Chapter 1
Introduction to NOAO Data

Version 1.1, May 2009

In This Chapter...

NOAO instruments routinely produce large quantities of digital data from a variety of science and survey programs directed by principal investigators (PIs). Data from some of these instruments are reduced and calibrated with automated pipelines, and both raw and reduced data products are published through the NOAO Science Archive. This chapter describes the scope of the data holdings in the Archive and the means to access them.

1.1 NOAO Data Holdings

NOAO collects a large quantity and variety of scientific data products, including images, spectra, catalogs, etc., from a large number of instruments on several telescopes that are deployed on three mountain tops on two continents. The primary repository for these data is the NOAO Science Archive (NSA). The data holdings can be categorized in a number of ways, including the source of the data, the type of data product, and the level(s) of processing required to produce them. Each of these views will be described in the following subsections. Since calibration pipelines exist for only a few instruments, most of the current archive holdings consist only of raw data products.

NOAO hosts two science data archives. The NOAO Science Archive for surveys (called Survey Archive) contains data products that were generated by NOAO Survey PI teams; it can be accessed through its custom Web interface.[1]

1. The Survey Archive data products can be accessed at http://archive.noao.edu/nsa/

The main NOAO Science Archive (called Archive), which contains all raw data, and calibrated data products for select instruments as produced by NOAO pipelines, is accessed through a Virtual Observatory (VO) compatible portal.[2] The NOAO VO Portal can also access all Survey Archive data products. The remainder of this chapter will describe the data holdings and access methods for the Archive (all data other than PI-produced survey products).

Warning: Users of NOAO data products should be aware that, with few exceptions, the data were obtained from observing programs that were designed and carried out by principal investigator (PI) teams, who obtained the data for the purpose of achieving their particular science objectives, and who obtained supporting calibration observations at their discretion. In addition, the quality of the data and the calibrations are strongly affected by the weather and other conditions that prevailed during the observing run.

1.1.1 Data Sources

The scope of the Archive data holdings can be measured by the variety of contributing data streams. Data from most instruments that are offered on NOAO facilities (with the exception of some visitor instruments) are captured and stored in the Archive. In addition, most data from partner institutions (i.e., observatories operated by consortia of institutions of which NOAO is a partner) are also captured, though in some cases only data obtained through the NOAO time allocation process may be stored in the Archive. The specific instruments and telescopes that contribute data to the Archive are listed in Table 1.1, along with the highest level of scientific data processing that is currently offered (see next subsection).

TABLE 1.1: Archive Data Holdings

Instrument	Type	Observatory/ Location	Highest Level Data Product[a]
Hydra	Optical fiber spectrograph	NOAO/Cerro Tololo	1
ISPI	IR imager	NOAO/Cerro Tololo	1
Mosaic-2 Camera	Optical imager	NOAO/Cerro Tololo	3
R-C Spectrograph	Optical spectrograph	NOAO/Cerro Tololo	1
4-m direct	Optical imager	NOAO/Kitt Peak	1

2. The NOAO VO Portal can be accessed at http://portal-nvo.noao.edu/

TABLE 1.1: Archive Data Holdings (Continued)

Instrument	Type	Observatory/ Location	Highest Level Data Product[a]
2.1-m direct	Optical imager	NOAO/Kitt Peak	1
Echelle Spectrograph	Optical spectrograph	NOAO/Kitt Peak	1
FLAMINGOS[b]	IR imaging spectrograph	NOAO/Kitt Peak	1
GoldCam	Optical spectrograph	NOAO/Kitt Peak	1
MARS	Optical spectrograph	NOAO/Kitt Peak	1
Mosaic-1 Camera[c]	Optical imager	NOAO/Kitt Peak	3
NEWFIRM	IR imager	NOAO/Kitt Peak	3
R-C Spectrograph	Optical spectrograph	NOAO/Kitt Peak	1
SQIID	IR imager	NOAO/Kitt Peak	1
0.9-m direct	Optical imager	SMARTS/Cerro Tololo	1
ANDICAM	Optical/IR imager	SMARTS/Cerro Tololo	1
1.5-m direct	Optical imager	SMARTS/Cerro Tololo	1
1.5-m Echelle	Optical fiber spectrograph	SMARTS/Cerro Tololo	1
1.5-m Cass Spectrograph	Optical spectrograph	SMARTS/Cerro Tololo	1
Y4KCam	Optical imager	SMARTS/Cerro Tololo	1
Goodman Spectrograph	Optical imaging spectrograph	SOAR/Cerro Pachón	1
OSIRIS	IR imaging spectrograph	SOAR/Cerro Pachón	1
SOI	Optical imager	SOAR/Cerro Pachón	1
Hydra/Bench	Optical fiber spectrograph	WIYN/Kitt Peak	1
Mini-Mosaic	Optical imager	WIYN/Kitt Peak	1
SparsePak	Optical IFU spectrograph	WIYN/Kitt Peak	1
WHIRC	IR imager	WIYN/Kitt Peak	1

a. See Table 1.2 on page 1-4 for a description of the data product levels.
b. Offered on both the 4-m and 2.1-m telescopes at KPNO.
c. Also offered on the WIYN Observatory 0.9-m telescope.

