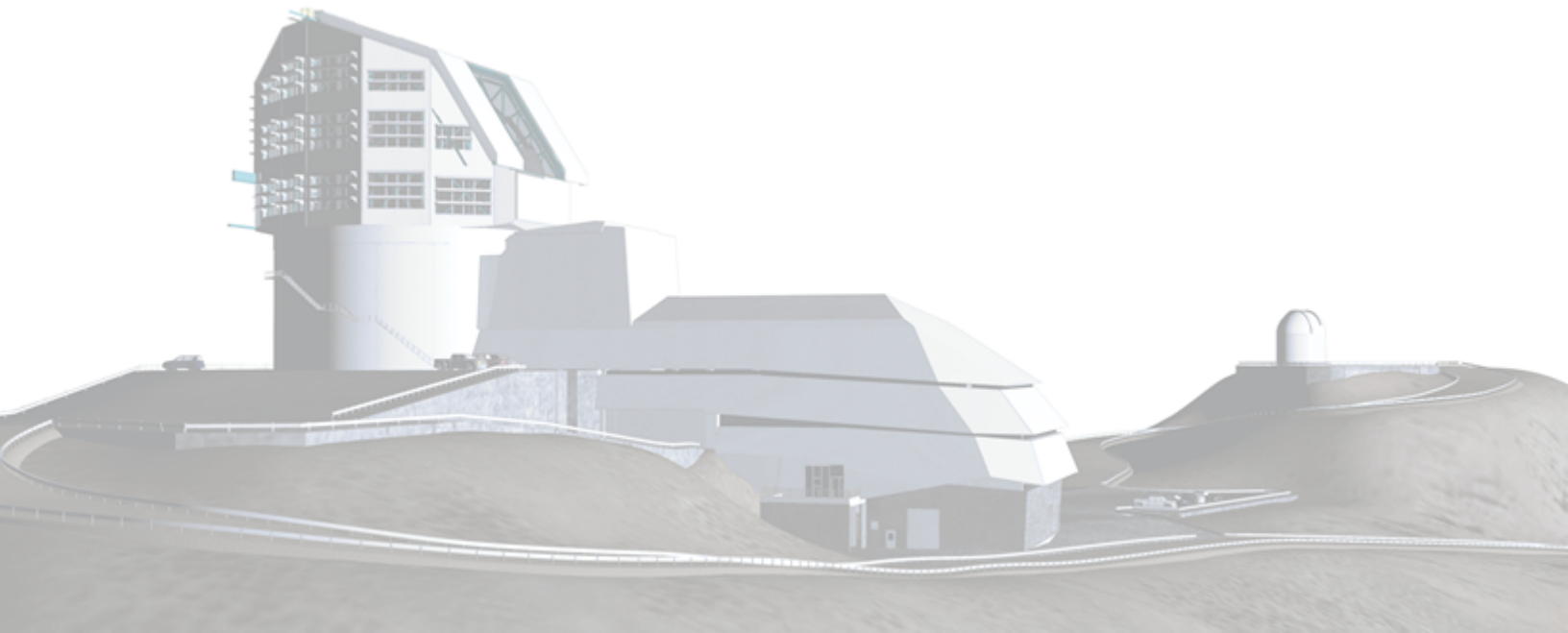

Vera C. Rubin Observatory
Data Management

Options for Alert Packets

M. L. Graham, L. P. Guy, E. C. Bellm, and the Data Management System Science Team

DMTN-248

Latest Revision:

Abstract

A review of the scientific impact of options for alert packets, with recommendations.

Change Record

Version	Date	Description	Owner name
1	2023-01-23	First version.	Melissa Graham

Document source location: `https://github.com/lsst-dm/dmtn-248`

Contents

Options for Alert Packets

1 Introduction

Alerts will play a major role in the scientific impact of the LSST. They are the only data product that is both world public (exempt from any proprietary period) and publicly accessible, the latter meaning that they are distributed to community brokers and not solely available via Data Access Centers (DACs) RDO-013.

The alert stream and its packets' contents are described in the Data Products Definitions Document (DPDD; LSE-163). The motivation behind their design is to rapidly distribute all of the LSST data about a difference-image detection that a broker (or their users) would need to assess, classify, and prioritize the alert for follow-up observations within minutes.

This document evaluates options that were considered for the alert stream and alert packet contents to maximize science while minimizing alert packet size and thus bandwidth usage. These considerations were most salient prior to the selection of the community alert brokers [LDM-612]. Most options look at either reducing the packet size, or improving the scientific utility of packet contents.

