

ANTARES: A Prototype Transient Broker System

Abhijit Saha,^a Thomas Matheson,^a Richard Snodgrass,^b John Kececioglu,^b
Gautham Narayan,^a Robert Seaman,^a Tim Jenness,^c Tim Axelrod^b

^aNOAO, 950 N Cherry Avenue, Tucson, AZ 85719 USA;

^bUniversity of Arizona, Tucson, AZ 85721 USA;

^cDepartment of Astronomy, Cornell University, Ithaca, NY, 14853, USA

ABSTRACT

The Arizona-NOAO Temporal Analysis and Response to Events System (ANTARES) is a joint project of the National Optical Astronomy Observatory and the Department of Computer Science at the University of Arizona. The goal is to build the software infrastructure necessary to process and filter alerts produced by time-domain surveys, with the ultimate source of such alerts being the Large Synoptic Survey Telescope (LSST). The ANTARES broker will add value to alerts by annotating them with information from external sources such as previous surveys from across the electromagnetic spectrum. In addition, the temporal history of annotated alerts will provide further annotation for analysis. These alerts will go through a cascade of filters to select interesting candidates. For the prototype, ‘interesting’ is defined as the rarest or most unusual alert, but future systems will accommodate multiple filtering goals. The system is designed to be flexible, allowing users to access the stream at multiple points throughout the process, and to insert custom filters where necessary. We describe the basic architecture of ANTARES and the principles that will guide development and implementation.

Keywords: Time-domain alert analysis, Event broker, LSST, VOEvent, transient response, big-data

1. THE PROBLEM

1.1 Background

An increasing number of large area astronomy surveys that probe time-variable phenomena are producing candidates that need rapid response follow-up (whether monitoring ongoing changes, or spectroscopic observations to probe physical characteristics). Salient examples are: the Lick Observatory Supernova Search^{1*}, the Catalina Real-Time Transient Survey^{2†}, the Panoramic Survey Telescope & Rapid Response System (Pan-STARRS)^{3‡}, the Palomar Transient Factory (PTF and iPTF)^{4§}, and the La Silla-Quest Variability Survey^{5¶}. These surveys are discovering new transient phenomena that are already taxing the available follow-up capacity of telescope facilities world-wide. These projects have developed tools to filter their discoveries to focus on events of interest to their research teams (e.g., supernovae, gamma-ray burst events, and so on), which can gather ancillary information about their ‘alerts’ from external catalogs, and use the available information to classify the sources associated with their alerts. A leading example of this is SkyAlert^{6||}, a system that has solved many of the astronomical issues associated with adding value to alerts. SkyAlert enables users to create filters on alerts, including ancillary information on these alerts, in order to find relevant events. The PTF system also employs tools to identify interesting alerts.⁷ The scale of time-domain alert generation, though, is quickly increasing. The Zwicky Transient Facility⁸ (ZTF) will have more than 6 times the field-of-view

Further author information: (Send correspondence to A. Saha)

A.Saha: E-mail: saha@noao.edu, Telephone: +1 520 318 8288

*http://astro.berkeley.edu/bait/public_html/kait.html

†<http://crts.caltech.edu/>

‡<http://pan-starrs.ifa.hawaii.edu/public/>

§<http://ptf.caltech.edu/iptf/>

¶<http://hep.yale.edu/lasillaquest>

||<http://skyalert.org/>

Observatory Operations: Strategies, Processes, and Systems V, edited by Alison B. Peck,
Chris R. Benn, Robert L. Seaman, Proc. of SPIE Vol. 9080, 908000 © 2014 SPIE
CCC code: 0277-786X/14/\$18 · doi: 10.1117/12.2056988

of PTF, while time domain surveys with DECam on the Blanco telescope benefit not only from the 3 deg² field-of-view, but the depth attainable with a 4m-class facility. Moreover, transients are generated across the electromagnetic spectrum, from radio facilities such as LOFAR^{9**} to high-energy space-based observatories such as Fermi^{10††}, making the overall problem that much more complex.

