



LSST Photo-z Virtual Forum Series

Data Management System Science Team

Monthly drop-in seminars for everyone.

These slides will be
available at ls.st/clo4683





Vera C. Rubin Observatory
Data Management

A Roadmap to Photometric Redshifts for the LSST Object Catalog

M. L. Graham, J. Bosch, L. P. Guy,
and the DM System Science Team.

DMTN-049

Latest Revision: 2021-02-03

full: ls.st/dmtn-049

summary: ls.st/clo4682

Introduction to Data Management (DM) Hosts

Leanne Guy, DM System Science Team (SST) Lead

Melissa Graham, DM-SST Science Analyst

Jim Bosch, DM Data Release Pipelines (DRP) Scientist

[+ any other DM staff helping out]

Reminder of Future PZ Virtual Forum Sessions

Monthly drop-in sessions; find schedule at ls.st/clo4683.

Today's Agenda

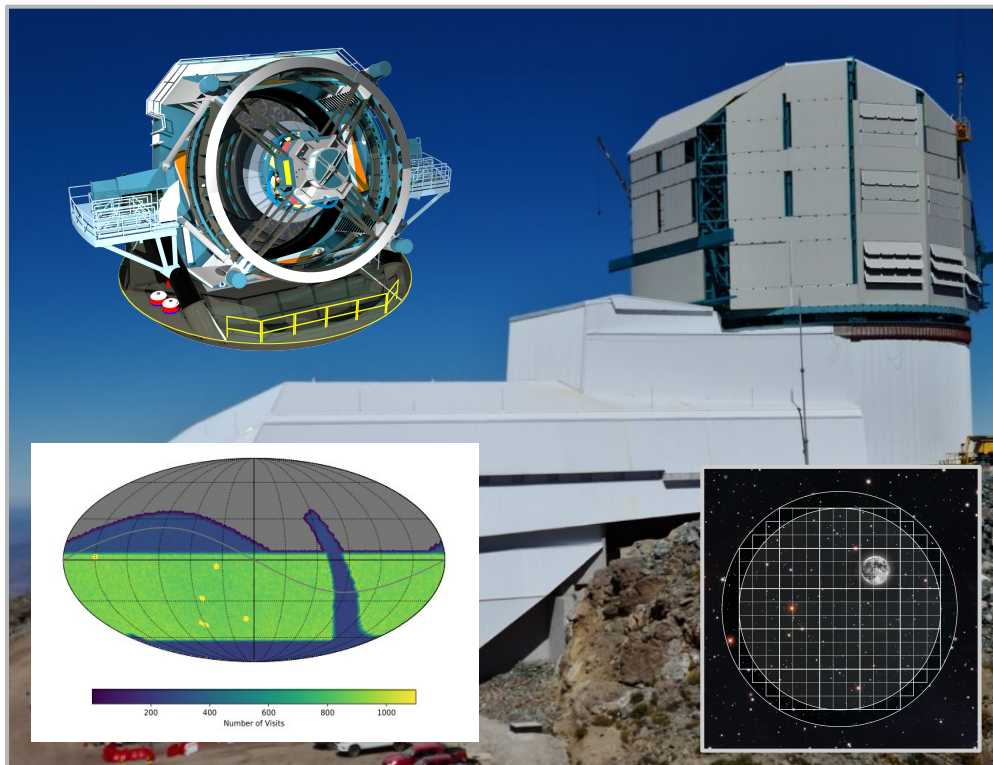
10 minutes -- presentation about the LSST PZ Roadmap

15 minutes -- Q&A about the roadmap

15 minutes -- discussion about evaluation criteria

10 minutes -- open time

The Vera C. Rubin Observatory



The Rubin Observatory, located in Chile, has an 8.4 meter diameter primary mirror and a 9.6 deg² field-of-view camera with six filters, *ugrizy*.

Once complete, Rubin Observatory will execute the **Legacy Survey of Space and Time (LSST)**.