Reduce packet size: The typical size of an alert packet is estimated to be about 82 KB, based on simulations of the alert contents specified in Section 3.5 of the DPDD (see also the Alerts Key Numbers document; DMTN-102). Reducing packet size was more of a concern in the earlier days of construction, in order to potentially enable alert distribution to additional brokers. However, even though the the decision to support seven full-stream brokers has been made, considerations of alert packet size are still relevant to, e.g., storage and bandwith usage.

Improve scientific utility: Other options below aim to improve the scientific utility of alert packet contents by providing additional or differently-formatted data.

This document also includes a brief discussion about the science impacts for the various ways of dealing with delayed alerts when, e.g., there are >40000 alerts per visit (Section 8).

2 Historical records

Proposal: Instead of including the history in every packet, brokers could retrieve it for only the fraction of the alerts for which it is needed.

Alert packets will include "*any* `DiaSource` *and* `DiaForcedSource` *records that exist, and difference image noise estimates where they do not, taken from the previous 12 months*" (Section 3.5.1 of the DPDD). The history would account for about 27 KB, or about 33% of the total alert packet size.

The motivation for including historical records in alert packets is to enable brokers to assess the *full* time-domain event in order to make a robust classification (or prioritization) for follow-up.

Brokers (and their users) would still need the list of the unique identifers for the `DiaSource` and `DiaForcedSource` records associated with the same `DiaObject` as the alert-triggering `DiaSource`, even if their full records were not included in the alert packet. Removing this list of unique identifiers is *not* being considered as part of this proposal for the following reasons:

1. The LSST Prompt Processing pipeline associates `DiaSources` into `DiaObjects`, and it would be challenging, time-consuming, and redundant to have brokers repeat this.
2. The association of `DiaSources` into `DiaObjects` is probabilistic, and in rare cases might change over time (e.g., strongly-lensed supernovae, or line-of-sight transient superpositions, where two distinct difference-image sources are blended/separated in poor/good seeing). In such cases the set of all `DiaSource` records which compose the history might change.

Brokers that save all past alerts in their own archives would be able to retrieve the full records for associated `DiaSources` locally (but not `DiaForcedSources`, as they would never have been in an alert packet). Brokers with authenticated RSP access would be able to query the PPDB to retrieve the full records for associated `DiaSources` and `DiaForcedSources`. Brokers that do not save all past alerts and/or do not have authenticated RSP access to the PPDB would have no historical information aside from the number of associated `DiaSources` (i.e., the number of past detections).

At best, the lack of historical records in alerts would add a layer of complication and a potential delay to the brokers' functionality, and inhibit the scientific assessment of time-domain events. At worst, it would be a severe risk to brokers' ability to process alerts.

Recommendation: Do not remove historial records, it imposes a risk to time-domain science.

3 Postage stamps

Proposal: Remove image stamps from the alert packets, in favor of one of the two options below, with which brokers could instead retrieve the postage stamps for only those alerts for which it is needed (e.g., unclassified time-domain events).

The postage stamps are, at minimum, 6×6 arcseconds (30×30 pixels) and contain flux, variance, and mask extensions for both the template and difference image, plus a header of metadata (DPDD). The stamps would account for about 18 KB, or about 20% of the total alert packet size.

The motivation for including postage stamps in the alert packets is to enable brokers to use image-based machine learning tools (e.g., custom real/bogus scores; host+supernova classifiers), 2-dimenstional flux distribution (e.g., trailed sources; cometary outbursts), or environmental context (e.g., field crowdedness) to classify or prioritze alerts. Postage stamps also enable visual inspection of the images by science users, which can be valuable in some contexts.

Two options to replace the inclusion of postage stamps in every alert are:

1. **Image cutout service**. Brokers with authenticated RSP access could use the image cutout service to create and retrieve stamps.
2. **URL to the postage stamp**. A URL could be put into the alert instead of the postage stamp, which points to an automated public cutout service or the pre-made postage stamps on a server.

Regarding the first option, the 80 hour embargo on all new images and difference images renders the option to use an image cutout service non-viable (regardless of whether it is via

the RSP or a public service).

The second option has the advantage that brokers (or downstream brokers) do not need to save the stamps and can retrieve them at any time, but a disadvantage that stamp retrieval could add a delay to brokers' alert processing and analysis.