On the horizon for beginning operation in 2021, is the Large Synoptic Survey Telescope.^{11,12} With its 10 deg² field-of-view and ~6m collecting area, the transient detection rate leaps by orders of magnitude. LSST will detect (with 5 σ significance) 10³ – 10⁴ alerts per image, or 10⁶ – 10⁷ per night. By going fainter, and covering an area of over 18,000 square degrees, this 10 year long survey will probe an unprecedented volume of space with a time cadence that can identify variability on time scales from tens of minutes to years. During the survey operation, the LSST facility will issue alerts of celestial transient events using VOAEvent and other IVOA protocols.^{13,14}

1.2 Variable Event Alerts

An alert is a notice triggered when an image shows that something is significantly different with respect to an archive image. A variable star may trigger an alert each and every time it is imaged: a supernova in a distant galaxy will trigger repeatedly against an archive image from before it erupted. A moving object will be seen typically over erstwhile blank sky, and move to a different location at subsequent epochs, triggering alerts at all these different locations at each respective epoch. While most alerts will be yet another incremental data point for a celestial object already known to vary, among these multitudes will lurk objects the likes of which have never (or extremely rarely) been seen before.

Prompted by the potential importance of early detection of short lived transient phenomena, the LSST survey will issue alerts with a latency of only about a minute. An alert contains essential information like the location on the sky, the passband in which the variability was detected, whether the change was in brightness or in position, the magnitude of the change, and the epoch of that particular trigger. It may or may not (depending on the facility that issues it) contain ancillary information about whether it is a recurrent alert, a history of all alerts at that location in the sky, or other similar ancillary information.

A good fraction of alerts from LSST will be known variable stars or moving objects,^{15,16} but hidden among them will be rare and interesting objects that have relatively short lifetimes. Only with additional follow-up will these objects reveal their nature. These could range from short-lived phases of stellar evolution such as the final helium flash^{17,18} to superluminous supernovae¹⁹ to electromagnetic counterparts of LIGO detections.^{20,21} Beyond these rare, but known or predicted, objects lies the great discovery space that awaits LSST. The superluminous supernovae were essentially unknown fifteen years ago and the discovery of dark energy was certainly surprising. Over its life, LSST will generate more than a billion alerts and some will be completely unknown and unanticipated objects. Without the ability to rapidly sort through millions of alerts each night and winnow them down to a reasonable number that can be studied in detail, we will lose these rare and potentially extraordinarily interesting objects. The astronomical community is becoming more aware of the necessity of such a tool.²²

1.3 Project Goals

We take note of (and are encouraged by) advanced methods for determining classification probabilities using small numbers of time-series measurements (for instance^{23–25}), using machine learning algorithms and techniques. They have been used to identify specific kinds of variables of interest to their respective investigations. Their successes notwithstanding, the alert rates that the LSST survey promises require us to look at the problem a little differently:

- The alert broker needs to handle alert data volumes at the rate that LSST is capable of generating, i.e. thousands of alerts per minute. They need to be processed with a latency that does not make them stale: i.e. since LSST will produce alerts with a latency of ~ 1 minute, the broker must process those alerts without introducing significantly larger delays to the alert stream.