The 10-year southern sky survey will make major advances in four core science areas:

1. Probing dark energy and dark matter
2. Taking an inventory of the solar system
3. Exploring the transient optical sky
4. Mapping the Milky Way

The LSST will cover $\sim\frac{1}{3}$ of the sky each night, detect billions of stars and galaxies, and millions of transients, variables, and moving objects -- a data set of unprecedented volume and complexity.

LSST Data Release (DR) Data Products

Data Release Contents

Images: processed visit images, coadded images

Catalogs: Sources/Objects, DIASources/DIAObjects, ForcedSources.

Source measurements such as PSF, shape, size, flux

Object characterization such as color, model fits, variability, photo-z

Data Release Timescale

DR1 includes data from the first 0.5 years.

DR2 includes data from the first year.

DR3 includes data from the first two years.

...

DR11 include data from the full 10-year survey.

It could take ~0.5 years to process and validate each DR.

Expect DR1 at 1 year, DR2 at 1.5 years, etc.

Definitions

Source = a detection in an image

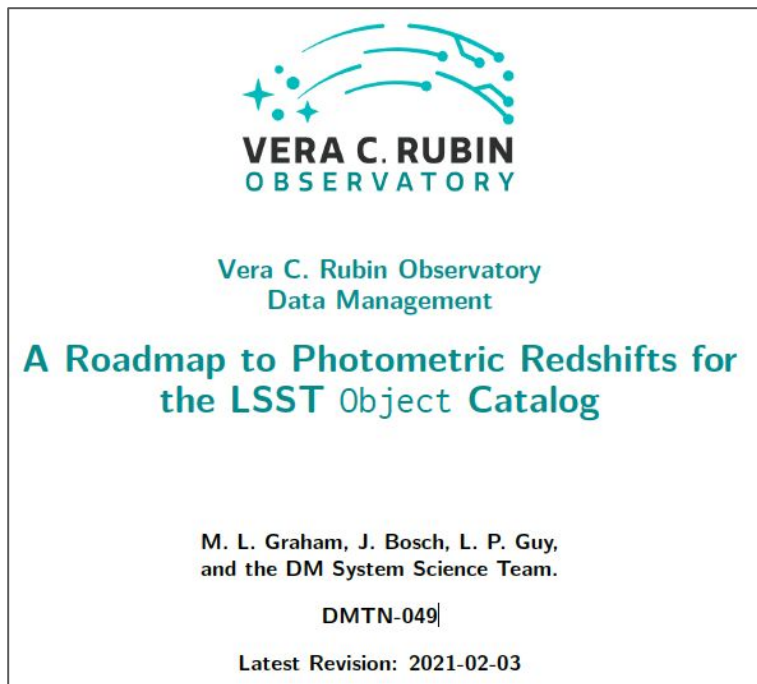
Object = Sources associated by their location/coordinates

DIA = Difference Image Analysis

ForcedSource = photometric measurements regardless of whether Source detected

PSF = point spread function

Roadmap Overview.



Roadmap Contents (*DMTN = Data Management Tech Note*)

This is a living document that will be updated over time.

Section 1 - Introduction

Section 2 - Roadmap Timeline

Section 3 - Technical Considerations from the DMS

Section 4 - Proposed Evaluation Criteria (*draft*)

Section 5 - Letters of Recommendation (*draft call*)

Section 6 - Photo-z Validation Cooperative (*draft*)

Appendix A - Examples of implementation/validation

Appendix B - Data Products Related to LSST PZ

Appendix C - Example Use-Cases for LSST PZ

full: ls.st/dmtn-049

summary: ls.st/clo4682

S.1. Why do we need a Roadmap?



The Data Management System (DMS) is required to compute a photometric redshift for all detected Objects (ls.st/lse-61).

The [science community](#) has a wealth of expertise in generating photo-z and will be the primary users of this data product.

Thus, DM will involve the [science community](#) in selecting, implementing, and validating one (or more) PZ estimator(s).

This is a complicated process that requires a lot of communication and coordination; the roadmap will be our framework.

