Astro2020 Activities, Projects, or State of the Profession Consideration White Paper

ANTARES: Enabling Time-Domain Discovery in the 2020s

Type of Activity:  ✓ Ground Based Project  □ Space Based Project  ✓ Infrastructure Activity  ✓ Technological Development Activity  □ State of the Profession Consideration  □ Other


Principal Author:
Name: Thomas Matheson
Institution: National Optical Astronomy Observatory
Email: matheson@noao.edu
Phone: 520-318-8517

Co-authors: C. Stubens (NOAO), M. Soraïsam (NOAO), G. Narayan (STScI), A. Saha (NOAO), C. Lee (NOAO), N. Wolf (NOAO), C. Merrill (NOAO), S. Ridgway (NOAO), A. Bolton (NOAO), R. Snodgrass (Arizona), C. Scheidegger (Arizona), J. Kececioglu (Arizona), J. Peek (STScI), A. Rest (STScI), A. M. Smith (STScI), I. Momcheva (STScI), D. Petravick (NCSA), E. Morganson (NCSA)

Abstract We describe the scientific goals and capabilities of the ANTARES project. This is a software infrastructure system designed to process time-domain alerts at the scale the Large Synoptic Survey Telescope will produce. Current and future time-domain surveys will produce events at a scale well beyond the capacity of individual astronomers. We are building a system that will allow everyone access to large-scale time-domain streams with real-time filters, machine learning, and other tools so that astronomers can find and study the objects that they want. Without such a system, we run the risk of losing the great potential of astrophysical time-domain discovery in the 2020s. ANTARES will ingest millions of alerts from wide-field surveys and then annotate, characterize, categorize, and rank them, producing a value-added product for the community. This enables active follow-up at all scales, from amateurs to robotic facilities to the next generation of extremely large telescopes.
1 What Is an Alert Broker and Why Do We Need One

The 2020s will see the culmination of the astronomical time-domain revolution that began in the last decade. For centuries, astronomers found objects by eye that changed in brightness or moved. In the last century, photographic plates provided a permanent record of observations, but also presented a new way to find changing objects by blinking the plates. The advent of CCDs enabled even greater capability for time-domain discovery, partly as a result of their greater sensitivity, but also because digital images could be subtracted in a machine. This digital image subtraction technique allowed for wholesale detection of all objects that had changed between the two images. Once wide-field cameras could be built, the stage was set for explosive growth in time-domain discovery.

In the past, astronomers only had a handful of time-domain events to consider on any given night. However, new wide-field surveys have dramatically increased the rate of discoveries over even the past 5 year. The Zwicky Transient Facility (ZTF, Bellm, 2014; Smith et al., 2014; Dekany et al., 2016; Bellm & Kulkarni, 2017) currently issues several hundred thousand alerts from its public survey every night. The anticipated output of the Large Synoptic Survey Telescope (LSST, Ivezić et al., 2008; Kantor, 2014; Gressler, 2016) is ten million time-domain alerts per night, every night, over the course of its ten-year survey. The volume and rate of astronomical time-domain alerts is already well beyond the capacity of individual astronomers to digest. LSST will increase the scale by orders of magnitude.

It is likely that a significant fraction of these alerts (\(\gtrsim 90\%\), Ridgway et al., 2014) will be repetitions of known variable stars and moving objects. These have to be identified in the alert stream so that unknown objects can be evaluated. Hidden among these millions of alerts each night will be rare and interesting objects that have relatively brief lifetimes. Only with additional follow-up, either photometric, but especially spectroscopic, will these objects reveal their nature.

In addition, the time scale in which objects must be identified has shrunk as early-time observations have proven of such great value (e.g., Khazov et al., 2016; Tanaka et al., 2016; Arcavi, 2018; Pursiainen et al., 2018; Kochanek, 2019). Finally, in the era of multi-messenger astronomy, the ability to detect and identify the correct electromagnetic counterpart among the thousands of candidates in a gravitational wave footprint is of critical importance.