** <http://www.transientskp.org/>

†† <http://fermi.gsfc.nasa.gov/>

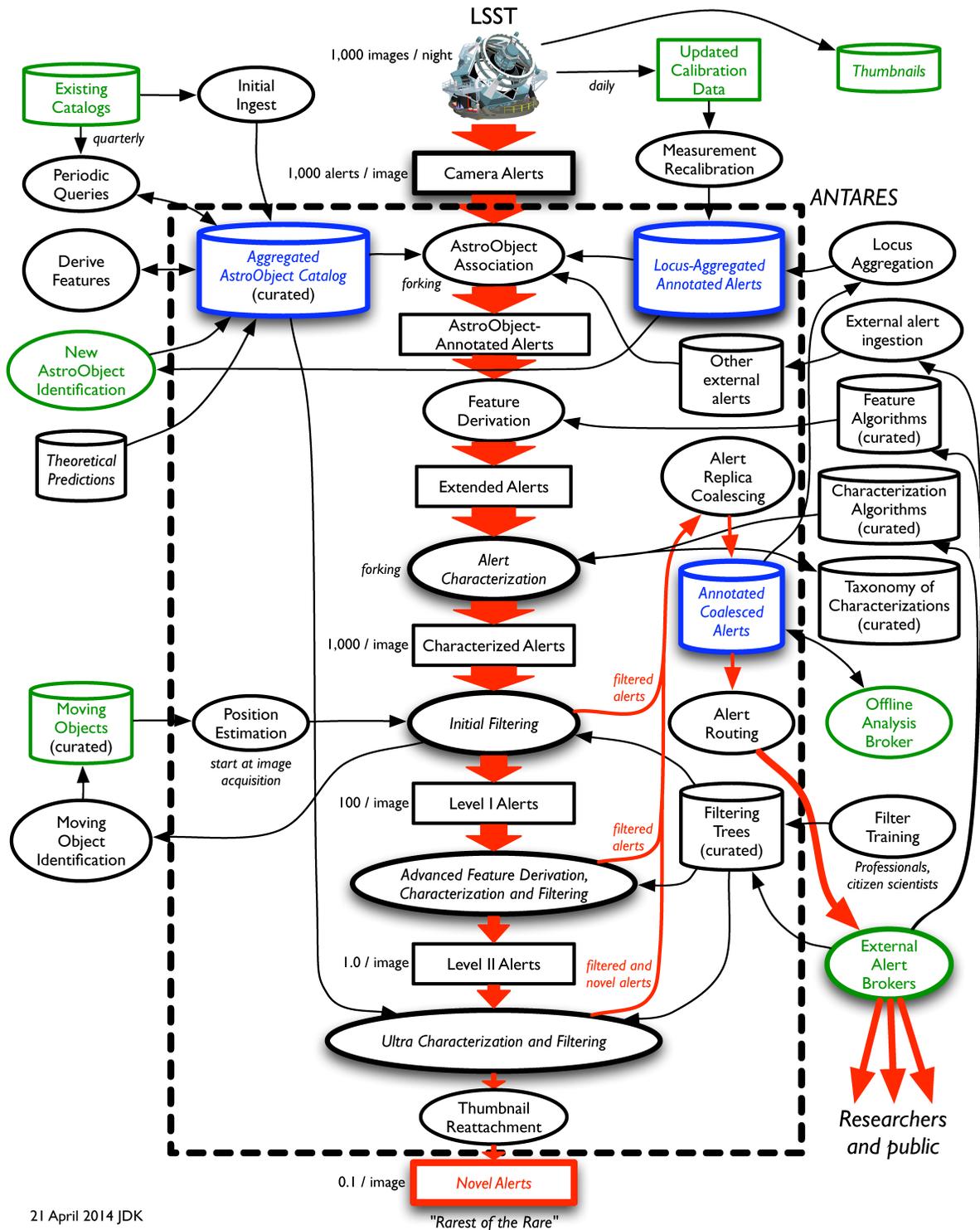


Figure 1. Basic architecture of the ANTARES system. The dashed box encompasses the processes that must keep up with the LSST frame-rate.

- The alert broker needs to serve generic needs: i.e. its design should not be limited to identifying specific pre-defined kinds of celestial sources.
- It should store and archive all alerts, and be able to append contextual information for any celestial source(s) associated with that alert. It should be able to integrate results from any followup investigation of sources.

Achieving these goals requires the combined efforts of astronomers with experience in time-domain astronomy, as well as computer scientists who can impose the design and methods necessary to achieve the necessary end-to-end speed and scalability for dealing with LSST scale data rates and volumes.

2. DESIGN AND ARCHITECTURE OF ANTARES

2.1 Proto-type design considerations

The knowledge we have about an alert, such as brightness, change in flux, Galactic coordinates, ecliptic coordinates, distance to nearest galaxy, etc., constitute features that can probabilistically characterize alerts. We re-iterate that this is a broad characterization, not a specific classification: the latter will have to come from software systems further downstream. Because of the time-scale of LSST exposures, with a new image every ~ 37 seconds, alerts must be processed rapidly to keep up with the data stream. Classification often requires more complex analysis and usually a more complete light curve.^{23,26}

For the prototype, we have selected the challenging problem of identifying the rarest of time domain phenomena: those that are least like things we know. Alerts that appear to come from more commonplace astronomical sources are diverted, but saved for further use. We will discuss later how ANTARES is structured so that it can be modified and applied to identify other kinds of phenomena, and thus become a generic tool. Identifying the ‘rarest of the rare’ leads us through the problem space that makes adapting to other needs relatively straight-forward.

