June 30, 2015

Summary of Data Management Principles: VERITAS

Experiment description:

The Very Energetic Radiation Imaging Telescope Array System, VERITAS, at the F.L. Whipple Observatory (FLWO) near Tucson, AZ, provides the ground-based capability to study extremely energetic gamma rays, ranging in energy from 50 GeV to 50 TeV, potentially produced from a variety of astrophysical sources. The gamma rays are observed from the light they induce as they interact with the Earth's atmosphere. VERITAS permits unprecedented elucidation of the properties of the sources of these gamma rays. The results of the VERITAS studies of these cosmic Tevatrons / Pevatrons often have broad often have broad implications beyond the physics of the sources themselves. Topics addressed include indirect dark matter searches, cosmology, black holes, fundamental physics, and the origin of cosmic rays. VERITAS studies have many multi-messenger (neutrino, cosmic-ray and gravitational wave) astrophysics implications, and complement the ongoing efforts of the Fermi Gamma-ray Space Telescope and the High-Altitude Water Cherenkov (HAWC) observatory.

The VERITAS consists of four 12 meter telescopes whose mirrors each reflect Cherenkov radiation produced by gamma- and cosmic-ray initiated air showers onto a detector, or camera, consisting of 499 2.5 cm photomultiplier tubes (PMT). Sophisticated hardware and software are used to digitize the PMT signals 500 million times per second, implement a high-rate/low dead-time data acquisition (DAQ) system and a multi-level event trigger, perform online analysis and diagnostics, and store the data for later analysis. VERITAS began regular observations with the full 4-telescope array in September 2007, was upgraded in Summer 2012, and plans to operate until at least Summer 2019.

DOE's roles in the experiment:

The US Department of Energy (DOE) provided ~42% of the observatory's $17.6 million dollar construction cost, and has historically provided one-third of the funds necessary for the maintenance and the routine operation of both the telescopes and the array facility. In addition, the DOE has historically funded several university / lab groups who have supported VERITAS research. Responsibilities for various aspects of the experiment are determined by the VERITAS collaboration, with oversight provided by a Project Office located at the experiment's site.

Partnerships:

The U.S. participants in the operation of VERITAS are supported by the Physics (PHY) division of the U.S. National Science Foundation (NSF), the Office of High Energy Physics (OHEP) in the U.S. Department of Energy (DOE), and the Smithsonian Astrophysical Observatory (SAO). Operations are supervised by the three agencies via the Joint Oversight Group (JOG). The VERITAS Project JOG has the responsibility to see that the VERITAS Project is effectively managed and executed, and will coordinate agency policies and procedures regarding project oversight and management. The JOG will serve as the point of contact for the VERITAS Project Office with SAO, NSF

and DOE. Specific responsibilities of the JOG, as well as its membership, are described in the MOU between SAO, NSF and the DOE, which relates specifically to the VERITAS Project. The foreign institutions involved in the VERITAS Collaboration are responsible for managing the operational funds that they bring to the project.

VERITAS Project Organization:

The collaboration that operates VERITAS is international in scope, and consists of ~100 scientists from 20 institutions in 4 countries. Ten of the institutions are signatories to the VERITAS Teaming Agreement. The remaining institutions who joined the collaboration during or after the construction phase and are not signatories to the Teaming Agreement. The members of the VERITAS collaboration can be found online (http://veritas.sao.arizona.edu/about-veritas-mainmenu-81/newpeople).

The VERITAS Executive Committee (VEC) serves as the collaboration's governing body. The Managing Organization, appointed by the VEC, is the organization that manages the VERITAS Project. The VERITAS Project Office is the organizational subunit of the Managing Organization that is directly involved with executing the Project. The managing organization is appointed by the Teaming Agreement and is currently SAO. The VEC established and determines the membership of a VERITAS Science Board which governs the science program of VERITAS. The collaboration spokesperson (Science Board chair), deputy spokesperson and VEC chair are all elected positions with 2-year terms. Only the Project Scientist, who is the head of the SAO group / Project Office, is effectively unchanging.

An experimental operation's plan (EOP) describes the overall project structure and management at all levels, including agency oversight and the partnership between DOE, SAO and NSF. It establishes the technical scope and baseline against which the agencies will monitor the execution of the project. Two organization charts from this EOP can be seen in Figures 1 and 2. The EOP is regularly updated, has active links to all high-level documents (e.g. MoUs, Teaming agreements, etc), and the most recent version can be found here: http://veritas.sao.arizona.edu/~benbow/VERITAS_EOP/VERITAS_EOP_Dec2014.pdf

