Tools to enable efficient and effective large-scale observing programs in the era of 'the alert firehose'

R.A. Street, M. Bowman, E. Saunders, T. Boroson
Las Cumbres Observatory

Modern astronomical surveys promise an unprecedented wealth of discoveries, if we can extract the most valuable information from a flood of alert data, and conduct effective and timely follow-up observations. Experience from existing programs has proven databases of targets and observations to be extremely valuable in conducting efficient follow-up - future projects handling a vast increase in potential targets will find such tools to be indispensable. However, astronomy's existing target and observation manager systems are hard to implement, and will not scale to meet the alert rate of future surveys such as ZTF or LSST. We propose that astronomy takes advantage of modern database technology to develop general-purpose, publicly-available tools to enable future follow-up programs.

1 Introduction

Current and near-future astronomical surveys promise dramatic increases in the sky area, frequency and depth to which we monitor the night's sky. Their scientific potential is vast, delivering a target sample which could transform many areas of astronomy, and one large enough to discover the "rarest of the rare", revealing new phenomena. These facilities will analyze their data in near-real-time and issue alerts to the community of new discoveries. The survey data by itself is, in many cases, insufficient to fully understand the detections, as discussed in detail by Najita et al. (2016). Follow-up observations must be triggered in response to the alerts to properly characterize a range of phenomena.

Historically, astronomers would travel to telescope facilities to conduct observations of selected targets within an allocation of specific nights. While this approach works well for some programs, it presents several important disadvantages to time domain astronomy in particular. The science often demands observations at times and at intervals dictated by physical phenomena rather than by a Time Allocation Committee - for example, at the time of a planetary transit or when an asteroid reaches its closest approach distance. Transient targets cannot be predicted in advance, demanding that observers respond rapidly to new target alerts or unpredictable behavior. Targets of all types often need to be monitored repeatedly over the course of days, months or even years. Observing facilities can, and in many cases are, responding to this by gradually adopting a number of policies to facilitate time domain science, such as queue-mode scheduling, target-of-opportunity overrides, remote- or robotic observing modes and automated observation-request systems. We recommend that these measures be made a priority wherever practically possible, for both ground- and space-based facilities at all wavelengths.

Regardless of the science goal, a follow-up program requires a sequence of distinct components, represented in Figure 1. A new target is usually discovered by a survey, which issues an alert. This is aggregated with catalog information by a broker in an initial attempt to classify the phenomena, at which point the object may be selected for further observations, according to the scientific criteria of one or more projects. Since observing facilities are limited, these projects often compete for time on the same telescopes to conduct these observations. Finally, analysis of the resulting data products completes the cycle by providing more information, which is used to determine the future analysis and/or observation strategy for that object.

Some science goals require more rapid responses than others. Observing supernovae at early times, or finding optical counterparts to gravitational waves and gamma-ray bursts, requires that observers react within minutes or hours of the first alert, then monitor and evolve their strategy with the behavior of their targets. As transient targets are unpredictable, the most effective way to answer these requirements is to automate the process as much as possible.

However, the sheer size of future target catalogs and high survey alert rates presents a serious challenge. For example, the Zwicky Transient Factory (ZTF) is expected to produce ~1 million alerts per night (Smith, 2014), whereas the Large Synoptic Survey Telescope (LSST) could issue up to 10 million alerts per night at the start of the survey (LSST Science Book, Ridgway et al. 2014). In this paradigm, simply selecting targets of interest will be computationally intensive, to say nothing of keeping track of the observations required in each case as they evolve over time and the associated workload.

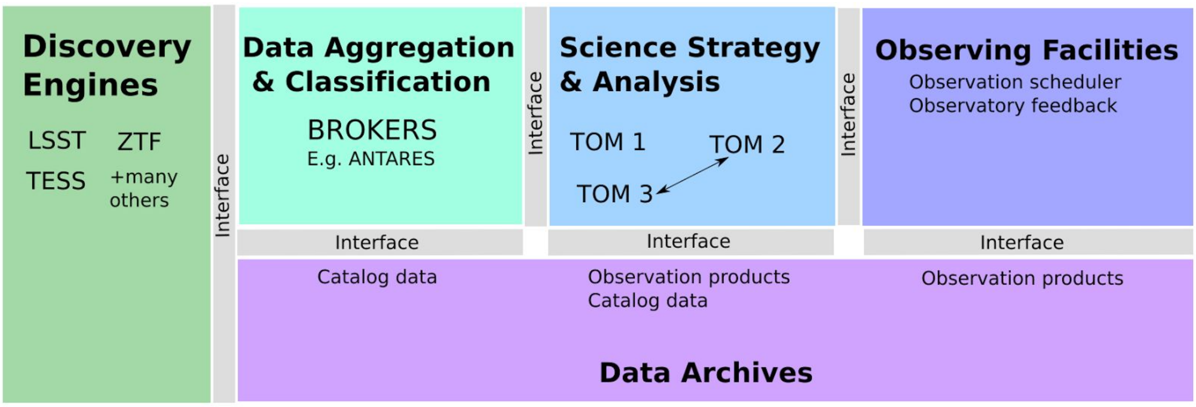

Figure 1: The stages, from left to right, of the astronomical discovery and follow-up process, showing the role of TOM systems. It is envisioned that different projects will have their own TOMs, and may choose to share data between them.