2.2 Architecture and Data Path

Figure 1 illustrates the main components of the ANTARES architecture.^{27,28} The overall design principles are open source and open access. The software will be available for anyone to implement and our implementation will be community driven. The alert stream can be tapped at many points throughout the system. In Figure 1 alerts enter the system from the top center.

The first stages provide annotation that add contextual value to the alerts. Source association is a critical step to incorporate relevant astronomical knowledge for each alert. Catalogs of astronomical information, as well as the LSST source catalog will be the basis for this source association. Examples include the 2MASS All-Sky Data Release^{29‡‡}, the Chandra Source Catalog^{30*}, the NRAO VLA Sky Survey^{31†}, the Sloan Digital Sky Survey^{32‡}, the NASA Extragalactic Database^{33§}, and GAIA^{34¶}, among many others. These external catalogs are collated and initially ingested to produce the *Aggregated AstroObject Catalog*, shown near the top-left of Figure 1. This catalog will be updated from time to time with periodic queries to the external catalogs, and with new ‘AstroObjects’ from the episodic data releases of the LSST survey. Even the proximity to known sources can provide useful constraints.

A new alert is also tested for association with past alerts (from a database maintained by ANTARES, and shown in the figure as *Locus-Aggregated Annotated Alerts*) and additionally from any other available external alert data sources. The history of flux measurements, such as a light curve, will be valuable annotation. The *Locus-Aggregated Annotated Alerts* is meant to be an efficient database that can be updated regularly is an

‡‡<http://www.ipac.caltech.edu/2mass/releases/allsky/>

*<http://cxc.harvard.edu/csc/index.html>

†<http://www.cv.nrao.edu/nvss/>

‡<http://www.sdss.org/>

§<http://ned.ipac.caltech.edu/>

¶<http://sci.esa.int/science-e/www/area/index.cfm?fareaid=26>

essential element of the system. This will be a valuable astronomical resource on its own. As mentioned before, the SkyAlert system provides a similar annotation. The problem for the future is the scale of alerts and the resulting necessity of this efficient database being integrated into the system for brokering alerts.

A central notion in our procedure is that of *alert characterization*. This is a discriminant activity which uses the features to determine what ‘kind’ of alert we have. We distinguish this task from ‘classification’, in that characterization is necessarily uncertain and probabilistic, while classification is a more certain association with a known astrophysical type. Examples of broad characterizations include known variable star, extragalactic source, active galaxy, or likely moving object. These require looking at all of the features, and as such is a holistic analysis, as contrasted with feature derivation, which can be performed independently for each added feature. For example, a small change in magnitude might imply a stellar variable, but if it has not been detected before, and it is near a galaxy, it may be a supernova, but caught when the brightness is changing slowly.

For many alerts, there will only be a small number of features available for characterization, especially for an initial detection. If there are not enough features for discrimination by filtering, we can apply a probabilistic expectation of variability based on position on the sky and known distributions of variability.¹⁵ For a position, we can construct a variability probability density function and predict the likelihood of the alert as observed.

With more data, more features become available and more complex filtering algorithms can be used. ANTARES will then use multiple layers of filters to sort the alerts and find the rarest or most interesting among them (the focus of the prototype project). The filtering will be based on feature vectors, either directly supplied by the alert and associated contextual information, or derived therefrom. These features are then compared against the features from known time-variable phenomena, using a variety of methods. Alerts that are likely to come from ‘common-place’ phenomena are diverted away from the main processing stream. Each stage lets through fewer and progressively less commonly characterized alerts. These may then be re-characterized in feature vector space that is different or of higher dimension and filtered again. The filtering stages are meant to be ordered so that most efficient (decisions on most alerts in the least time) filters are staged first. More time consuming and in-depth probing is reserved for the later stages, where the alerts have already been winnowed to a smaller number. Experimentation will show us the most efficacious and efficient feature combinations and algorithms. The training of filters and algorithms will be aided using machine based experimentation with AMELIE (see below in section 2.3), and is an integral part of the development of the ANTARES system.

The **diverted alerts are not discarded**: in each filtering stage they are diverted from the main filtering stream but are still accessible to other filtering systems, including, potentially, copies of the ANTARES system itself that are tuned to other specific goals. Thus an *External Alert Broker* (shown in the bottom right of Figure 1) can utilize the value added material from ANTARES to filter according to alternative needs and priorities. Custom filters can be applied, allowing users to isolate exactly which of the alerts is of interest to them and thus address many different goals. These community-derived filtering algorithms will be applied in a multi-step process, allowing for better management of computational resources. By characterizing the alerts, the number of dimensions of feature space can be reduced. More complex filters can be applied to the smaller number of alerts after initial filtering stages.