To address this issue, the National Optical Astronomy Observatory (NOAO), the University of Arizona Department of Computer Science, the Space Telescope Science Institute (STScI), and the National Center for Supercomputing Applications (NCSA) are developing a software infrastructure system (the Arizona-NOAO Temporal Analysis and Response to Events System, ANTARES, Saha et al., 2014, 2016; Narayan et al., 2018) designed to effectively and efficiently filter time-domain alerts at LSST scale. These software tools are commonly referred to as brokers (e.g., Borne, 2008) as they sit between producers and consumers of data while adding value. The ANTARES system is already processing the public alert stream from ZTF. The capabilities highlighted below can be seen in action at https://antares.noao.edu/.

We believe that there are no serious technological impediments to deploying this system. What we need is funding to support the underlying infrastructure and the personnel necessary to ensure effective operation of ANTARES. The 2020s will present a great opportunity for time-domain discovery. We should make sure that the US community can take full advantage.
2 Scientific Goals of ANTARES

The overarching goal of ANTARES is to serve the broadest possible scientific interests of our community. We are designing the system to be flexible enough to support a wide range of filtering and classification tools. We envision ANTARES as a generic broker that can be tuned to many possible goals and adds value to every alert. It will act at a high-level, enabling downstream brokers, Target and Observation Managers (TOMs, e.g., Street et al., 2018), and astronomers to operate in more detail on substreams of interest. The scientific aims of the broker will be determined largely by the demands and interests of the community that uses it. It can act as a time-domain science platform that allows users to deploy filters and other tools to fulfill their science needs in the LSST era. Our archive of alerts will provide the training sets users will need to build their filters. We are building a system that will enable astronomers to maximize the scientific output of time-domain surveys.

We have several science use cases that guided development of the system and illustrate the breadth of the capabilities ANTARES enables. These include:

1) Identifying electromagnetic (EM) counterparts to gravitational wave (GW) sources. The recent tremendous success in recognizing and following the kilonova associated with GW170817 (e.g., Abbott et al., 2017a,b; Chornock et al., 2017) has opened an entirely new field of astronomy. Given the relatively poor localization of GW targets, wide-field surveys will be used to locate the EM source, although clever targeting of likely host galaxies can mitigate the need for wide-field observations. Kilonovae are generally predicted to be short-lived, so rapid identification is critical. ANTARES can use the GW localization and predicted characteristics of kilonovae to locate potential candidates. This demonstrates the use of features to distinguish known classes of objects.

2) Tidal disruption events (TDEs) can shed light on the nature of black holes, including accretion mechanisms and jet formation (e.g., Metzger & Stone, 2016; Hung et al., 2018). Alerts at the centers of galaxies are potential TDEs, but could also be active galactic nuclei (AGN). ANTARES will include multiwavelength catalogs that can help to distinguish AGN (and identify them for the AGN community), showing the value in object association beyond optical wavelengths. In addition, there is a statistical preference for TDEs in post-starburst galaxies (French et al., 2016; French & Zabludoff, 2018), although the physical basis for this is not known. The probability distribution of objects across host types will be incorporated into the filtering process.

3) Relatively common transients may not require follow up for every target. Astronomers may want to implement a filter that provides transients on demand. For example, LSST will detect millions of Type Ia supernovae (SNe), far beyond the world’s capacity to obtain spectra. A cosmology campaign may operate for some period of time during which they have classically scheduled nights with large telescopes for spectroscopic follow up. They don’t need to be notified about every Type Ia SN, just the ones visible during their campaign. This highlights the flexibility of the ANTARES pipeline structure to accommodate changes in filter order and use. Alternately, filtering could be basic, sending the alert stream with a few, well-characterized cuts to the group, allowing downstream systems to select objects. ANTARES will maintain provenance of all alerts ingested and issued along with the code that operated on them, allowing detailed analysis of the selection function.