1.1.2 Processing Levels

The basic data files that are generated by the instruments in the observing environment are referred to as *raw* data products. Data from certain instruments have been processed to remove the instrumental signature, and to provide some type of calibration. Such advanced, higher-level science data products, along with concomitant data, are also available from the archive; detailed descriptions of these data products may be found in subsequent chapters, which are specific to each instrument. The degree to which data have been processed is expressed in terms of *levels*, as described in Table 1.2.

TABLE 1.2: Levels of Data Processing

Level	Description	Example Products
1	Raw data, as it is formatted by the data collection software at the telescope	Raw Mosaic camera frames, including overscan; may include approximate WCS predicted from model.
2	Single-frame reduced data	Single Mosaic camera images with instrument signature removed; WCS calibration applied; image possibly re-projected.
3	Combinations of two or more frames	Combined and/or tiled images of a region of sky derived from multiple, spatially overlapping exposures with the same filter. Artifacts and gaps between detectors in a focal-plane array are identified and removed.
4	Measurements from one or more level-3 processed data frames	Star catalogs, including position, magnitudes, colors; identifications of moving objects; time-series of object brightnesses.

Note that not all data products can be usefully described in terms of processing level, including exposures obtained to support calibration and ancillary data that are described in the next subsection.

1.1.3 Types of Data Products

The Archive contains a variety of data products that are relevant for a scientist end-user. The following categories of products are part of the vocabulary used throughout this *Handbook* to describe NOAO data. Generally, each type of data product (apart from metadata) is stored in its own file.

At present there are only rudimentary ways of associating reduced science data files with related files such as the original raw data and data quality masks. The value of the header keyword `DTNSANAM` will be identical for raw science data files and products derived directly from them. For stacked images and other products with more than one parent, `DTNSANAM` will correspond to the first raw file that was used in creating the stack. Spatial coincidences of the instrument FoV may be useful, but bear in mind that the WCS is updated from raw to reduced data products.

Science Data

Frames of data obtained with an instrument of an astrophysical source are generally referred to as *science* data. Such data include images and spectrograms, and consist primarily of contiguous blocks of pixel values from one or more detector arrays in the instrument. The science data are packaged with headers of metadata (see below) within a file; these metadata describe essential details of the science data. Science data products may also include high-level products, such as catalogs or time-series, that may or may not contain pixel values.

Science data in raw form are produced from the instruments listed in Table 1.1. Science data in processed form are generated either by dedicated processing pipelines that are part of the NOAO End-to-End (E2E) system, or they are provided by PI teams to the Survey Archive as advanced science products. In the first case, the data products are described in this *Handbook*. In the second case, PI teams provide the documentation necessary to understand the archived data products; links to this documentation are provided on the NSA Survey Archive Web site.[3]

Metadata

Metadata, or data that describe data products, are collected or created at various points in the planning, acquisition, and processing of science data. Most of the metadata that are available to a user are stored in the headers of science data files as keyword–value pairs. Metadata fall into a variety of categories, but perhaps the most practical for the end-user are the following (see Shaw 2007):

Telescope/Environment. These metadata include telescope name and configuration information, celestial coordinates of the telescope pointing, airmass/ zenith distance, tracking information, ambient temperature and wind speed, ADC use, etc.

Instrument/Detector. These metadata are required for generating Level-2 products from Level-1 raw data and include the instrument configuration, detector/amplifier designations, gain settings, operating temperatures, photo-active regions of the image, filter/disperser or other optical elements, and pixel scale.

Science/Calibration. These metadata include the World Coordinate System (WCS) mapping, field of view, resolution or delivered image quality, start of the exposure and its duration, effective bandpass and sensitivity of the system, brightness units and dynamic range, background level, and transformation to a standard photometric system.

Pedigree/Provenance/Quality Assurance. These metadata include the observer name and proposal identifier, version identifier of the control software, version of the data processing/calibration software, the processing steps performed, calibration reference files used, etc.

3. See the NSA Survey Archive Web site at http://archive.noao.edu/nsa/.

Critical metadata in raw science data files may be missing or have incorrect values. At present, no mechanism is in place to validate and correct corrupted metadata, apart from the proposal identifier (`PROPID`) and the date of the observation (the *date* portion of `DATE-OBS`). However, certain of the metadata in the headers of calibrated data, notably the WCS information and the filter name, are corrected during the course of pipeline processing. Metadata *added* by the pipeline during processing are not in question.

Calibration Data

Calibration data in raw form are often very similar to (or indistinguishable from) science data products. These include exposures obtained for bias (zero duration), dark (finite duration with shutter closed), flat-field (including flats taken of the twilight sky), and standard star calibration. The *Master Calibration* reference files are usually generated by the processing pipeline from exposures obtained during the same observing run as the science data. The methods by which master calibration files are generated and how they are applied during processing is specific to each instrument, and they are discussed in detail in subsequent chapters.

Concomitant Data

Concomitant data quantify the detailed quality of the science data, generally at the pixel level, and are generated during the course of pipeline processing. These data are intimately related to the scientific interpretation of the data products. Examples of concomitant data include data quality masks (DQMs) which codify one or more pathologies that affect (usually) small areas of the image, exposure masks that quantify the exposure level as a function of position in an image mosaic, and variance arrays that quantify the uncertainty associated with pixel values.