An important design consideration throughout the architecture of ANTARES is the structured provision of community input. While ANTARES will provide the overarching design of the alert analysis, it is the role of the astrophysical community to provide the specific algorithms used at various places along the filtering process.

2.3 AMELIE

The Arizona Machine-Experimentation Laboratory (AMELIE), Figure 2), provides a system for constructing and testing structural-causal models.³⁵ This essentially automates the scientific process and allows us to run experiments to test relationships among features, including relationships that have not yet been apparent. It can observe the operation of ANTARES and make it more efficient.

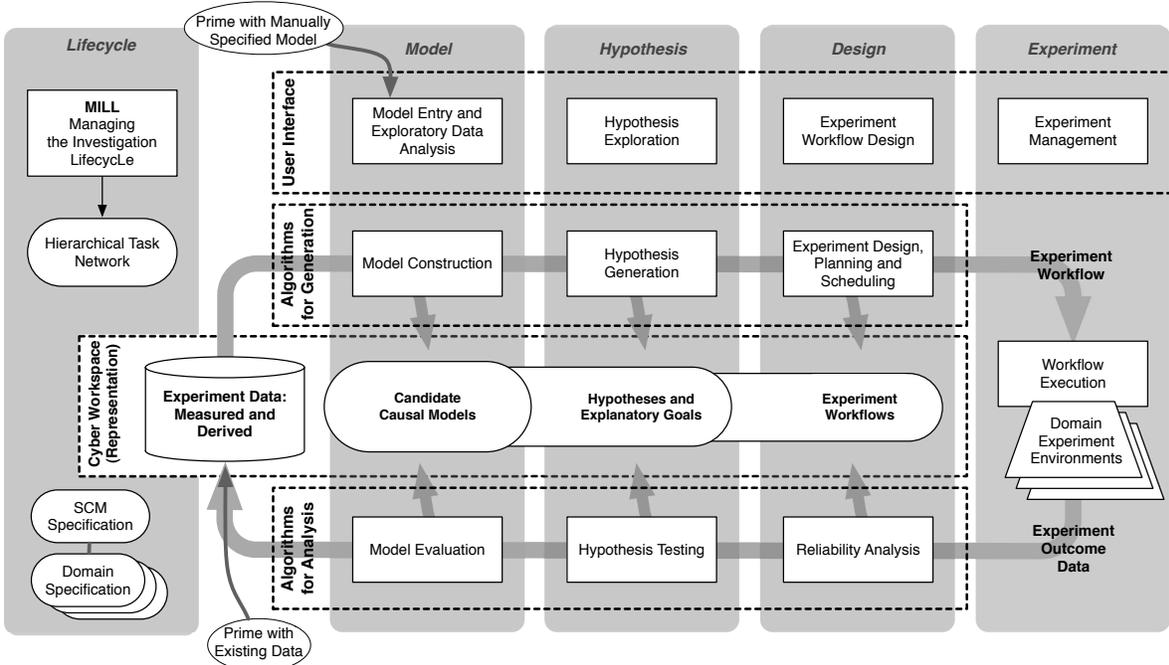


Figure 2. Basic architecture of AMELIE.

2.4 From Proto-type to a Generic Tuneable Broker

The goal for the prototype is to distinguish rare and unusual objects. Once it is operational, the next stage is to expand the scope to allow users to find any type of alert of interest to them. We foresee that there will be many stages of the ANTARES system itself, processing different data streams over different time scales. The overall alert ecosystem could accommodate multiple alert input streams and thus find a general way to serve the astronomical community's needs.

ACKNOWLEDGMENTS

We acknowledge the NSF INSPiRE grant (CISE AST-1344204, PI:Snodgrass) that supports this work. AS acknowledges many conversations with Kirk Borne, which seeded much of the thinking from which this project was born.