4) Known Solar System objects can be flagged by the alert generation systems, so ANTARES can select this specific stream and redirect it to a special-use broker that assesses the objects for
activity. This illustrates how ANTARES can fit into a larger time-domain ecosystem where there are many brokers and TOMs. The interconnection between these systems can produce a more effective overall system. We are already sending known Solar System objects from the ZTF public alert stream to a team of planetary astronomers at Northern Arizona University and Lowell Observatory.

5) For objects that vary on longer timescales, immediate notification of any change may not be necessary. Variables (periodic or aperiodic) will be more useful for study once full light curves are available. To enable research and analysis on these classes of objects, we will make our alert history database available. Because this has the full history from all LSST alerts, filters can act on all available data, rather than just the twelve-month history provided by LSST alerts themselves. One can already use the ANTARES client within the NOAO DataLab (https://datalab.noao.edu/). The DataLab offers a full-service data-science platform that enables exploration of ANTARES alerts.

6) For some objects, astronomers will want notifications of any alert. We will provide a capacity for watch lists so that anyone can follow alerts on objects of interest to them. These watch lists are private and only visible to the users that create them. We can provide direct notifications for an alert on a watched object.

7) There is still a large space for potential discovery. ANTARES will incorporate a “Touchstone” that represents the feature-space distribution of known objects. Alerts that fall well out of the populated regions of this feature space could be something completely unknown (see, e.g., the well-known Kasliwal, 2012, energy/timescale diagram). Anomaly detection will be a key part of ANTARES filtering. We can implement at least partial solutions to all these cases right now. If an astronomer can write a filter within our system to identify their targets of interest, we can accommodate it, subject to resource (compute, storage, or bandwidth) limitations. In addition, downstream brokers could take advantage of our annotations to fulfill other use cases.

3 Relevant Astro2020 Science White Papers

More than forty Astro2020 Science White Papers have direct or indirect requirements fulfilled by the ANTARES project. Several white papers call specifically for funding and support of broker and broker-like infrastructure. These include:

- Chang, P., et al., Cyberinfrastructure Requirements to Enhance Multi-messenger Astrophysics
- Chanover, N., et al., Triggered High-Priority Observations of Dynamic Solar System Phenomena
- Cowperthwaite, P., et al., Joint Gravitational Wave and Electromagnetic Astronomy with LIGO and LSST in the 2020s
• Graham, M., et al., Discovery Frontiers of Explosive Transients: An ELT and LSST Perspective
• Kupfer, T., et al., A Summary of Multimessenger Science with Galactic Binaries
• Olsen, K., et al., Science Platforms for Resolved Stellar Populations in the Next Decade (mentions ANTARES)
• Siemiginowska, A., et al., The Next Decade of Astroinformatics and Astrostatistics

Many of the white papers that focused on multi-messenger astrophysics will benefit greatly from a broker. As described in our science use cases, this is a function that we designed ANTARES to perform. These white papers include:

• Burns, E., et al., A Summary of Multimessenger Science with Neutron Star Mergers
• Burns, E., et al., Opportunities for Multimessenger Astronomy in the 2020s
• Caldwell, R., et al., Cosmology with a Space-Based Gravitational Wave Observatory
• Chornock, R., et al., Multi-Messenger Astronomy with Extremely Large Telescopes
• Foley, R., et al., Gravity and Light: Combining Gravitational Wave and Electromagnetic Observations in the 2020s
• Littenberg, T., et al., Gravitational Wave Survey of Galactic Ultra Compact Binaries
• Metzger, B., et al., Kilonovae: nUV/Optical/IR Counterparts of Neutron Star Binary Mergers with TSO
• Reitze, D., et al., The US Program in Ground-Based Gravitational Wave Science: Contribution from the LIGO Laboratory
• Sathyaprakash, B., et al., Multimessenger Universe with Gravitational Waves from Binary Systems
• Shawhan, P., et al., Multi-Messenger Astrophysics Opportunities with Stellar-Mass Binary Black Hole Mergers
• Shoemaker, D., et al., Gravitational-Wave Astronomy in the 2020s and Beyond: A View across the Gravitational Wave Spectrum
• Shoemaker, D., et al., Gravitational Wave Astronomy with LIGO and Similar Detectors in the Next Decade