Ancillary Data

Most ancillary data are generated contemporaneously with science data, but are not part of the science or engineering data streams. Some ancillary data, such as preview images, are generated by the processing pipeline. Ancillary data are used to establish the context of an observation, or to enable a rough, qualitative assessment of science data quality or applicability. Examples of ancillary data include observation logs and environmental data such as that from seeing and atmospheric transparency monitors, images from all-sky cameras at or near the observing site, and weather satellite images.

At present no contemporaneous weather data are available from the Archive. However, weather data and images from all-sky cameras are available for KPNO;[a] statistical weather data are also available for CTIO,[b] as well as historical cloud coverage information[c] dating back to 1975.

a. KPNO weather and sky conditions: http://www-kpno.kpno.noao.edu/Info/Mtn_Weather/
b. CTIO weather: http://www.ctio.noao.edu/environ/environ.html
c. CTIO historical cloud coverage: http://www.ctio.noao.edu/site/phot/sky_conditions.php

1.1.4 Data File Nomenclature

Raw science data as obtained at the telescope are captured from the observing environment by the Data Transport System, and persisted as FITS files with unique names. Each name consists of a prefix, which is a two-letter code for the observing site (one of "`cp`," "`ct`," or "`kp`"), and a numeric serial number that increments with each successive observation (though not necessarily monotonically). Calibrated data files are prefixed with a code for the archive center where they were generated ("`tu`" for Tucson, "`ls`" for La Serena), plus a serial number that increases as each calibration data product is made available for Archive ingest. When users obtain data products through the Portal, the filenames are further prepended with a code for the Archive version number (currently, "`NSAR3_`"). The files are postpended with the file type, which will also reflect the type of compression applied to the file, e.g., "`.fits.gz`" for a g-zipped FITS file. ***Note that these file names are not related in any way to the names assigned by observers to data files at the telescope.***

Participants in observing programs should note that the filenames they assigned via the data taking system are preserved in the data headers, as the value of the keyword `DTACQNAM`. For some instruments this name includes the file pathname. After raw data are retrieved from the Archive, it is possible to restore the file names to those assigned in the observing environment using a script, an example of which can be found in the VO Portal contributed tools area.[a]

a. Available at http://nvo.noao.edu/noaonvo/contrib.shtml

1.2 NOAO Data Formats

Most data from the Archive will be provided to users in FITS format (FITS 2008), which is the well known data interchange format for astronomy. As most users know, the FITS standard allows for many complex representations of data. The general features of FITS file structures that are used for data in the Archive are described here, but implementation details for each instrument are deferred to subsequent chapters. Note that most data products as published by the Archive have been compressed in one way or another, which minimizes the time required for users to download their data. Depending upon the software that is used to examine or analyze the data, users may need to uncompress or otherwise transform their data after download.

Data files from the Archive are stored in one of a few compressed formats. Science images and calibration reference files are currently compressed with ***gzip*** (see http://www.gzip.org/); most systems have a native ***gunzip*** or equivalent program for decompressing the files after download. FITS tile compression (Pence et al. 2009) is planned for a future release of the Archive. Data quality masks are stored as a pixel list in BINTABLE extensions in a FITS file (this is actually one form of FITS compression). Some software (such as IRAF) treats these files as logical images transparently to the user; other software (e.g., DS9) currently requires that these masks be transformed to FITS images (or image extensions) before they can be imported.

1.2.1 Imaging Data

In modern imaging cameras it is common to use multiple detectors, arranged in a contiguous array to cover a large fraction of the available focal plane of a telescope focal station. The detectors in these *focal plane arrays* (FPAs) are operated simultaneously, and are read out in parallel. (Some detector electronics allow for parallel read-out of portions of a single detector.) The signal from each of the detectors (or portions thereof) is generally stored in raw form as a separate image, along with the associated metadata; all of the images from a single exposure with the FPA are collected into a single FITS file for convenience. On the other hand, images from cameras with single CCDs, or composite images from, e.g., multiple, overlapping pointings that have been combined into a single image, are stored in simple FITS images in the Primary Header-Data Unit.

A schematic representation of the structure of NOAO FITS files for storing data from FPAs is given in Figure 1.1. Generally, the pixel values for multiple components of an image are stored in image extensions, rather than in the primary header/data block at the beginning of the file. The headers of each image extension contain metadata that pertain specifically to that image (e.g., the WCS information). Global metadata that apply to all the images (e.g., filter name, exposure duration, etc.) may be stored in the primary header, in which case the keyword `INHERIT = T` will be found in the extension headers.[4]

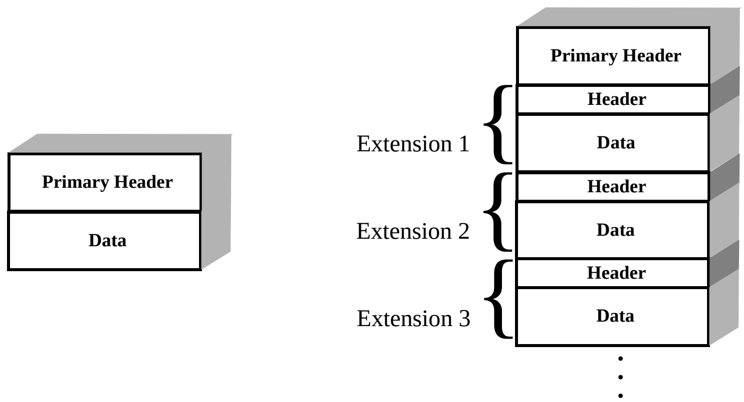


FIGURE 1.1: Schematic of the structure of a simple FITS file that stores a single image (*left*) and a Multi-Extension Format file that stores one or more components of an image (*right*). Metadata that apply to all extensions are often stored in the Primary Header.