REFERENCES

- [1] Leaman, J., Li, W., Chornock, R., and Filippenko, A. V., "Nearby supernova rates from the Lick Observatory Supernova Search - I. The methods and data base," *Monthly Notices of the Royal Astronomical Society* **412**, 1419–1440 (2011).
- [2] Drake, A. J., Djorgovski, S. G., Mahabal, A., Beshore, E., Larson, S., Graham, M. J., Williams, R., Christensen, E., Catelan, M., Boattini, A., Gibbs, A., Hill, R., and Kowalski, R., "First Results from the Catalina Real-Time Transient Survey," *Astrophysical Journal* **696**, 870–884 (2009).
- [3] Kaiser, N., Aussel, H., Burke, B. E., Boesgaard, H., Chambers, K., Chun, M. R., Heasley, J. N., Hodapp, K.-W., Hunt, B., Jedicke, R., Jewitt, D., Kudritzki, R., Luppino, G. A., Maberry, M., Magnier, E., Monet, D. G., Onaka, P. M., Pickles, A. J., Rhoads, P. H. H., Simon, T., Szalay, A., Szapudi, I., Tholen, D. J., Tonry, J. L., Waterson, M., and Wick, J., "Pan-STARRS: A Large Synoptic Survey Telescope Array," in *[Survey and Other Telescope Technologies and Discoveries]*, Tyson, J. A. and Wolff, S., eds., *Proc. SPIE* **4836**, 154–164 (2002).

- [4] Law, N. M., Kulkarni, S. R., Dekany, R. G., Ofek, E. O., Quimby, R. M., Nugent, P. E., Surace, J., Grillmair, C. C., Bloom, J. S., Kasliwal, M. M., Bildsten, L., Brown, T., Cenko, S. B., Ciardi, D., Croner, E., Djorgovski, S. G., van Eyken, J., Filippenko, A. V., Fox, D. B., Gal-Yam, A., Hale, D., Hamam, N., Helou, G., Henning, J., Howell, D. A., Jacobsen, J., Laher, R., Mattingly, S., McKenna, D., Pickles, A., Poznanski, D., Rahmer, G., Rau, A., Rosing, W., Shara, M., Smith, R., Starr, D., Sullivan, M., Velur, V., Walters, R., and Zolkower, J., “The Palomar Transient Factory: System Overview, Performance, and First Results,” *Publications of the Astronomical Society of the Pacific* **121**, 1395–1408 (2009).
- [5] Hadjiyska, E., Rabinowitz, D., Baltay, C., Ellman, N., Nugent, P., Zinn, R., Horowitz, B., McKinnon, R., and Miller, L. R., “La Silla-QUEST Variability Survey in the Southern Hemisphere,” in [*New Horizons in Time-Domain Astronomy*], Griffin, E., Hanisch, R., and Seaman, R., eds., *IAU Symposium* **285**, 324–326 (2012).
- [6] Williams, R. D., Djorgovski, S. G., Drake, A. J., Graham, M. J., and Mahabal, A., “Skyalert: Real-time Astronomy for You and Your Robots,” in [*Astronomical Data Analysis Software and Systems XVIII*], Bohlender, D. A., Durand, D., and Dowler, P., eds., *Astronomical Society of the Pacific Conference Series* **411**, 115 (2009).
- [7] Bloom, J. S., Richards, J. W., Nugent, P. E., Quimby, R. M., Kasliwal, M. M., Starr, D. L., Poznanski, D., Ofek, E. O., Cenko, S. B., Butler, N. R., Kulkarni, S. R., Gal-Yam, A., and Law, N., “Automating Discovery and Classification of Transients and Variable Stars in the Synoptic Survey Era,” *PASP* **124**, 1175–1196 (2012).
