# Report from "Building the Infrastructure for Time-Domain Alert Science in the LSST Era"

Thomas Matheson (NOAO) & Rachel Street (LCO)

## 1 Background

Time-domain astronomy has experienced decades of consistent growth. From the first systematic attempts at time-domain observational programs using photographic plates through the current high-cadence, wide-field programs using digital detectors, the field has enjoyed a tremendous increase in the number of objects detected, as well as the quality of the data, including the number of epochs and range of wavelengths observed. There are dozens of optical time-domain projects running now and their output already stresses an observational infrastructure that was mainly built for static objects.

In just a few short years, the Large Synoptic Survey Telescope (LSST) will come online and generate $\sim 10^7$ alerts per night from objects that have changed in brightness (or moved). To be ready for that deluge of alerts, astronomers must prepare now. Part of that preparation involves developing the infrastructure astronomers will need to take full advantage of the time-domain science LSST will enable. To that end, the National Optical Astronomy Observatory (NOAO) and the Las Cumbres Observatory organized a workshop on 'Building the Infrastructure for Time-Domain Alert Science in the LSST Era' that was held from May 22-25, 2017 in Tucson, Arizona. The overall goal was to bring together the diverse set of stakeholders who will be necessary to build a full-scale time-domain system and identify what needs to be done and how we might get there.

The following document summarizes the presentations and discussions at the meeting. The presentations themselves are available at the meeting web site and we refer to them throughout for the details. Here is the URL for the presentations:

`https://www.noao.edu/meetings/lsst-tds/agenda.php`

## 2 Sources of Alerts

This session included a presentation about the expected alert streams from the Zwicky Transient Facility (ZTF) and LSST. The expectation from ZTF is $\sim 10^6$ alerts per night, with another factor of 10 for LSST. The ZTF public alerts should begin in the first half of 2018 and thus provide an excellent test bed for time-domain infrastructure systems that plan to scale to the LSST level. There was also a presentation on the possible formats for the alerts from LSST as well as a presentation on how the Dark Energy Survey processes their time-domain alerts from deep-drilling fields.

One of the subjects that came up in the discussion of these topics was how to understand the provenance of the data. The potential evolution of the algorithms used, the code used to implement the algorithms, the underlying language and associated compiler, and even the hardware on which all of this runs could alter how alerts are evaluated. Keeping track of this is difficult and it isn't clear the depth to which the provenance has to be maintained.

Another topic discussed was the time scale necessary for follow up. Although there will be millions of alerts per night, not all require immediate or even same-day follow up. It is important to recognize that infrastructure to facilitate time-domain astronomy can and should take advantage of the range of time scales associated with transient and variable objects.

# 3 Observatories and Their Policies

One of the goals of the meeting was to assess the current and near-term capabilities and policies of astronomical observatories that could play a role in follow-up strategies in the era of LSST. Several observatory directors and other staff attended the meeting, representing Las Cumbres, NOAO, CTIO, SOAR, Gemini, Keck, Apache Point, Chandra, and MMT. All presentations are available at the meeting web site.

Each observatory representative was asked to provide a brief summary of capabilities. This included the current and future instrumentation suite (where 'future' meant instruments in development, not far-term plans without specification). They also described the current time-allocation process. This covered time sharing with other facilities as well as their target-of-opportunity policy. Given the potentially dramatic increase in time-domain follow-up requests, they were also asked to speculate on possible modifications to time allocation. The observatory representatives also described their scheduling modes, especially those related to time-domain astronomy. While many observatories are still scheduled classically, there are interrupt modes for time-domain observations. In addition, some facilities offer a queue-based system. Some of these queue-based systems are also robotically operated. There was much discussion by the classically scheduled facilities about how to adopt queue or queue-like systems to accommodate future time-domain demand.

The observatories were asked to address issues related to data and software. The issue of who has access to data and when will become ever more complicated when there is a high demand for time-domain follow up. This is also tied to how observatories award and schedule time, so clear policies will need to be developed to avoid conflicts. The software systems necessary to manage a time-domain system cover many aspects of observatory operations. Data reduction software, especially automated pipelines, would be extremely useful for time-domain astronomy, but few of the observatories had such capabilities. In addition, there is a desire for observing control system interfaces that would allow for automatic (or at least rapid) configuration of instruments and scheduling/execution of observations. The observatories at the meeting covered the range from fully automatic configuration to no current method to incorporate this kind of interrupt without direct human involvement.

The details from each observatory are available in the individual presentations on the meeting web site. Two points to highlight from those presentations are the NSF Mid-Scale Innovations Program (MSIP) opportunity with Las Cumbres Observatory and the plan being developed by US federal observatories and Las Cumbres to build a time-domain follow-up system.