Many other white papers discussed a wide variety of science with supernovae, ranging from basic physics of explosion mechanisms to cosmology. Again, ANTARES can filter time-domain alerts to identify supernovae. In addition, it can monitor strongly lensed systems and report potentially lensed supernovae. These white papers are:
• Brown, P., et al., Keeping an Ultraviolet Eye on Supernovae

• Hložek, R., et al., Single-Object Imaging and Spectroscopy to Enhance Dark Energy Science from LSST

• Kim, A., et al., Testing Gravity Using Type Ia Supernovae Discovered by Next-Generation Wide-Field Imaging Surveys

• Milisavljevic, D., et al., Achieving Transformative Understanding of Extreme Stellar Explosions with ELT-enabled Late-time Spectroscopy

• Perlmutter, S., et al., The Key Role of Supernova Spectrophotometry in the Next-Decade Dark Energy Science Program

• Scolnic, D., et al., The Next Generation of Cosmological Measurements with Type Ia Supernovae

• Slosar, A., et al., Dark Energy and Modified Gravity

• Wheeler, J. C., et al., ELT Contributions to The First Explosions

• Wood-Vasey, M., et al., Type Ia Supernova Cosmology with TSO

• Zingale, M., et al., MMA SAG: Thermonuclear Supernovae

There are several white papers that describe follow-up science cases for a wide variety of targets, from stellar binaries to microlensing to tidal disruption events. ANTARES has the capacity to integrate filters across all these cases. The white papers are:

• Blakeslee, J., et al., Probing the Time Domain with High Spatial Resolution

• Kara, E., et al., X-ray Follow Up of Extragalactic Transients

• Lu, J., et al., From Stars to Compact Objects: The Initial-Final Mass Relation

• Maccarone, T., et al., Populations of Black holes in Binaries

• Pasham, D., et al., Probing the Cosmological Evolution of Super-massive Black Holes using Tidal Disruption Flares

• Rix, H.-W., et al., Binaries Matter Everywhere: from Precision Calibrations to Re-Ionization and Gravitational Waves

• Schaefer, G., et al., High Angular Resolution Astrophysics in the Era of Time Domain Surveys

• Wheeler, J. C., et al., ELT Contributions to Tidal Disruption Events

Finally, a few white papers propose new time-domain surveys for which a broker like ANTARES is generically useful. These include:
• Kasliwal, M., et al., The Dynamic Infrared Sky
• Ross, N., et al., Opportunities in Time-Domain Extragalactic Astrophysics with the NASA Near-Earth Object Camera (NEOCam)
• Wang, L., et al., JWST: Probing the Epoch of Reionization with a Wide Field Time-Domain Survey
• Zemcov, M., et al., Opportunities for Astrophysical Science from the Inner and Outer Solar System

In addition, Ntampaka et al. (The Role of Machine Learning in the Next Decade of Cosmology) point to machine-learning aspects of ANTARES.

4 Technical Overview

4.1 Data Products & Services

ANTARES will annotate the alert stream, adding information from other astronomical catalogs and from evaluations performed in filters. We cross-match each alert with astronomical catalogs (e.g., Sloan Digital Sky Survey, 2MASS, NASA Extragalactic Database, Chandra Source Catalog, etc.). We also associate alerts with any previous history at the alert location from our database. This means that prior associations or calculations do not need to be repeated and that alert history from other time-domain surveys will be available.