- [8] Kulkarni, S. R., “Cosmic Explosions (Optical Transients),” *ArXiv e-prints* (2012). arXiv:1202.2381.
- [9] van Haarlem, M. P., Wise, M. W., Gunst, A. W., Heald, G., McKean, J. P., Hessels, J. W. T., de Bruyn, A. G., Nijboer, R., Swinbank, J., Fallows, R., et al., “LOFAR: The LOW-Frequency ARray,” *Astronomy & Astrophysics* **556**, A2 (2013).
- [10] Atwood, W. B., Abdo, A. A., Ackermann, M., Althouse, W., Anderson, B., Axelsson, M., Baldini, L., Ballet, J., Band, D. L., Barbiellini, G., et al., “The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission,” *Astrophysical Journal* **697**, 1071–1102 (2009).
- [11] Krabbendam, V. L. and Sweeney, D., “The Large Synoptic Survey Telescope Preliminary Design Overview,” in [*Ground-based and Airborne Telescopes III*], Stepp, L. M., Gilmozzi, R., and Hall, H. J., eds., *Proc. SPIE* **7733** (2010).
- [12] LSST Science Collaboration, Abell, P. A., Allison, J., Anderson, S. F., Andrew, J. R., Angel, J. R. P., Armus, L., Arnett, D., Asztalos, S. J., Axelrod, T. S., et al., “LSST Science Book, Version 2.0,” *ArXiv e-prints* (2009). arXiv:0912.0201.
- [13] Williams, R. D. and Seaman, R., “VOEvent: Information Infrastructure for Real-Time Astronomy,” in [*Astronomical Data Analysis Software and Systems XV*], Gabriel, C., Arviset, C., Ponz, D., and Enrique, S., eds., *Astronomical Society of the Pacific Conference Series* **351**, 637 (2006).
- [14] Seaman, R., Williams, R., Allan, A., Barthelmy, S., Bloom, J., Brewer, J., Denny, R., Fitzpatrick, M., Graham, M., Gray, N., Hessman, F., Marka, S., Rots, A., Vestrand, T., and Wozniak, P., “IVOA Recommendation: Sky Event Reporting Metadata Version 2.0,” *ArXiv e-prints* (2011). arXiv:1110.0523.
- [15] Ridgway, S. T., “How Many Galactic Variables will LSST Detect?,” in [*American Astronomical Society Meeting Abstracts #219*], *American Astronomical Society Meeting Abstracts* **219**, #156.12 (2012).
- [16] Grav, T., Jedicke, R., Denneau, L., Chesley, S., Holman, M. J., and Spahr, T. B., “The Pan-STARRS Synthetic Solar System Model: A Tool for Testing and Efficiency Determination of the Moving Object Processing System,” *PASP* **123**, 423–447 (2011).
- [17] Herwig, F., “Evolution of Asymptotic Giant Branch Stars,” *ARAA* **43**, 435–479 (2005).
- [18] Iben, Jr., I., Kaler, J. B., Truran, J. W., and Renzini, A., “On the Evolution of Those Nuclei of Planetary Nebulae That Experience a Final Helium Shell Flash,” *ApJ* **264**, 605–612 (1983).
- [19] Gal-Yam, A., “Luminous Supernovae,” *Science* **337**, 927 (2012).
- [20] Sigg, D. and LIGO Scientific Collaboration, “Status of the LIGO detectors,” *Classical and Quantum Gravity* **25**(11), 114041 (2008).
- [21] Nissanke, S., Kasliwal, M., and Georgieva, A., “Identifying Elusive Electromagnetic Counterparts to Gravitational Wave Mergers: An End-to-end Simulation,” *ApJ* **767**, 124 (2013).