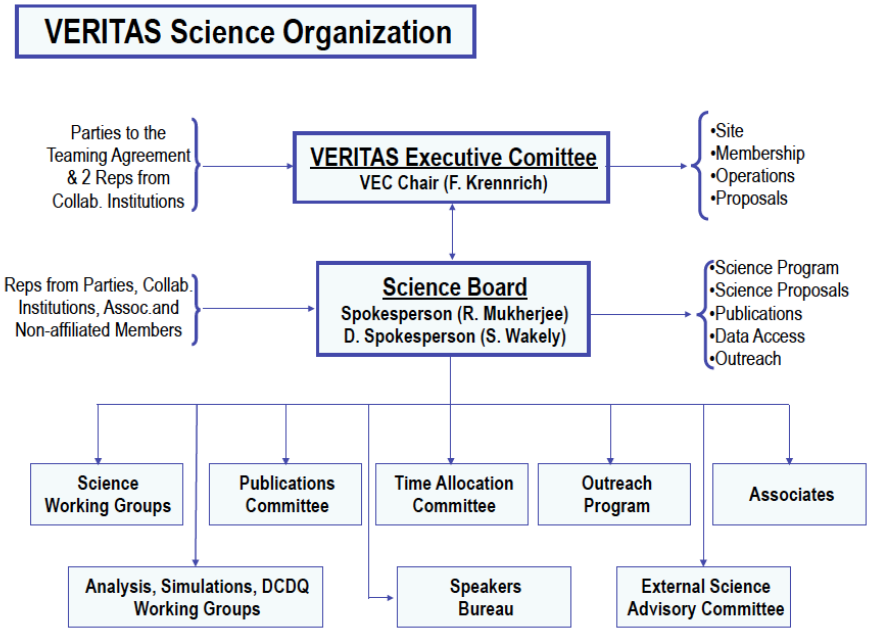

Figure 1: VERITAS Science Organization

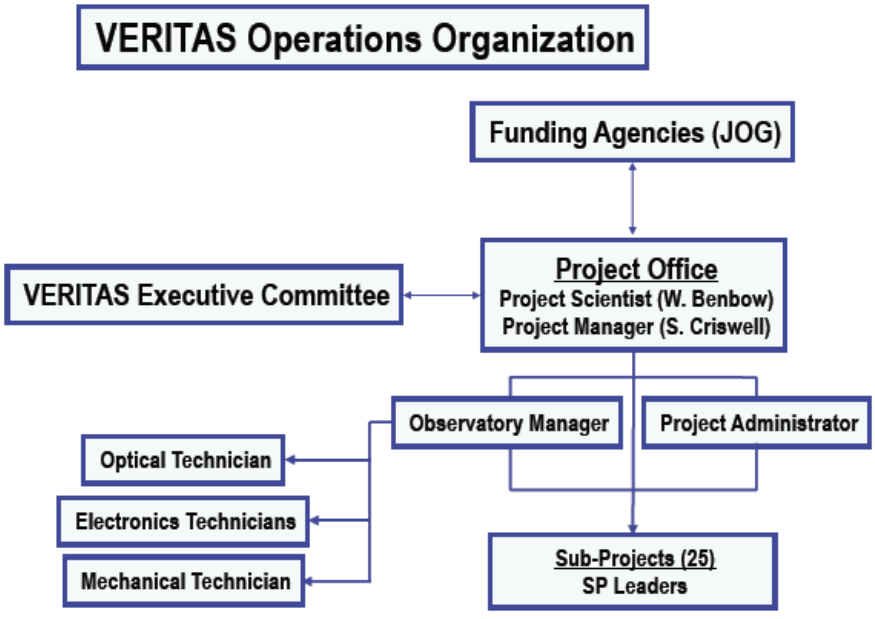

Figure 2: The organization of the VERITAS Project Office.

Data policy management:

The VERITAS Science Board is responsible for setting the VERITAS data policy and has delegated various aspects of its implementation to members and committees of the collaboration. The VERITAS Project Office ensures that the Data Management Plan described below is implemented. There are significant quantities of various forms of data produced by VERITAS each year (~45 TB / year; 311 TB total as of Sept 2014). The data storage methodology is detailed in a management plan that was presented to, and reviewed by, both the various funding agencies and previous review committees. The VERITAS data management plan has functioned well and without incident for the past 8 years of operation. The following briefly summarizes the data management plan for VERITAS.

Types of data, their archive plan and preservation of access:

VERITAS work will generate numerous types of scientific data, software, electronic logbooks, technical notes, and other documentation. The scientific data are largely the electronic readouts of all the extensive, custom VERITAS instrumentation. These numerical data are either stored in a custom, compressed data format that is archived at UCLA (and funded via a sub-award from the Project Office), or for simpler instruments (e.g. meteorological) in a data-base hosted by UC Santa Cruz and mirrored at FLWO, or for documentation on a centralized wiki / logbook system. The UCLA data storage facility consists of numerous commercial raid-based disk arrays, ensuring electronic back ups of the data, as well as high-density physical media (2 tape copies) stored in a different area. Data are transferred daily from the experiment to UCLA and to a back-up archive at the University of Utah (few month capacity; most recent data only). The back-up archive exists for redundancy as well as to guard against temporary connectivity issues at UCLA that might prevent access to the most recent, highest-priority data. Recent data are also stored at the experimental site (which has limited bandwidth) for at least 1 month to ensure no transfer problems to the permanent UCLA archive. All data on the UCLA archive are easily downloaded using standard linux software by VERITAS collaboration members. Back-up media (tapes) are annually returned to the VERITAS site to ensure a comprehensive, permanent archive exists in a geographically distinct location to guard total loss of the archive to catastrophic regional events (e.g. massive earthquakes).