The core function of ANTARES is to filter and classify alerts in real-time. Filters can be simple (such as finding bright or rapidly changing objects) or complex (such as the implementation of RAPID, a machine-learning classifier, Muthukrishna et al., 2019). The ANTARES team will develop several filters to serve a broad range of science interests. The exact complement of filters will depend on user interest. We expect development, testing, profiling, and optimization of these filters and maintenance of the catalogs for cross-matching to be conducted at NOAO.

In addition, users of ANTARES will be able to submit their own filters. STScI will host a service to develop and test filters on a subset of the alert stream, as well as cross-matching against HST and other holdings at the Mikulski Archive for Space Telescopes (MAST). Users will prototype their

Figure 1: The ANTARES web portal showing some of the features of the current system.
filters using our “devkit” and this sandbox environment at STScI or NOAO before the filter is run against the full LSST alert stream. We will provide a watch list service where users can upload targets of interest so that they can be notified of any alert on those targets to simplify follow-up studies with other facilities. Additionally, filters will be executable on archives of light curves both at NOAO and STScI, allowing tasks such as anomaly detection on past and on-going missions such as TESS. This will also allow ANTARES to be prototyped for future missions such as WFIRST.

The output from ANTARES can be accessed in a variety of ways. The ANTARES portal provides a web-based interface to the streams produced by filters (see Figure 1). We also send stream notifications to the ANTARES Slack workspace (antares-noao.slack.com). The Slack channel does not have the alert data, but rather links to the ANTARES portal. Watch list matches also appear as Slack messages to specific users or channels. The Slack workspace is open to registered users of ANTARES. We will support other notification systems based on user demand. We make the full data set for alerts in filtered streams available via Apache Kafka. We provide subscription keys to users along with a Python-based toolkit so that they can automatically receive alert data directly to their own machines. In addition, our alert database, with all the annotations, will be queryable to enable analysis or discovery of objects with longer timescales. We implement this via ElasticSearch technology and also provide an API for automated interaction.

We make no restrictions on the nature of users of our system. If we reach a point where there are resource limitations, we will rely on a science advisory committee made up of community members to provide us with a ranking for compute and bandwidth priority, similar to how a time-allocation committee evaluates proposals for observing resources.

The ANTARES team does not anticipate coordinating follow-up observations. We will, however, incorporate filters and APIs to enable efficient communication between our system, astronomers, downstream brokers, and TOMs. We expect to be able to connect to ground-based resources through tools such as the Astronomical Event Observatory Network (AEON) being developed by Las Cumbres Observatory (LCO) and NOAO. STScI will develop APIs to facilitate follow-up observations with space-based resources such as HST and JWST. Where existing TOMs infrastructure exists (e.g., Swift), we will work with the relevant teams to connect ANTARES and these resources. The generic TOM ToolKit being developed at LCO can already interface directly with ANTARES. Critically, we will work with the community to ensure that the data they need to make follow-up decisions is available via ANTARES-annotated alerts, adding requisite features as appropriate.

### 4.2 Implementation

An ANTARES installation consists of three hardware clusters: a Kubernetes cluster, a Kafka cluster, and a database cluster. A fourth cluster is needed for non-realtime functionality, such as archival search of alerts (see Figure 2). Kubernetes is an open source technology developed by Google that orchestrates the deployment and runtime operations of containers across clusters of machines. ANTARES uses Kubernetes to run the alert-processing pipeline, web services, and other processes with the exception of Kafka and the database. Apache Kafka is a streaming technology

---

1. [https://lco.global/tomtoolkit/](https://lco.global/tomtoolkit/)
2. [https://kubernetes.io/](https://kubernetes.io/)
3. [https://kafka.apache.org/](https://kafka.apache.org/)
that ANTARES uses for outputting alerts to other systems and will use for maintaining internal work queues. Both Kubernetes and Kafka are used by ANTARES in production now, and are also used extensively by LSST data management. The ANTARES database will be a distributed database cluster.