An example serves to put the challenge in context: currently there are ~700 alerts of new supernovae per year. The Global Supernova Project, lead by A. Howell, brings together over 150 astronomers from institutions worldwide and their combined follow-up resources include time allocations on 12 ground-based facilities and the Swift spacecraft. This project finds it obtains adequate follow-up observations for ~80-100 targets per year (Howell, priv. comm.), or ~11-14% of the total. LSST is expected to detect ~100,000 supernovae per year for 10 years. Each and every target must be monitored as it evolves over time, to ensure appropriate observations are obtained at each stage, with their relative priority being continually re-assessed in comparison to new alerts.

Keeping track of targets and follow-up observations is therefore no easy task even with current surveys. In the LSST era, the alert rate will *far* outstrip current capacity - yet follow-up observations are essential to achieve the science goals. Without technological development, we will be unable to capitalize on the scientific potential of the new surveys.

2 Target and Observation Manager Systems (TOMs)

Teams responding to alerts from earlier surveys have, of necessity, developed software tools which act as "v1.0" for these technologies. Examples range from the Palomar Transient Factory Marshall system to the Exoplanet Follow-up Observing Program run by IPAC[1], but in particularly we draw from experience with the RoboNet Project[2] (exoplanet discovery via microlensing, PI: Street, Tsapras), SNEx[3] (supernova follow-up, PI: Howell) and NEOExchange[4] (Near Earth Object follow-up, PI: Lister).

Though written independently to achieve very different goals, these v1.0 tools show a number of striking similarities. Each project found it necessary to build a database to keep track of their targets and observations and to develop associated software to:

a) Select new targets to observe based on survey alerts or input catalogs
b) Specify and submit science-appropriate observation requests to telescope facilities
c) Retrieve and analyze the resulting follow-up data
d) Share their results with collaborators
e) Plan new observations in the light of the new data.

We refer to this category of software as *Target and Observation Manager systems (TOMs)*.

Fig. 1 shows how TOM systems fit within the follow-up process, essentially providing the scientific 'intelligence' to control and organize a large-scale and/or rapid follow-up program. A TOM system could apply a human-driven target selection and data analysis, a fully algorithmic process, or a hybrid of the two. Realistically though, even projects that prefer manual analysis will still greatly benefit from the database tools we outline below.

The resources that astronomers will call on to receive survey alerts, to conduct their observations and to retrieve their data, will often be common-user facilities that serve large communities - for example, the ANTARES broker [Saha et al. 2014 Proc. SPIE 9149, arXiv:1409.0056], the Minor Planet Center[5], telescopes accessible through NOAO, and data archives such as IPAC[6]. Many of these resources already have or are developing online interfaces which users can interact with programmatically, and many science projects will need tools to do so, for the reasons discussed above. Rather than have every project waste

[1] https://exofop.ipac.caltech.edu/
[2] http://robonet.lco.global/
[3] http://supernova.exchange/public/
[4] https://lco.global/neoexchange/
[5] http://www.minorplanetcenter.net/
[6] https://www.ipac.caltech.edu/

time and money reinventing these tools, it would be better to make available a general-purpose toolkit that can interface to these facilities.

While some elements of the TOM systems are necessary regardless of the science goals, many aspects need to be customized. For example, astronomers need to be free to experiment with new target selection criteria or data analysis algorithms, change their observing strategy or modify how they display information on their targets. They most commonly work in collaboration, often with geographically-distributed team members, all of whom need to access the data held in the project's TOM. This argues for a highly flexible 'toolkit' approach that enables astronomers to easily build the system they need with an online user interface.

Though we draw on the experience cultivated with v1.0 TOMs, the existing systems are not sufficient to meet the challenges of the next decade. All of the existing systems were written to address a specific science case, by astronomers rather than software engineers, and are heavily customized. Furthermore, they suffer similar limitations in the number and rate of alerts they were designed to handle: they will not scale to the size needed for future surveys (they would be unusably slow). At the same time, industry has made rapid progress in database performance and data visualization tools thanks to their widespread applications. We recommend that astronomy take advantage of this, and call on professional software engineers to design a tools appropriate to these tasks.

Functions of a TOM

Regardless the science goal behind it, TOM systems need the following components, shown in Fig 2.