As a side note, the second option is unlikely to cause a bottleneck when multiple brokers attempt to retrieve large numbers of stamps simulataneously. For example, if 5 brokers retrieve 500 stamp sets per visit, that's 2500 stamp sets every 35 seconds, and at 18 KB per stamp set the data rate would be 10.5 Mbps (5% of one full alert stream). However, the second option would require Rubin Observatory to create and maintain a public server of pre-made postage stamps, which is not currently planned for development.

Recommendation: Do not remove postage stamps, it imposes a risk to time-domain science.

4 Large multi-resolution postage stamps

Proposal: Provide larger multi-resolution postage stamps in the alert packets by binning the pixels with a bin size that increases towards the edges of the images, such that several arcminutes can be included.

The motivation for this proposal is to enable time-domain science that relies on the rapid association of low-redshift transients with their wide-area host galaxies, using algorithms such as "DELIGHT: Deep Learning Identification of Galaxy Hosts of Transients using Multiresolution Images" (Förster et al. 2022).

The default 6×6 arcsecond postage stamp size is large enough to contain the angular diameter (30 kpc) of a Milky Way-like galaxy at a redshift of 0.5, which will be a fairly typical supernova host galaxy in the LSST data set. The fraction of all alerts that will be low-redshift transients with wide-area host galaxies will be <1% [DMTN-102].

Host association will already be included in the alert packet as part of the `DiaObject`, as described in DMTN-151 and in Table 3 of the DPDD, including potential nearby extended and low-redshift galaxies (see columns nearbyExtObj and nearbyLowzGal).

It is expected that this type of host-galaxy identification could be done by accessing the template images (which are not subject to the 80 hour embargo) via the image cutout service, via which large-area cutouts could be made.

Although the context provided by larger postage stamps have also been shown to better identify `DiaSources` that are glints (rotating debris in low earth orbit that appears as a row of point sources; e.g., Karpov & Peloton 2022), ways to identify and flag `DiaSources` as potential glints in the AP pipeline are underway, instead of leaving this identification solely to brokers.

Furthermore, although the multi-resolution aspect of the proposal aims to provide a larger area without significantly adding to the size of the alert packet, it is currently unclear whether serialization of the multi-resolution cutouts could be achieved without requiring custom display tools and/or increasing overheads from image headers.

Recommendation: At this time, do not increase the size of postage stamps, but potentially reconsider in the future.

5 Packet compression

Proposal: Compress alert packets to reduce their size.

The application of gzip compression could further reduce the size of a full alert to 65 KB (80%). This might help to avoid alert distribution bottlenecks, lower the cost for brokers' storage needs, and potentially enable additional full streams in the future – all of which could positively impact time-domain science with the LSST.

At this point we don't believe we need to sacrifice latency in order to achieve bandwidth savings, but we may revisit this tradeoff as we begin testing alert distribution at scale in the USDF.

Recommendation: Do not compress alert packets, it imposes a risk to rapid time-domain science.

6 Multiple packet formats

Proposal: Allow brokers to specify whether their stream should have histories or postage stamps removed.

The main motivation here is not scientific, but to help reduce brokers' processing costs spent removing unneeded information from alert packets.

There are a few drawbacks to this proposal, but they are mostly technical:

1. New scope, as the current plan is for Prompt Processing to create a uniform alert packet, and this would create additional provenance work for Rubin (i.e., tracking which alerts were sent to which broker).
2. Non-identical packets might cause bookkeeping issues for downstream brokers subscribing to multiple brokers.

These drawbacks could impose a risk to the particular science goals of brokers using full-sized or non-identical alerts. Further, during the broker proposal process the approved broker teams all expressed a desire to receive the complete alert packets. **Recommendation**: Do not offer multiple packet formats.

7 Pre-filtered streams

Proposal: Allow brokers to request a pre-filtered alert stream, for example to only include alerts in a certain sky region or which meet other criteria (e.g., brightness limits, number of past detections).

As with the multiple packet formats, the main motivation here is not scientific, but to help reduce brokers' processing costs spent receiving unwanted alerts, or to enable Rubin Observatory to distribute alerts to more brokers.

For example, due to the exponential relationship between the number of variable stars and their variability amplitude, and also that of volume and distance modulus, an apparent mag-

nitude limit that is 1 mag brighter than the nominal 5-sigma `DiaSource` detection limit could reduce the alert stream data rate by 50%.

However, all seven full-stream brokers have committed to receiving the full stream, and the term downstream broker refers to brokers which will ingest a filtered stream of alerts from one or more full-stream brokers.

Recommendation: There is no need for Rubin Observatory to provide pre-filtered streams to brokers.