The opportunity with Las Cumbres arises from an MSIP award from the NSF to Las Cumbres that provides public access to their facilities. It includes ∼1200 hours of 1m time and ∼200 hours of 2m time per semester for eight semesters. Time allocation is being run through the regular NOAO process. This began in 2017AB (Apr 1 – Nov 30, 2017) with time already awarded through NOAO. There will be an opportunity for longer-term key projects starting in 2018. (Some new key projects have already begun with public participation.)

Along with this public time, Las Cumbres is also working with the NSF-funded AURA facilities (NOAO, SOAR, Gemini) to develop a comprehensive follow-up system that includes all these facilities. The system would include software development that would enable brokering of alerts (tools for selecting alerts from event streams) as well as scheduling observations at all the included facilities. Reduced data would then be returned to users. The goal is an end-to-end system, initially with limited modes. For details on both of these concepts, see the presentations from Bob Blum and Todd Boroson on the meeting web site.

During the discussion, both in this session and at the end of the meeting, one of the issues that arose was how knowledge about observations could be disseminated. In other words, how would anyone know if an object had already been observed and what the results of that observation might be? It would be useful to the community in general if observatories could publish what they are observing. There are issues related to proprietary data, but eliminating duplicate observations in an era when targets are plentiful but observing resources are scarce would be useful. One could envision a system similar to a concept named the 'future footprint server' by Michael Wood-Vasey. All surveys could publish a plan for their nightly operations, as well as up-to-date information about where they are pointed currently. That way, other surveys or targets observations could follow behind (or proceed ahead) of any given survey. Follow-up systems could also publish where they are pointed to ensure observations are not duplicated. The system could include the resources available at follow-up facilities so that users could make informed decisions about where to activate observations.

# 4 Brokers

The term 'broker' can have a wide range of meanings and capabilities associated with it. In the time-domain astronomy context, it generally refers to a software system to process time-domain alerts. Any time-domain broker should be careful to define exactly what the system does to avoid potential misunderstanding. In this session, two broker systems were described. The first was NEOfixer, a system in use by the Catalina Sky Survey to process their alerts on moving objects. It actually combines some functions of a broker with those of a target/observation manager (described later). The second broker was the ANTARES system, a broker designed to scale to the LSST alert stream. It will annotate, characterize, rank, and distribute value-added alerts. See the presentations about these two brokers on the meeting web site for details.

In addition, the ANTARES team provided a demonstration of the brokers current capabilities. They also ran a hack session with Jupyter notebooks. Attendees could either write simple filters in a mini version of ANTARES to select interesting objects from a simulated stream or they could choose to use a machine-learning tool to explore characterization of light curves.

# 5 NOAO Data Lab

There was a presentation describing the NOAO Data Lab and a hack session allowing attendees to explore the capabilities it provides. Initial development has not focused on time-domain astronomy, but many of the catalog-based tools will be applicable to databases of time-domain objects.

The overall goal of the NOAO Data Lab (datalab.noao.edu) is to enable efficient exploration and analysis of large-scale data sets with particular emphasis on DECam-based catalogs and images. The Data Lab aims to:

- Connect users to high-value catalogs from NOAO and external sources (e.g. SDSS, GAIA) and NOAO-based images linked to catalog objects
- Enable users to discover the data that they need for their science
- Allow users to develop intuition through interaction with selected catalog and image sets
- Allow users to automate their analysis to aid discovery in large data sets

The Data Lab provides services to enable as much work as possible close to the data, Current Data Lab services include tools for sky exploration, user authentication, large catalog queries, image queries and cutouts, query result storage, file transfer, and scientific analysis through Jupyter Notebooks. Services under active development include a job-managed compute service, a data publication service, and a multi-purpose catalog cross-match service. The services are available through both server-side and locally running clients, including web-based interfaces, command-line tools, Python interfaces, APIs, and legacy tools such as

TOPCAT. These services provide a flexible way for users to interact with large catalog and pixel data sets, including for analyses of time series. At the time-domain workshop, Robert Nikutta and Stephanie Juneau demonstrated example uses of Data Lab services for specific science cases.

Data sets available through the Data Lab currently include DECaLS DR3, SMASH DR1, and several reference data sets including GAIA DR1 and select tables from SDSS DR13. Future data sets will include additional DECaLS releases, DES DR1, and the NOAO Source Catalog (NSC), which David Nidever described at the workshop. The NSC is an aperture photometry-based catalog of all public DECam images, Mosaic-3 images, and images from Bok 90Prime taken as part of the DESI Targeting Survey. It currently contains 20 billion individual measurements of 2 billion unique objects, and will make it possible to search for variable sources over most of the sky. Procedures and tools for working with time series from the NSC are under current development.