Exact hardware requirements for LSST are have not been finalized, but current load testing does provide rough estimates. We are in the process of acquiring a new test cluster that we expect to be capable of LSST-scale throughput tests, but not LSST-scale data storage capacity. Results of these tests will further inform the exact hardware requirements for LSST. Work is also underway to determine exact hardware needs for the LSST-scale database.

Figure 2: Schematic architecture of the ANTARES alert processing pipeline, highlighting some of the technologies used in system.

ANTARES has been running live on the ZTF alert stream since December 2018. The alert processing pipeline has had no unintentional downtime yet as of May 2019. A web interface to the system is also live, and is being actively used by 30-40 users per day (out of ~140 registered users). There are user-contributed filters and several private watch lists. ANTARES uses an unmodified version of Kafka. The ANTARES database is the only component still in a prototyping phase. Our prototype ZTF database will not be sufficient for LSST. We expect the capacity and I/O requirements for the ANTARES database to be comparable to the LSST Prompt Products Database (PPDB) requirements of ~0.4 petabytes\(^4\), with a throughput of ~5000 queries per second. We are experimenting with the Cassandra distributed Key-Value system and Google Cloud technologies.

ANTARES is designed with three potential hosting providers in mind: NCSA, Google, and Amazon Web Services (AWS). The containerized nature of ANTARES will allow us to remain

\(^4\)https://dmtn-018.lsst.io/
flexible in the choice of a hosting site. NCSA is developing experience in hosting Kubernetes and Kafka systems through their work with LSST DM and will likely be providing such services continuously through LSST operations. NCSA may also host the LSST PPDB, which will be very similar to ANTARES database. Google and AWS both offer fully-managed Kubernetes services, and allow customers to deploy Kafka, although it is not offered as a fully-managed service.

5 Organization, Partnerships, and Current Status

NOAO is the lead organization in managing, developing, and operating the ANTARES project. These efforts are hosted within the Community Science and Data Center (CSDC). At the start of fiscal year 2020, CSDC will become a part of NSF’s new National Center for Optical-IR Astronomy (NCOA), which incorporates current NOAO programs as well as Gemini Observatory and LSST Operations. Faculty in the Department of Computer Science at the University of Arizona provide expertise in database design, algorithm development, machine learning, and visualization. Scientific staff at STScI develop user-facing tools and enable connections to space-based observing resources and archives. Staff at NCSA will help with integration of the system into the data center environment. The ANTARES project is already live and processing the public alerts stream from ZTF. We have funding for near-term goals, but further resources will be necessary to operate at LSST scale for the duration of the LSST survey.

6 Schedule

ANTARES is operational at the ZTF scale. We will continue to add functionality in response to community demand. Major goals over the next few years are:

- FY20, Conduct scaling tests to determine infrastructure requirements for LSST operations
- FY20, Determine technology for Alerts database
- FY20, Develop internal database representation for astronomical feature space (classification tool)
- FY21, Assess data center/cloud computing options and select host site for LSST operations
- FY22, Deploy fully functional LSST-scale system

7 Cost Estimates

We have not yet completed scaling tests that would specify the full infrastructure needs in the LSST era, so cost estimates are somewhat speculative. Based on estimates of the size and use of the alerts database (the cost driver for technology), we anticipate $1-2 million over the ten-year LSST survey. Considering the scale of the project, we estimate 6-8 FTE distributed across software engineers and scientists for the operations staff. This implies an annual cost near $1.4
million, although there could be some trade-off in engineer costs vs. data center support. The total is \(\sim$16\) million over ten years. This value is uncertain, but we expect that the lifetime cost is well within the Small category for ground-based projects as defined in the Call for Activities, Projects, or State of the Profession Consideration White Papers from Astro2020.


Kasliwal, M. M. 2012. “Systematically Bridging the Gap Between Novae and Supernovae.” PASA, 29, 482