8 When and how to distribute delayed alerts

The Data Management System (DMS) is required to support the distribution of at least 40,000 alerts per single standard visit. Furthermore, for visits producing ≤40,000 alerts, no more than 1% of them may fail to have at least 98% of their alerts distributed within 60 seconds of image readout (based on LSR-REQ-0101 in LSE-29; OSS-REQ-0193 in LSE-30; and DMS-REQ-0392 and -0393 in LSE-61).

Furthermore, alert distribution should degrade gracefully beyond that limit, meaning that visits resulting in an excess of 40,000 of alerts should not cause any downtime for the Data Management System (DMS; LSE-30, LSE-61). It is also a requirement that all alerts be stored in an archival database and be available for retrieval (OSS-REQ-0185 in LSE-30).

This leaves the open question of what, from a science perspective, is the optimal way of dealing with delayed alerts. (Aspects of the technical implementation of a graceful degradation, such as distributing delayed alerts and alert archive storage access, are out of scope for this document).

There are three main options:

1. **Next-opportunity distribution via the alert stream**. Distribute delayed alerts as soon as possible. There are plenty of science goals that do not absolutely require alert distribution in 1 minute, and so distributing delayed alerts via the stream would still enable plenty of science. The brokers might prefer to have delayed alerts clearly flagged to properly process them (e.g., some filtering and processing done by brokers might only

be appropriate for alerts delivered within a given latency).

2. **Next-morning distribution via the alert stream**. Collect all delayed alerts during the night and then release them (perhaps on a new topic) all at once in the morning, after survey operations have ended for the night. From a science perspective this is not as useful as next-opportunity distribution, but if it is preferred for technical reasons it would enable more science than the option below.

3. **Do not distribute delayed alerts; send directly to archive**. There is no scientific merit in *not* distributing delayed alerts. Four futher drawbacks include: the alert archive update timescale is 24 hours (significantly slower than next-morning distribution); the alert database would only be accessible by brokers with authenticated RSP access; alerts might only be able to be queried by alert ID; and bulk download capabilities might be limited.

Recommendation: Flag and distribute delayed alerts as soon as possible to enable time-domain science. Alerts should also have a processing timestamp added so that brokers can gauge delay timescales.

A References

[DMTN-151], et al., M.L.G., 2021, *Host Galaxy Association for DIAObjects*, DMTN-151, URL `https://dmtn-151.lsst.io/`,
Vera C. Rubin Observatory Data Management Technical Note

[LDM-612], Bellm, E., Blum, R., Graham, M., et al., 2020, *Plans and Policies for LSST Alert Distribution*, LDM-612, URL `https://ldm-612.lsst.io/`,
Vera C. Rubin Observatory Data Management Controlled Document

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

[LSE-29], Claver, C.F., The LSST Systems Engineering Integrated Project Team, 2017, *LSST System Requirements (LSR)*, LSE-29, URL `https://ls.st/LSE-29`

[LSE-30], Claver, C.F., The LSST Systems Engineering Integrated Project Team, 2018, *Observatory System Specifications (OSS)*, LSE-30, URL https://ls.st/LSE-30

[LSE-61], Dubois-Felsmann, G., Jenness, T., 2019, *Data Management System (DMS) Requirements*, LSE-61, URL https://lse-61.lsst.io/,
Vera C. Rubin Observatory

[DMTN-102], Graham, M.L., Bellm, E., Guy, L., et al., 2020, *LSST Alerts: Key Numbers*, DMTN-102, URL https://dmtn-102.lsst.io/,
Vera C. Rubin Observatory Data Management Technical Note

[LSE-163], Jurić, M., Axelrod, T., Becker, A., et al., 2022, *Data Products Definition Document*, LSE-163, URL https://lse-163.lsst.io/,
Vera C. Rubin Observatory

B Acronyms

Acronym	Description
AP	Alert Production
DM	Data Management
DMS	Data Management Subsystem
DMS-REQ	Data Management System Requirements prefix
DMTN	DM Technical Note
DPDD	Data Product Definition Document
KB	KiloByte
LDM	LSST Data Management (Document Handle)
LSE	LSST Systems Engineering (Document Handle)
LSR	LSST System Requirements; LSE-29
LSST	Legacy Survey of Space and Time (formerly Large Synoptic Survey Telescope)
OSS	Observatory System Specifications; LSE-30
PPDB	Prompt Products DataBase
RDO	Rubin Directors Office
RSP	Rubin Science Platform
URL	Universal Resource Locator
USDF	United States Data Facility