# 6 Target/Observation Managers

For science projects to select follow-up targets of interest to them and to coordinate appropriate observations, a critical link in the workflow will be software tools, generically described as Target and Observation Managers or TOMs. These software systems provide users with an interface (such as a web-based GUI or an API) that enables ingestion of filtered alert streams, scheduling and management of observations, storage and exploration of data, and communication and collaboration tools. Further capabilities can be included, depending on the science case and desires of the science users. In this session, the science drivers and essential functionality required by TOM systems were identified. A number of existing systems that fulfill this role (or aspects of it) were presented as examples, including those running the LCO microlensing, supernovae and Near-Earth Object projects as well as several from a range of community projects including PESSTO, PTF, ExoFOP, and AGN Agent. A sample of the wide array of science results derived from TOM systems were presented, and there was general agreement regarding the benefits they bring, particularly in the overwhelmingly target-rich era that astronomy is entering. While different TOM systems all require some elements which are heavily customized to the science they perform, it was striking that a core set of functions are required by almost all of the current TOMs. A proposal was therefore outlined by T. Boroson, to develop a software "toolkit" that will enable astronomers to easily build systems to suit their projects, without re-inventing the wheel. This will be a joint program by LCO and NOAO, with significant input from the community.

# 7 Breakout Sessions

An afternoon session of the meeting was run as an ‘unconference.’ Based on suggestions from the meeting attendees, five topics were chosen for the breakout sessions. Details from each are available on the meeting web site with summaries here.

## 7.1 Classification Challenge

A critical problem for large-scale time-domain surveys is how to recognize what an astronomical alert might be when little information is available. There are several projects underway trying to address this issue, most notably, the Photometric LSST Astronomical Time-series Classification Challenge (https://plasticcblog.wordpress.com/). Some issues brought up in the breakout session included the lack of high-quality, multi-band data for almost all kinds of variable objects and the lack of attention to early-time identification.

## 7.2 Observing Modes/Time Allocation

This breakout session looked at ways observatories could change observing modes or time-allocation methods to improve time-domain follow up. They reported on two different kinds of issues that will arise. The first is cultural, with many astronomers being resistant to change, especially if they haven’t been involved in time-domain astronomy. Changes will have to be gradual to accommodate this resistance. The second category of issues is structural, including added cost, scheduling, and accommodation of static-sky astronomers.

## 7.3 Protocols/APIs for Requesting Observations and Receiving Data/Scheduling

This breakout session focused mainly on requesting and scheduling observations, with many examples using the current Las Cumbres API. It is a challenge to manage observation requests over a large number of observing resources. This will only get more complicated with the addition of heterogeneous resources compared to the current Las Cumbres network. Some of the points stressed at the breakout included the ability of the system to provide useful feedback to the user and that the system requires solid documentation so that the users can understand the capabilities and limitations of the system.

## 7.4 Spectroscopy Reduction

Everyone agreed that modern spectroscopy reduction tools are needed, but the pathway was not obvious. There is an effort beginning within ASTROPY to address this, but the need for basic tools, especially for automatic reduction, exists now.

### 7.5 Tools for TOMs

This session focused on what tools are necessary for TOMs. There was a wide range of capabilities discussed, including plotting tools, scheduling tools, and methods to visualize available resources. There was also a recognition that some functionality of TOMs is useful to observatory staff (as opposed to the science users) to facilitate the functional tasks involved in time-domain observation management and execution. Some TOM capability to visualize the scheduling at an observatory helps the science users to understand where the pressure is for observing time and how to optimize their follow-up strategy. There was some discussion about the boundaries between brokers and TOMs and how TOMs might interact with each other. It is important for any software tool to have clearly defined boundaries so that all users will know what they can and cannot do, and what other tools may be necessary to bridge between them. There was also discussion about how TOMs can serve citizen science and how they might interact with data archives. Finally, a minimal TOM was described.

## 8 Summary

The meeting was attended by a diverse group, including astronomers and computer scientists, observatory directors and graduate students, and software developers and instrument builders. All of these people and more will be needed to develop, build, and operate the infrastructure necessary to get the most out of the time-domain science that LSST will enable. Many attendees had not been exposed to the challenges faced by the other groups, so seeing the all the cross-disciplinary issues laid out was helpful to all.

Participants made connections that had not yet existed and laid the groundwork for further development. The AURA/Las Cumbres effort to build a follow-up system grew stronger out of this workshop. The great advantage provided by TOMs was directly apparent to all who attended. Overall, there is still an enormous amount of work yet to do, but, on many smaller scales, the initial steps are happening now.