issues – are all potential delegate actions (Section 4.3).

4.3 Example Delegate Actions

As described in Section 2.2, the most important responsibility of DPO delegates is to learn how to do scientific analyses in the RSP, and then take action that informs or enhances the DMS, the RSP, or a Science Collaboration.

The objective is for these delegate actions to extend the benefits of DPO beyond the limited number of DPO participants that Rubin Observatory is able to support, and to maximize the impact of DPO with respect to the goals of testing and informing development of the RSP and preparing the scientific community for early science (Section 1.1).

These actions are envisioned to be simple and enjoyable, and they are completely voluntary and will not be tracked. Group work is encouraged. It is left to individual delegates to decide what kind of actions best match their interests and expertise.

Examples of potential delegate actions include but are not limited to:

- participate in feedback surveys managed by Rubin Observatory
- provide a public demo of some science analysis (e.g., Jupyter Notebook)
- small written reports on activities (e.g., in Community Forum)
- short presentation on activities during topical DP0-related virtual workshops
- video demonstrations of “how-to” with the RSP
- work through an “RSP Test Checklist” provided by Rubin Observatory
- leading working groups composed of other DP0 delegates
- helping to resolve reported issues or support other delegate’s work

A References

[RDO-013], Blum, R., the Rubin Operations Team, 2020, *Vera C. Rubin Observatory Data Policy*, RDO-013, URL <https://ls.st/RD0-013>

LSST Dark Energy Science Collaboration, Abolfathi, B., Alonso, D., et al., 2020, arXiv e-prints, arXiv:2010.05926 (arXiv:2010.05926), ADS Link

[RTN-001], O’Mullane, W., 2020, *Data Preview 0: Definition and planning.*, RTN-001, URL <http://RTN-001.lsst.io>

B Acronyms

Acronym	Description
B	Byte (8 bit)
CET	Community Engagement Team
ComCam	The commissioning camera is a single-raft, 9-CCD camera that will be installed in LSST during commissioning, before the final camera is ready.
DC2	Data Challenge 2 (DESC)

DESC	Dark Energy Science Collaboration
DM	Data Management
DMS	Data Management Subsystem
DOE	Department of Energy
DP0	Data Preview 0
DP1	Data Preview 1
DP2	Data Preview 2
IDF	Interim Data Facility
LSST	Legacy Survey of Space and Time (formerly Large Synoptic Survey Telescope)
NCSA	National Center for Supercomputing Applications
NOIR	NSF's National Optical-Infrared Astronomy Research Laboratory; https://nationalastro.org
RDO	Rubin Directors Office
RSP	Rubin Science Platform
RTN	Rubin Technical Note
SAC	Science Advisory Committee
US	United States