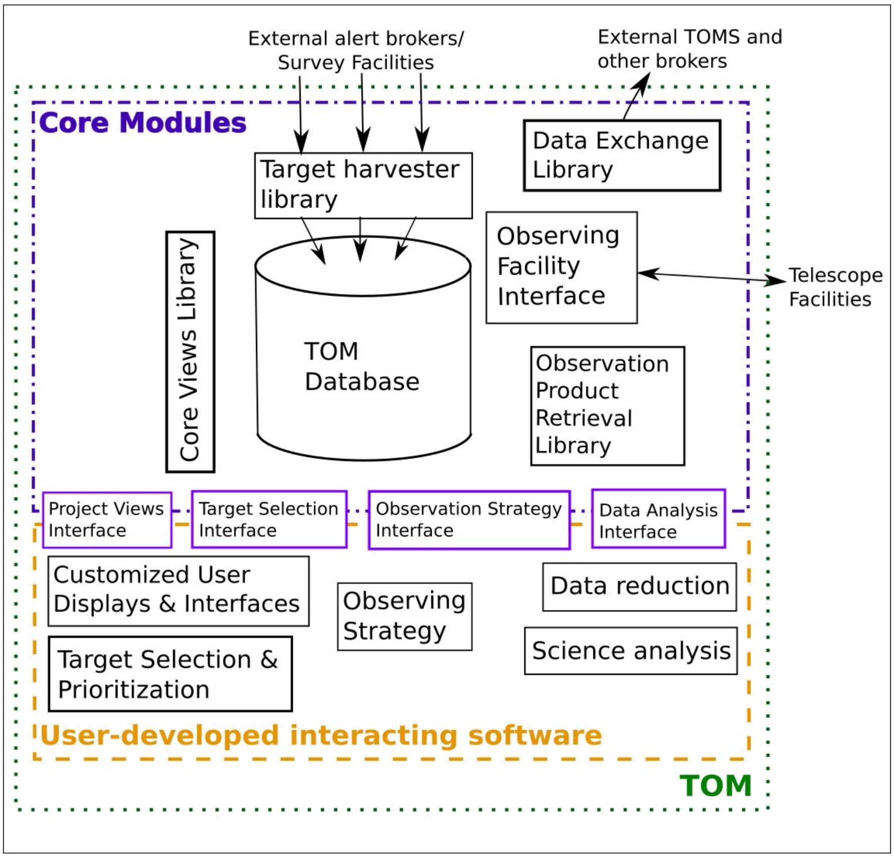

Figure 2: Schematic diagram showing the different components of a TOM system needed to run an astronomical observing program.

A central database ensures the necessary information on targets and observations is stored and is fully searchable. Targets are ingested by means of software that harvests information from sources of interest (labeled above as the **target harvester library**), and a range of display tools provide users with the means to view, edit, plot and discuss the data. Some display tools will be common to all projects **(core views library)**, such as user account management, while users will need tools to enable them to build displays customized to their science **(project views interface)**. Ideally, these tools should be accessible through online portals, to facilitate geographically-separated teams.

Many packages and data pipelines already exist to perform astronomical analysis and astronomers should be free to use whatever they prefer. However, these packages are likely to need to access information in the TOM, and it's important that the TOM be able to receive the results of this analysis, since it is likely to impact observing strategy, target selection etc. Therefore the TOM must provide interfaces which can be called by this science-specific software.

TOM systems can do more than simply record when observations are taken. Telescope facilities are increasingly able to accept requests for observations submitted remotely or robotically. TOM systems therefore need to support interfacing between manual, remotely-operated and robotic common-user observing facilities **(observing facility interface)**. Experience from the LCOGT Network has shown that it is also highly desirable for the facility to return information regarding both the status of the telescope (e.g. is it open, or closed for bad weather?) and the specific observation (e.g. has my request been scheduled?).

Having obtained data, rapid access to the data products is essential for some science and desirable for others. This argues for encouraging facilities to make their data available through online and ideally programmably-accessible archives, which may require investment in the development of in-house data pipelines. The TOM system should incorporate tools to interface with the online archives for common-user facilities **(observation product retrieval library)** and also catalog archives such as IPAC.

Data Sharing

As noted above, existing projects already have more targets than they can observe, and the number of targets is set to increase far more rapidly than resources to follow them. It is therefore likely we will "saturate" the capacity of our community-accessible telescopes, particularly if competing teams with similar science goals request similar observations for the same targets. We can potentially reduce the congestion and improve efficiency overall if astronomers are encouraged to share data wherever possible. TOM systems can be designed to encourage this by providing tools to make this easy.

We acknowledge that there are sometimes legitimate arguments to retain data, particularly for early-career researchers who need to lead high-impact papers, and different research areas have different common practices. We recommend exploring ways to provide incentives to share data.

References:

LSST Science Collaborations and LSST Project 2009, LSST Science Book,
Version 2.0, arXiv:0912.0201

Najita, J. et al. (2016), Maximizing Science in the Era of LSST: A Community-Based Study of Needed US Capabilities, arXiv:161001661.

Ridgway, S. et al. (2014), ApJ, 796, 53

Smith, R. et al. (2014), Proceedings of the SPIE, 9147, id. 914779.