1.2.2 Spectroscopic Data

The storage format for raw spectrograms is conceptually similar to that for imaging cameras, i.e., pixel values from multiple detectors are stored in simple FITS or FITS MEF format. At present, no Level-2 or higher (i.e., processed) data products for spectra are offered in the Archive.

1.2.3 Concomitant Data

Concomitant data are always stored in compressed form, although they are logically equivalent to ordinary FITS images. Since DQMs usually contain only a small fraction of pixels with values different than zero (zero indicates no data quality problem), it is much more efficient to represent the image internally as a list of regions containing particular nonzero values. These lists are stored as tables in FITS BINTABLE extensions: one extension per image, similar to the structure illustrated in the right panel of Figure 1.1. Similarly, exposure maps mostly consist of large regions with a common effective exposure time; these are also stored in BINTABLE extensions.

4. In practice, the most critical metadata in the primary header often also appear in the extension headers because not all data analysis systems support the FITS `INHERIT` convention, which is described at http://fits.gsfc.nasa.gov/registry/inherit.html.

1.2.4 Ancillary Data

Ancillary data come in a variety of formats, depending on the specific product. At present the only ancillary data offered through the Archive are down-sampled image previews. These data are generated from science or Master Calibration images, and are stored in PNG[5] format. Note, however, that these preview images are encapsulated in a FITS file in extensions of type `FOREIGN`[6] (Zárate, Seaman & Tody 2007), and must be unpacked before they will be of use to most application software.

1.3 ACCESSING NOAO DATA

Access to and use of nonproprietary data is available to everyone, consistent with the *Conditions of Use* policy described in Section 1.4, and requires neither registration nor login. Authorized users who wish to download *proprietary* data may do so only after registering with the Virtual Observatory (this need be done only once) and logging in to the NOAO VO Portal. The VO provides secure *authentication* of users by issuing an electronic security certificate to their browser when they log on to the VO; this mechanism enables access to any service or data in the VO that accepts single sign-on certificates. The Archive makes use of such certificates to provide you with access to all proprietary data to which you are authorized.

The NOAO VO Portal can only be used successfully with compatible browsers, which currently are Mozilla 1.7 or greater, Netscape 9.0 or greater, and Firefox 1.4 or greater. Use with any other browsers will likely result in missing or erroneous functionality that may or may not be apparent to the user!

1.3.1 User Registration and Login

PIs wishing to retrieve *proprietary* data from the Archive must complete a one-time registration process with the VO to obtain a user ID and password. Retrieving nonproprietary data does *not* require registration. Investigators who have been awarded observing time on NOAO or partner facilities will receive an invitation via email (from `archive-noreply@noao.edu`) to register with the VO. This email will be sent to PIs prior to the actual observing run.

5. http://www.libpng.org/pub/png/
6. http://fits.gsfc.nasa.gov/registry/foreign.html

Registration is a straightforward process and consists of the following steps:

1. **View the invitation email** from `archive-noreply@noao.edu` (with the subject line "NOAO Science Archive Registration Invitation"), and enter the registration URL into your compatible browser.[7]
2. **Complete the Portal registration form.** This form will be prefilled with information that is on file with NOAO: your name, institution, and email. If any of this information is incorrect or out of date, now is the time to fix it. If you have previously registered with the VO, click the I'm already registered button and skip to step 5. Otherwise, click Start registration with VO, and continue with step 3.
3. **Complete the VO registration form.** Information from the Portal registration form will be transferred to this form. Fill in the remaining information, including your choice for *username* and a *password*. **Note:** *it is important to remember your VO username and password*. Click Submit to advance to the confirmation phase of the VO registration process.
4. **Confirm your identity.** A confirmation token will be sent to the email address you provided in step 2. Copy this token from the email message into the dialog box on the form.[8] Clicking Submit will first display a congratulatory page from the VO, and within about 20 seconds your browser will be redirected to the VO login page.
5. **Log in to the VO.** Your Portal registration will not be complete until you perform this step. Click the Login button after entering your username and password (selected in step 3 for first time registrants). Your browser will automatically be redirected to the Portal **sky** view.

Returning users need only log in to the NOAO VO Portal at http://portal-nvo.noao.edu/ by selecting **Sign in** from the menu bar, using the username and password they established during the registration process.[9] Users will automatically have access to all proprietary data from all NOAO programs for which they have been authorized. Users who have previously registered with the VO,

> If an error message appears in your browser indicating that a "Secure Connection Failed" or that "sso.us-vo.org uses an invalid security certificate," you will need to add a security exception in your browser. Adding such an exception is easy, but the details are browser-specific.

but who have not been invited to do so by the NOAO Science Archive via the aforementioned invitation email, will not have secure sign-in privileges with the NOAO VO Portal.[10] They may still use the Portal without logging in, however.