- [22] Matheson, T., Fan, X., Green, R., McConnachie, A., Newman, J., Olsen, K., Szkody, P., and Wood-Vasey, W. M., “Spectroscopy in the Era of LSST,” *ArXiv e-prints* (2013). arXiv:1311.2496.
- [23] Richards, J. W., Starr, D. L., Butler, N. R., Bloom, J. S., Brewer, J. M., Crellin-Quick, A., Higgins, J., Kennedy, R., and Rischard, M., “On Machine-learned Classification of Variable Stars with Sparse and Noisy Time-series Data,” *ApJ* **733**, 10 (2011).
- [24] Angeloni, R., Contreras Ramos, R., Catelan, M., Dékány, I., Gran, F., Alonso-García, J., Hempel, M., Navarrete, C., Andrews, H., Aparicio, A., Beamín, J. C., Berger, C., Borissova, J., Contreras Peña, C., Cunial, A., de Grijs, R., Espinoza, N., Eyheramendy, S., Ferreira Lopes, C. E., Fiaschi, M., Hajdu, G., Han, J., Helminiak, K. G., Hempel, A., Hidalgo, S. L., Ita, Y., Jeon, Y.-B., Jordán, A., Kwon, J., Lee, J. T., Martín, E. L., Masetti, N., Matsunaga, N., Milone, A. P., Minniti, D., Morelli, L., Murgas, F., Nagayama, T., Navarro, C., Ochner, P., Pérez, P., Pichara, K., Rojas-Arriagada, A., Roquette, J., Saito, R. K., Siviero, A., Sohn, J., Sung, H.-I., Tamura, M., Tata, R., Tomasella, L., Townsend, B., and Whitelock, P., “The VVV Templates Project. Towards an Automated Classification of VVV Light-Curves. I. Building a database of stellar variability in the near-infrared,” *Astronomy & Astrophysics in press* (2014). arXiv:1405.4517.
- [25] Masci, F. J., Hoffman, D. I., Grillmair, C. J., and Cutri, R. M., “Automated Classification of Periodic Variable Stars detected by the Wide-field Infrared Survey Explorer,” *Astronomical Journal in press* (2014). arXiv:1402.0125.
- [26] Graham, M. J., Djorgovski, S. G., Mahabal, A. A., Donalek, C., and Drake, A. J., “Machine-Assisted Discovery of Relationships in Astronomy,” *MNRAS* **431**, 2371–2384 (2013).
- [27] Matheson, T., Saha, A., Snodgrass, R., and Kececioglu, J., “ANTARES: A Prototype Transient Broker System,” in [*American Astronomical Society Meeting*], *American Astronomical Society Meeting Abstracts* **223**, #343.02 (2014).
- [28] Matheson, T., Saha, A., Snodgrass, R., and Kececioglu, J., “ANTARES: The Arizona-NOAO Temporal Analysis and Response to Events System,” in [*Hot-Wiring the Transient Universe 3*], Wozniak, P., Graham, M., and Mahabal, A., eds. (2014).
- [29] Skrutskie, M. F., Cutri, R. M., Stiening, R., Weinberg, M. D., Schneider, S., Carpenter, J. M., Beichman, C., Capps, R., Chester, T., Elias, J., Huchra, J., Liebert, J., Lonsdale, C., Monet, D. G., Price, S., Seitzer, P., Jarrett, T., Kirkpatrick, J. D., Gizis, J. E., Howard, E., Evans, T., Fowler, J., Fullmer, L., Hurt, R., Light, R., Kopan, E. L., Marsh, K. A., McCallon, H. L., Tam, R., Van Dyk, S., and Wheelock, S., “The Two Micron All Sky Survey (2MASS),” *Astronomical Journal* **131**, 1163–1183 (2006).
- [30] Evans, I. N., Primini, F. A., Glotfelty, K. J., Anderson, C. S., Bonaventura, N. R., Chen, J. C., Davis, J. E., Doe, S. M., Evans, J. D., Fabbiano, G., Galle, E. C., Gibbs, II, D. G., Grier, J. D., Hain, R. M., Hall, D. M., Harbo, P. N., (Helen He, X., Houck, J. C., Karovska, M., Kashyap, V. L., Lauer, J., McCollough, M. L., McDowell, J. C., Miller, J. B., Mitschang, A. W., Morgan, D. L., Mossman, A. E., Nichols, J. S., Nowak, M. A., Plummer, D. A., Refsdal, B. L., Rots, A. H., Siemiginowska, A., Sundheim, B. A., Tippetts, M. S., Van Stone, D. W., Winkelman, S. L., and Zografou, P., “The Chandra Source Catalog,” *Astrophysical Journal Supplement* **189**, 37–82 (2010).
- [31] Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., and Broderick, J. J., “The NRAO VLA Sky Survey,” *Astronomical Journal* **115**, 1693–1716 (1998).
- [32] Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., Allende Prieto, C., An, D., Anderson, K. S. J., Anderson, S. F., Annis, J., Bahcall, N. A., et al., “The Seventh Data Release of the Sloan Digital Sky Survey,” *Astrophysical Journal Supplement* **182**, 543–558 (2009).
- [33] Helou, G., Madore, B. F., Schmitz, M., Bica, M. D., Wu, X., and Bennett, J., “The NASA/IPAC extragalactic database,” in [*Databases and On-line Data in Astronomy*], Albrecht, M. A. and Egret, D., eds., *Astrophysics and Space Science Library* **171**, 89–106 (1991).
- [34] Cacciari, C., “GAIA: the mission and (some of) its scientific applications,” *Memorie della Societ Astronomica Italiana* **80**, 97 (2009).
- [35] Morrison, C. and Snodgrass, R. T., “Computer Science Can Use More Science,” *Communications of the ACM* **54**, 36 (2011).