All custom algorithms for processing the VERITAS data are archived at UCLA, and these algorithms are easily available for CVS download by members of the VERITAS Collaboration, which ensures many backups and change logs exist. Processing these data requires significant computing resources and occurs at the various VERITAS institutions as needed by the interested VERITAS groups; processed data can be easily regenerated from the raw data stored at UCLA. Simulated detector data, required for generating instrument response functions, are also stored at UCLA for both gamma-rays (signal) and protons (background); generation of these simulated data is funded by separate VERITAS institutional (i.e. university) base grants. The simulated data are processed with the same software as the actual data. Instrument response functions, for standard analysis methodology, are also stored at UCLA and software exists to generate these for custom projects. All VERITAS software, as well as data

access methodology, is documented and available from the VERITAS internal webpages. Software changes are documented via standard CVS logging and bugs are reported / tracked via commercial software (Bugzilla). All the extensively-used electronic logbooks, technical notes and documentation are stored on the VERITAS internal web pages, which are mirrored to ensure backups exist. Uploads and changes to the VERITAS web pages are logged and archived versions are stored. Data produced by VERITAS will be maintained and electronically available from the relevant storage centers for at least 5 years after the end of the project.

Data and metadata standards:

There is currently no metadata standard in the field of gamma-ray astronomy. However, all data are logged in a standard (custom) manner that ensures the all data products can be retrieved and cross-referenced easily. Indeed, all VERITAS data, software, documentation, etc. are easily located and accessible to all VERITAS members. Most of the electronic documentation is performed using commercial software (primarily wiki software). UCLA has managed the storage of the VERITAS scientific data and software without incident for eight years. The same can be said for the various institutions which host the data bases, web pages and wikis, as well as their mirrors. As it is virtually impossible for an outsider to handle all but the highest level of the VERITAS data products, the VERITAS collaboration takes extensive measures to ensure the accuracy of its results; This includes a requirement that all results of data analysis be reproduced by at least one other completely independent analysis chain, and that the analysis chain be able to reproduce results from the standard gamma-ray reference source (the Crab Nebula). Particularly high-impact results often have more cross checks.

Policies for data access and sharing:

We will disseminate the results of our work through appropriate peer-reviewed journals, technical reports and conference presentations. All these items will be made available on a public website, along with simple summaries of the results. The vast amounts of data generated by the proposed work will be largely specific to the VERITAS effort, and will not be posted to the public website for external access, with the exception of the highest level data products (e.g. spectral information, light curves, sky maps). We do not expect many, if any, requests for lower-level data products. While unlikely, the VERITAS collaboration is willing to share other data products with other researchers as reviewed on a case-by-case basis by the VERITAS Science Board (cost, hardship, usefulness, etc). We do not expect patentable inventions to result from the efforts funded by this proposal; Should there be such inventions, the related data will be made available after the patent application is made. Some of the developmental efforts for VERITAS were, and conceivable could be, conducted by private companies; data and documents from these efforts may remain the intellectual property of the companies and may not be freely available.

Policies and provisions for re-use or re-distribution:

No human or animal subjects are involved, nor does any of the data have privacy issues. In principle, all VERITAS data can be made available to external researchers by request, subject to a reasonable delay while publications are prepared, and so long as this does not incur undue cost or hardship on the project. We note that low-level VERITAS data requires considerable storage space and is of limited use without extensive experiment-specific software algorithms to translate the numerical data into useful quantities. This mid-level calibrated data (e.g. pixel signal & location, trigger time) is also of no practical use without vast amounts of instrument-specific software and expertise, as well as the related simulated, instrument-response functions. Even this level of data still requires considerable storage. As there are few experts in the world who can handle these data, and their effective use would require many such experts, only exceptional circumstances would justify the release of the low- or mid-level data, and there are no plans to make this publicly available. It could be possible to make event-level information public (e.g. arrival time, energy, direction, background rejection parameters), however given the event rate (300 Hz; 1300 hours per year) such an endeavor, or at least a comprehensive one, would require significant additional personnel and IT funding that is not presently requested. All high-level data products (e.g. sky maps, spectra, light curves) will be made public immediately after their related publication via the VERITAS web site. The published figures will be available for easy download in standard publishing (.eps) or bitmap (.png) formats. The sky maps will also be made available in “FITS” format for astronomers, and the numerical data from the plots will minimally be provided in ASCII format for ease of use by other scientists. All VERITAS journal publications will also be posted to the academic pre-print servers (subject to any potential copyright constraints).

Responsiveness to SC Statement on Digital Data Management

This data management plan fully follows SC Statement on Digital Data Management.