7. Be sure to enable pop-ups, at least for the site http://portal-nvo.noao.edu/.
8. Alternatively, you may click the direct URL in your email message **provided that the browser you are using for email is compatible with the Portal.**
9. Attempting to follow the registration link in the invitation email *after you have registered* will result in an error.

1.3.2 VO Portal Interface

The NOAO Science Archive was designed to be compatible with the VO, and users can access NOAO data through a variety of tools and services that are VO-compatible. One such access tool is the NOAO VO Portal, which is composed of a rich set of services that are accessed through a Web interface, as shown in Figure 1.2. The initial view is that of an all-sky Aitoff projection of the celestial sphere, with the footprints of various surveys indicated with shaded regions (in color). The suite of functionality includes:

- panning and zooming on the sky to identify surveys that overlap spatially with regions of interest;
- searching for datasets from selected repositories that are connected to the VO, via a spatial region (e.g., a cone search), epoch of observation, or type of data (image, spectrum, etc.);
- staging data in the VO for retrieval; and
- cross-matching a target location against major VO catalogs.

The NOAO VO Portal is particularly well tuned to retrievals of NOAO data from the Archive. The Portal's browser-based graphical user interface enables the generation of rich queries of the underlying database through mouse actions and simple forms-based user input. The auto-generated SQL (the structured query text) may be saved for later use in advanced queries, or for access to NOAO data through mechanisms outside the browser-based NOAO VO Portal (see "Special Case: Co-I access to proprietary data" on page 1-21).

10. Users who are not NOAO PIs and who wish to have an account for secure sign-in to the NOAO VO Portal may email their request to vohelp@noao.edu.

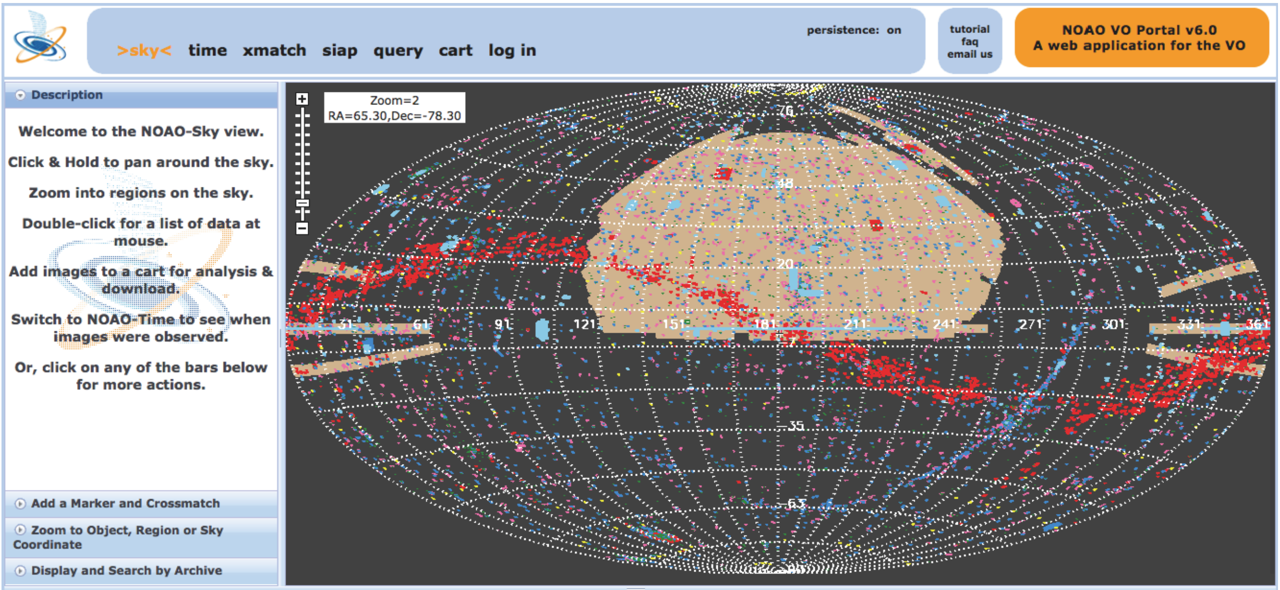


FIGURE 1.2: Web interface for the NOAO VO Portal, showing an all-sky view of the footprints of major VO holdings that can be accessed (*color shaded areas*). Additional markers will appear for NOAO datasets that are identified in the user's cart.

1.3.3 Obtaining Data via the Portal

Obtaining data from the Archive is straightforward and consists of the following steps:

1. **Register as a user of VO resources** (one time only, and only for access to proprietary data).
2. **Access the NOAO VO Portal** at http://portal-nvo.noao.edu/ (retrieving proprietary data requires login).
3. **Construct a query** for data that meet your science criteria.
4. **Select data products** of interest from the Data Grid, and add them to the cart.
5. **View the cart** contents, and select those data products of interest for retrieval.
6. **Stage the selected data** to the Archive ftp area.
7. **Download the data** to your machine from the ftp area.

The following examples illustrate the above process for some common cases. It is assumed throughout that the user has registered with the VO (if needed), has opened the Portal interface in a compatible browser, and has selected the **query** panel from the top menu bar.

