

**National Optical Astronomy Observatory
Data Management Plan
18 August 2016**

Adam S. Bolton, Ph.D.
Associate Director, System Science and Data
bolton@noao.edu

Sean G. McManus
Head, Data Management Operations
mcmanus@noao.edu

Overview

NOAO Data Management Operations (DMO) will maintain an archive facility for the long-term preservation and distribution of digital telescope data to the astronomical community. Responsibilities include development and operation of computer hardware/software infrastructure that facilitates raw telescope data capture, transport and archive ingest, pipeline data reduction, and web-based tools (NOAO Science Archive) that connect researchers with science data products.

Data Capture and Transport

Raw data capture involves the collection of images and calibration data from telescopes and associated instrumentation, concurrent with observing. DMO will capture data from telescopes located at Kitt Peak National Observatory (KPNO) and Cerro Tololo Inter-American Observatory (CTIO). This includes the Blanco and Mayall 4-meter telescope observing programs, as well as community observing programs at SOAR, SMARTs and WIYN partner facilities.

During an observing night, data files will be captured from a computer system or observer console at the telescope where files are constructed. DMO software will automatically detect new observations and copy them to mountaintop disk caches, to be staged for transport to mass storage facilities at NOAO-North (Tucson) and NOAO-South (La Serena).

Data accrued through an observing night will be tagged with a proposal ID assigned to the Principal Investigator (PI) of the observing program. The proposal ID notifies the Science Archive of file ownership and proprietary period. Typically, observations will be proprietary to the PI for a period of 18 months (or 12 months for survey programs), as agreed upon when telescope time is allocated through Telescope Allocation Committee (TAC). Calibration data is non-proprietary.

Data will be transported from the telescope-instrument to the Science Archive with open source software tools (e.g. rsync), as well as in-house tools (e.g. NOAO Data Transport System). A checksum is verified at various points within the system to ensure file integrity while passing through various networks and devices.

Data Ingest

The NOAO Science Archive has accrued over 40 different telescope-instrument combinations in its history, each with a unique data and metadata structure. Imaging and spectroscopic data are stored in FITS format, which is a commonly used digital file format in astronomy (http://fits.gsfc.nasa.gov/fits_primer.html). Files are normally compressed in fpack format (<https://heasarc.gsfc.nasa.gov/fitsio/fpack/>) to conserve disk storage resources. Additionally, thumbnail and preview images are stored in standard web graphics formats, such as JPEG and PNG.

The Telescope Automatic Data Archiver (TADA) software, developed in-house at NOAO will normalize FITS keyword name/value pairs as required for proper indexing inside the Science Archive backing database. TADA will rename original raw data into a standardized naming scheme across all instruments.

In addition, the TADA framework will monitor the status of raw files between the telescope and Science Archive, and provide feedback to Operations personnel to diagnose and correct errors. Further, TADA will provide DMO with tools to re-ingest and normalize historic datasets that are of high value to the Science Archive.

Raw data will be typically available to the PI via the Science Archive, within an hour of observing, depending on file sizes and network throughput, but complete reconciliation of a full night of observations may be deferred until the next business day.

Reduced Data

NOAO will operate data-reduction pipelines for the Mosaic, NEWFIRM, and DECam imagers. Both the reduced science data (single exposures, stacks, source catalogs and image previews) and their associated stacked master calibration data (masks, biases, darks, and flats) will be archived. The root filenames remain the same as their original raw data files, providing an easy association between the reduced and raw data.

NOAO will also host the enhanced reduced data products from the NOAO large survey program delivered by the survey teams. Enhanced data products may include enhanced stacks, object catalogues, and spectroscopic data.

Depending on staff resources, an effort will be made to provide reduced data products to the PI within 2 weeks of observing.

NOAO Science Archive

Archived raw and reduced data products will be accessible from the NOAO Science Archive web page (<http://archive.noao.edu>). All data will be publicly available after proprietary periods elapse. Otherwise, principal investigators and their co-investigators will be permitted access to their proprietary data through a secure account.

Assistance will be made available to all users of the Science Archive, via a Help link on the Science Archive web pages.

The overarching goal of the the Science Archive is to enable science across a broad spectrum of the scientific community. To that end, the Science Archive web pages will aim to deliver flexible web-based and programmatic interfaces to users of the system. These interfaces will adapt to evolving user requirements and changes to telescope instrumentation and datasets over time.

NOAO Data Lab

NOAO will deploy high-level tools and interfaces through the Data Lab initiative, with a focus on maximizing the community science return from public data products of large and homogeneous imaging surveys being conducted with DECam and other instruments. Data Lab services will enable users to discover coherent survey datasets hosted by NOAO, according to queries based on multiple constraints. Data Lab will also provide sophisticated query capabilities for selection and analysis of samples from within the large catalogs delivered by these surveys, and capabilities for tying these samples back to cutouts of the pixel-level data from which they were derived. Finally, Data Lab will provide collaborative virtual storage and computing facilities in co-location with the NOAO archive, to enable users to bring their analysis to the data. Most of these Data Lab capabilities will be exposed through programmatic interfaces as well as through high-level applications.

Data Access and Usage Policy

The published NOAO Conditions of Use policy provide for unrestricted use of non-proprietary data sets, provided that proper acknowledgment is given (http://www.noao.edu/noao/library/NOAO_Publications_Acknowledgments.html).

NOAO engineers, scientific and data management staff have access to all data for engineering or telescope/instrument monitoring and trending, as well as to help users.