



LARGE SYNOPTIC SURVEY TELESCOPE

# Large Synoptic Survey Telescope (LSST) Lossy Compression Working Group Charge

**Mario Juric, Robert Gruendl**

**LDM-582**

**Latest Revision: 2017-10-23**

This LSST document has been approved as a Content-Controlled Document by the LSST DM Change Control Board. If this document is changed or superseded, the new document will retain the Handle designation shown above.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 LSST digital archive and not printed versions. Additional information may be found in the corresponding DM RFC.

## **Abstract**

This is the charge for the Lossy Compression Working Group to be convened in September 2017.



## Change Record

Version	Date	Description	Owner name
---------	------	-------------	------------

*Document source location:* <https://github.com/lsst/LDM-582>



## Contents

<b>1 Scope</b>	<b>1</b>
<b>2 Background</b>	<b>1</b>
<b>3 Responsibilities</b>	<b>1</b>
<b>4 Specific Tasks</b>	<b>2</b>
4.1 Acceptable Loss . . . . .	2
4.2 Benchmarks . . . . .	2
4.3 Recommendations . . . . .	3
<b>5 Membership</b>	<b>3</b>
<b>6 Reporting</b>	<b>3</b>
<b>7 References</b>	<b>3</b>

# Lossy Compression Working Group Charge

## 1 Scope

This Working Group is to evaluate and recommend options for lossy compression algorithms to be used on LSST images in a way that continues to satisfy LSST science use cases. All major types of images shall be considered as candidates for compression, including raw data, Processed Visit Images, co-adds, and templates. This working group is to convene immediately and finish its investigation by 2017 November 30.

## 2 Background

This group has emerged as the response to RFC-325 that recognized that user experience will be unacceptably impacted by the long latency required to access the LSST data from tape media. Unfortunately, preliminary analysis indicated that retaining all processed images on disk would be too costly and therefore not feasible, unless lossy compression is applied. The same analysis indicated that storing all raw data on disk (w/o lossy compression) is feasible.

The LSST has traditionally avoided lossy compression for any of its image data products (including the large co-added images as well as templates retained for each data release). Anecdotal experience from DES and other surveys indicates that lossy compression can be applied, without loss of scientific fidelity. If this is the case, the reduced disk space needs may enable us to retain on low-latency media more data that we otherwise would (rather than regenerate or pull from tape. This group has been convened to study the problem and report on the results.

The working group should rely as much as possible on prior art found in the literature, and prefer applications of off-the-shelf solution rather than developing custom LSST-specific compression tools.

## 3 Responsibilities

The Working Group (WG) has the following responsibilities:

- Define criteria for “science-usable” lossy-compressed processed images across all LSST image types,
- Collect compression algorithm candidates (preferring existing in “off-the-shelf” tools/libraries),
- Evaluate their compression ratios (at “science-usable” quality),
- Evaluate constraints on processing that usage of compression may impose (e.g., avoidance of repeated re-compressions),
- Quantify the savings from application of lossy compression, in the context of the LSST Sizing Model ([LDM-144]),
- Make recommendations on which image types to lossy-compress, the algorithms to apply, and the description of processing constraints these would impose.

It should be noted that it is **not** intended that this group specifically investigate or recommend file formats (e.g. FITS, HDF5) excepting the utility of methods that are available to provide compression within a given format.

## 4 Specific Tasks

### 4.1 Acceptable Loss

The SRD outlines requirements for LSST. Those requirements should be met (or the impact understood) if lossy compression were implemented. Those requirements need to be considered in order to define what is considered an “acceptable loss.”

### 4.2 Benchmarks

The Working Group will obtain benchmarks based on existing astronomical data in order to provide a measure of lossy compression algorithms’: speed, compression factor, and fidelity. Where fidelity means a systematic comparison of image and catalog products with and without lossy compression applied.

## 4.3 Recommendations

The Working Group will provide a recommendation on the suitability of the implementation of a lossy compression algorithm of LSST products with a figure of merit that can be applied with the LSST Sizing Model ([LDM-144]).

## 5 Membership

Membership of roughly four people is optimal and should include persons familiar with weak-lensing and difference imaging concerns. The proposed membership is:

- Robert Gruendl (NCSA; **Chair**),
- Paul Price (Princeton),
- Bob Armstrong (Princeton),
- Krzysztof Findeisen (UW; replacing John Parejko),
- Sophie Reed (Princeton),
- Eric Morganson (DES/NCSA; observer)
- Ben Emmons (EPO Tucson; observer)

## 6 Reporting

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

## 7 References

- [1] **[LDM-144]**, Freemon, M., Pietrowicz, S., Alt, J., 2016, *Site Specific Infrastructure Estimation Model*, LDM-144, URL <https://ls.st/LDM-144>