Retrieve Nonproprietary Science Data

Construct a Query. This example describes how to retrieve nonproprietary reduced data from the observing program 2002B-0007, obtained in November 2005. The approach for building the needed query is to use pull-down menus to construct an approximate query, and then make refinements to the search criteria. At the **Query** panel, perform the following steps, as illustrated in Figure 1.3 on page 1-15:

1. Select the **Default Simple Query** view from the tab at the upper left of the Query panel.
2. Choose the following from the pull-down menus at the upper right:
   - Search Type: Search All Data
   - Data Type: Reduced Images
   - Example Queries: Get Reduced Data and Masks by PropID
3. Type the proposal ID into the text box (in this case "`2005B-0007`").
4. Enter a Date-Obs string to match any date in the year of the observation (in this case, "`2005%`").
5. Select the desired type of reduced data from the pull-down menu (in this case, InstCal).
6. Check the **Public Data** box to match only nonproprietary data.
7. Press the **Update Query** button (note how the text in the left panel changes).
8. Press the **Submit Query** button at the top left.

It is of course possible to select by other criteria rather than the proposal ID, for instance on a range in RA and Dec (using the "BETWEEN" operator in the pull-down to the left of the text boxes). In this case, one would leave the "Proposal ID" text box blank.

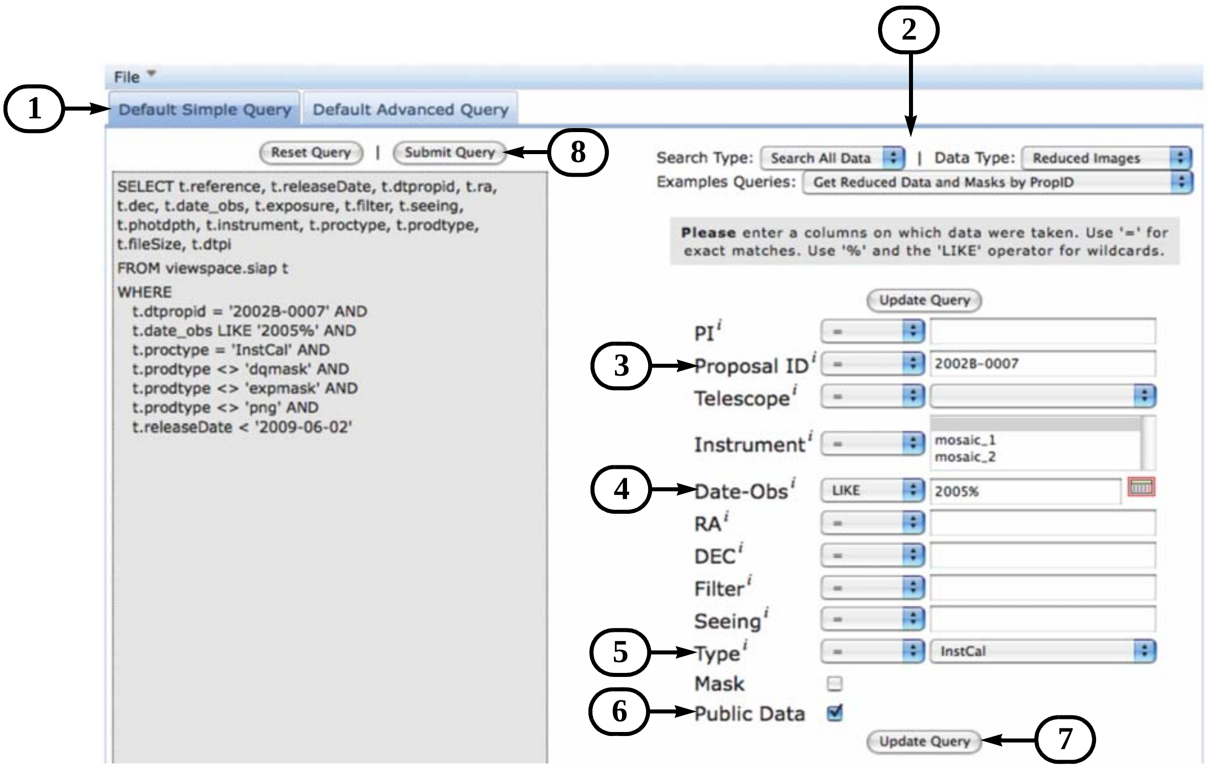
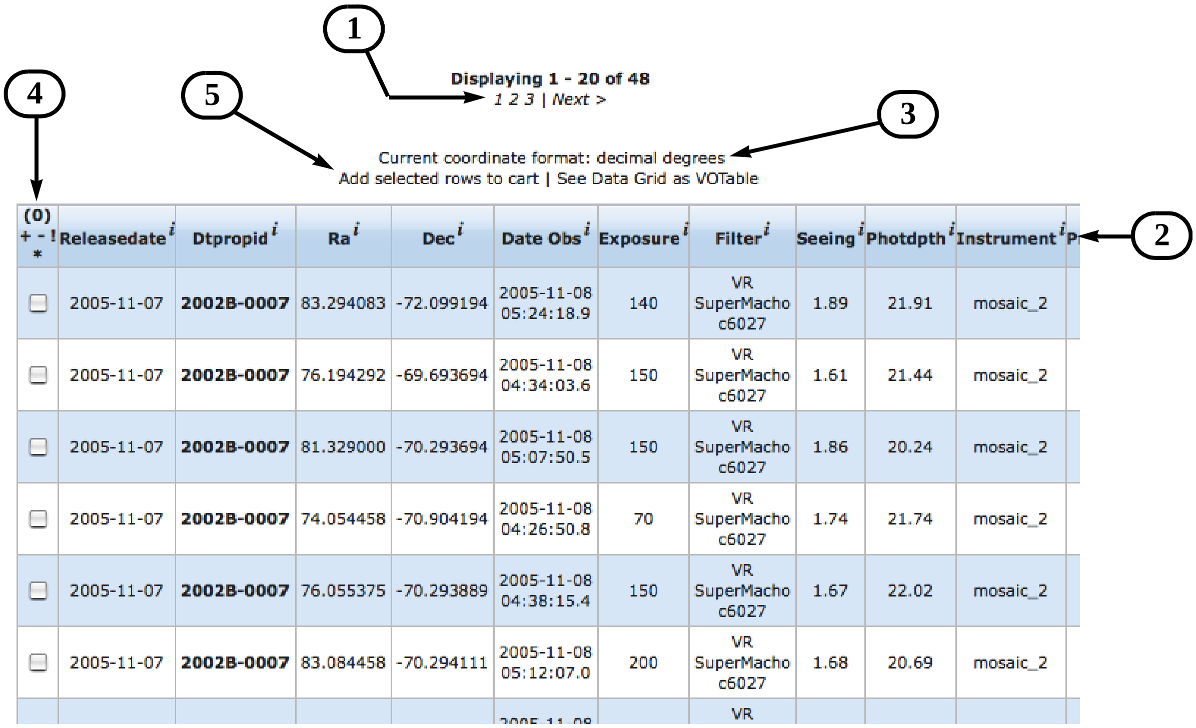