[The science community is everyone interested in LSST photo-z, including members of the Science Collaborations and members of International Programs proposing in-kind contributions related to LSST photo-z.](#)

S.2. Timeline of Milestones (DM Actions)



2021-02/04: Host LSST PZ Virtual Forum Set 1.

2021-05-01: Ingest community input about evaluation criteria for PZ estimators, update Section 4 of DMTN-049, and issue a formal call for “Letters of Recommendation”

2021-05/08: Host LSST PZ Virtual Forum Set 2.

2021-09-30: Accept “Letters of Recommendation” advocating for certain estimators or types of estimators.

Before commissioning: generate a shortlist of community-vetted estimators and update Section 5 of DMTN-049.

During commissioning: facilitate **community*** participation in a “Photo-z Validation Cooperative” to validate shortlisted estimators.

*data rights are needed to participate in commissioning

S.3. Technical Considerations from the DMS

Scalability -- must be able to scale to ~37B Objects (i.e., millisec/object)

Inputs/Outputs -- catalog-level inputs only; outputs made consistent with DPDD (ls.st/lse-163)

Language -- C++/Python easy to adapt; others should be adaptable to run from a Python harness

External Data Sets -- must exist by commissioning; must be public and formatted for the estimator

Estimator Training -- requires domain-specific knowledge beyond the scope of DM; training must be provided by community teams prior to data release (who will be supported by DM)

Storage - volume of input/output must conform to budget

Compute -- should have a low-memory-footprint; compute budget to be revised as dev. progresses

S.4. Proposed Evaluation Criteria

Community input on these proposed evaluation criteria is solicited by DM.

1. **Scientific Utility** -- meet the basic science needs for communities which would/could not generate custom photo-z; serve a wide variety of science use-cases; enable early science
2. **Type of Estimator** -- template fits preferred over machine learning to enable scientific use-cases that rely on additional derived galaxy properties
3. **Data Products** -- posterior distribution function and point estimates; reliable uncertainties or flags to help novice users avoid misinterpreting results
4. **Scientific Performance** -- e.g., point-estimate accuracy of ~10% would meet science needs
5. **User Experience** -- broad community adoption, publicly accessible documentation
6. **Technical Considerations** -- meet the DMS specifications

S.5. Letters of Recommendation (*draft call*)

May 1 2021: DM updates the evaluation criteria based on community input, call for LoR released.

Sep 30 2021: Deadline for Letters of Recommendation.

The LoR are a formal chance for the science community to state their scientific needs regarding LSST Object catalog photo-z (to state to DM, and to share with the science community).

The goal is for DM to create a shortlist of several community-vetted PZ estimators to coalesce the joint DM-community efforts during the “Photo-z Validation Cooperative” during commissioning.

Letters of Recommendation:

- should be short (1-3 pages) and not constitute an onerous task
- should address the evaluation criteria, and the technical considerations if possible
- could focus on many, a few, or just one estimator, type of estimator, or science use-case
- could be qualitative or quantitative in their recommendation
- will be openly shared in the Rubin community

S.6. Photo-z Validation Cooperative (*draft*)

The science community has a considerable wealth of expertise in generating photo-z catalogs, and are best suited to scientifically validate the LSST Object catalog photo-z.

DM will facilitate a “Photo-z Validation Cooperative” by providing infrastructure and technical support for community teams to collaboratively assemble training sets and validation metrics, apply them to commissioning data, and document and share the results.

Exactly how this will work and on what timescale remains TBD.

See Section 6.1 of DMTN-049 for a brief overview of the potential commissioning data that will be relevant to photo-z validation.

How do we communicate and work together?

Many teams within the Science Collaborations and International Programs proposing in-kind contributions are already working on LSST photometric redshifts.

The goal of this roadmap is to *let all that development progress* and at the same time *provide a transparent process for DM's activities* and a *consistent means of communication for everyone*.