FIGURE 1.3: Portal **Default Simple Query** panel, showing the input text boxes (*right*) and the resulting query string (*left*) that can be updated with values in the boxes by pressing the **Update Query** button (*bottom*). Numbered callouts are explained in the text. Pressing the **Submit Query** button (*top*) will send the query to the Archive, the results for which are displayed in the panel described in Figure 1.4 on page 1-16.

The pending query that will be passed to the Archive database is shown on the left portion of the panel in Figure 1.3. The results will appear in the *Data Grid* panel just below the query form, shown in Figure 1.4 on page 1-16; you may want to adjust the Data Grid's viewable area with the frame boundary widget by dragging upwards the bar that separates the Data Grid from the query panel.

Select Data. Each row in the Data Grid corresponds to one unique data product. There are up to 20 rows on each page, and pages can be navigated through links near the text labeled (1) in Figure 1.4. Various attributes of the datasets are given in the header row, labeled (2), and a brief description of the attribute is displayed by clicking the "*i*" at the upper-right of each column name; clicking the header text of any column sorts the rows by value in that column. The coordinates for the dataset are given by the Right Ascension and Declination columns, which by default are in units of decimal degrees; to view the coordinates in sexagesimal format, click the text labeled (3). You may select datasets of interest by doing one or more of the following actions in the column labeled (4):

- check the box to the left of individual datasets;
- click the plus (+) sign to select all rows on the displayed page (rows will remain checked even if you display another page);
- click the asterisk (*) to select all rows on all pages;
- click the minus sign (–) to de-select all rows on the displayed page; and/or
- click the exclamation point (!) to de-select all rows on all pages.



(0) + - ! *	Releasedate	Dtpropid	Ra	Dec	Date Obs	Exposure	Filter	Seeing	Photdpth	Instrument	P
☐	2005-11-07	**2002B-0007**	83.294083	-72.099194	2005-11-08 05:24:18.9	140	VR SuperMacho c6027	1.89	21.91	mosaic_2	
☐	2005-11-07	**2002B-0007**	76.194292	-69.693694	2005-11-08 04:34:03.6	150	VR SuperMacho c6027	1.61	21.44	mosaic_2	
☐	2005-11-07	**2002B-0007**	81.329000	-70.293694	2005-11-08 05:07:50.5	150	VR SuperMacho c6027	1.86	20.24	mosaic_2	
☐	2005-11-07	**2002B-0007**	74.054458	-70.904194	2005-11-08 04:26:50.8	70	VR SuperMacho c6027	1.74	21.74	mosaic_2	
☐	2005-11-07	**2002B-0007**	76.055375	-70.293889	2005-11-08 04:38:15.4	150	VR SuperMacho c6027	1.67	22.02	mosaic_2	
☐	2005-11-07	**2002B-0007**	83.084458	-70.294111	2005-11-08 05:12:07.0	200	VR SuperMacho c6027	1.68	20.69	mosaic_2	

FIGURE 1.4: Portal Data Grid, showing query results from Figure 1.3. Numbered callouts are explained in the text. Clicking the link labeled (5) will add all *selected* datasets to your cart, the result of which is shown in Figure 1.5 on page 1-17.