Communication Modes

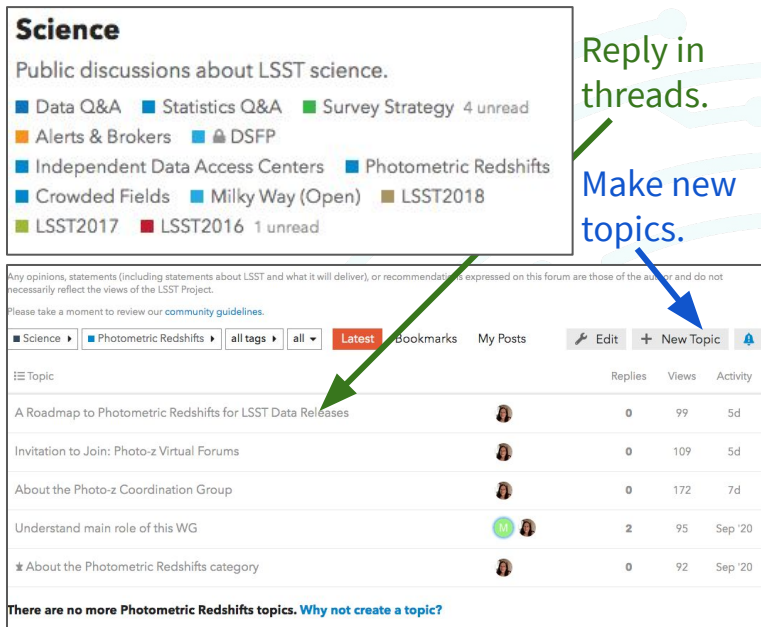
Attend these forums and participate in the discussions.

Photo-z Coordination Group in [Community.lsst.org](https://community.lsst.org) →

Emails to Melissa are OK too (mlg3k@uw.edu).

[Community.lsst.org](https://community.lsst.org)

“Photo-z Coordination Group” is the Photometric Redshifts category.



Science
Public discussions about LSST science.

- Data Q&A
- Statistics Q&A
- Survey Strategy 4 unread
- Alerts & Brokers
- DSFP
- Independent Data Access Centers
- Photometric Redshifts
- Crowded Fields
- Milky Way (Open)
- LSST2018
- LSST2017
- LSST2016 1 unread

Any opinions, statements (including statements about LSST and what it will deliver), or recommendations expressed on this forum are those of the author and do not necessarily reflect the views of the LSST Project.
Please take a moment to review our [community guidelines](#).

■ Science ▶ ■ Photometric Redshifts ▶ all tags ▶ all ▶ Latest Bookmarks My Posts Edit + New Topic

Topic	Replies	Views	Activity
A Roadmap to Photometric Redshifts for LSST Data Releases	0	99	5d
Invitation to Join: Photo-z Virtual Forums	0	109	5d
About the Photo-z Coordination Group	0	172	7d
Understand main role of this WG	2	95	Sep '20
★ About the Photometric Redshifts category	0	92	Sep '20

There are no more Photometric Redshifts topics. Why not create a topic?

Discussion Session

- ~15 minutes: questions about the roadmap
- ~15 minutes: discussion about evaluation criteria (next slide)
- ~10 minutes: open Q&A or discussion

Type in the Zoom chat, or unmute to speak (please introduce yourself!).

Questions later? Raise it as a new topic in Community.lsst.org
“Photometric Redshifts” category (<https://community.lsst.org/c/sci/photoz/42>).

S.4. Proposed Evaluation Criteria

Community input on these proposed evaluation criteria is solicited by DM.

1. **Scientific Utility** -- meet the basic science needs for communities which would/could not generate custom photo-z; serve a wide variety of science use-cases; enable early science
2. **Type of Estimator** -- template fits preferred over machine learning to enable scientific use-cases that rely on additional derived galaxy properties
3. **Data Products** -- posterior distribution function and point estimates; reliable uncertainties or flags to help novice users avoid misinterpreting results
4. **Scientific Performance** -- e.g., point-estimate accuracy of ~10% would meet science needs
5. **User Experience** -- broad community adoption, publicly accessible documentation
6. **Technical Considerations** -- meet the DMS specifications

Thank you for joining us today.