After selecting the datasets of interest, click the link labeled (5) to add them to the cart. At this point, you may either query for additional data (e.g., search for higher-level data products or Master Calibrations from this same dataset) and add them to the cart, or go directly to the **cart** view.

View the Cart. Data in the cart may be examined or used for a number of purposes using the Portal, such as viewing the spatial location on the **sky** view. To download the datasets in the cart, select the **cart** entry in the top menu bar.

To actually stage the data for ftp download, select datasets of interest in the cart Data Grid pane using much the same process as described above (i.e., checking boxes in the first column individually or in aggregate), as shown in Figure 1.5.

Stage the data. The status (1) of the staging process is updated continuously once the Start staging text (2) is clicked. Alternatively, individual data files may be downloaded directly by clicking the Retrieve text (3) in the **URL** column. Staging progress for individual files is displayed as progress bars (4). To view the stage area (see Figure 1.6 on page 1-18), click the Show stage area text (5). The staging progress is summarized in the **Image stage status** box, with files moving from being *enqueued*, to *staging* (i.e., actively transferring data from mass store to the temporary staging area), to *staged* (transfer complete).

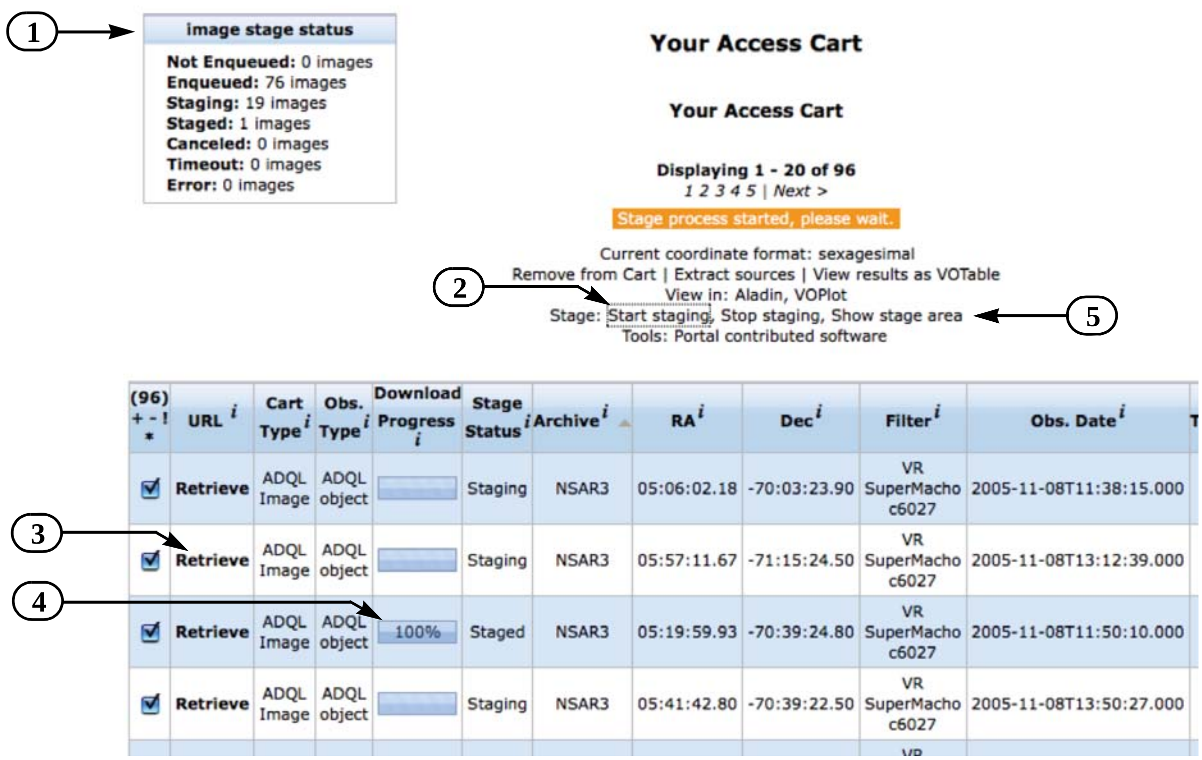


FIGURE 1.5: Portal Access Cart. Numbered callouts are explained in the text. The status (1) of the staging process is updated continuously, once the Start staging text is clicked. You may view the stage area (see Figure 1.6 on page 1-18) by clicking the Show stage area text.

If staging takes an unusually long time (perhaps hours), the status may change to “Error.” Problems may have their cause in hardware or software failures in the Archive, network problems, or a staging area that has reached full capacity. The system attempts to recover from such errors, so the status should be monitored for some time before seeking help from the NOAO Help Desk. If staging fails, the affected files are noted in the Error row at the top; the word “Error” then becomes a link which, if followed, will ask you if these files should be restaged. Note that if you log out of a Portal session before staging completes, you can resume monitoring the progress at any time by logging in again and clicking the Show stage area text on the **cart** view.

The stage area (Figure 1.6) shows which products have been (or are being) successfully staged for download, as well as other information needed for access. The files that reside in the stage area are listed on the right (1) with the standard file nomenclature (see “Data File Nomenclature” on page 1-7); if staging is not yet completed, the list may be updated manually by clicking the Update my stage area text (2).

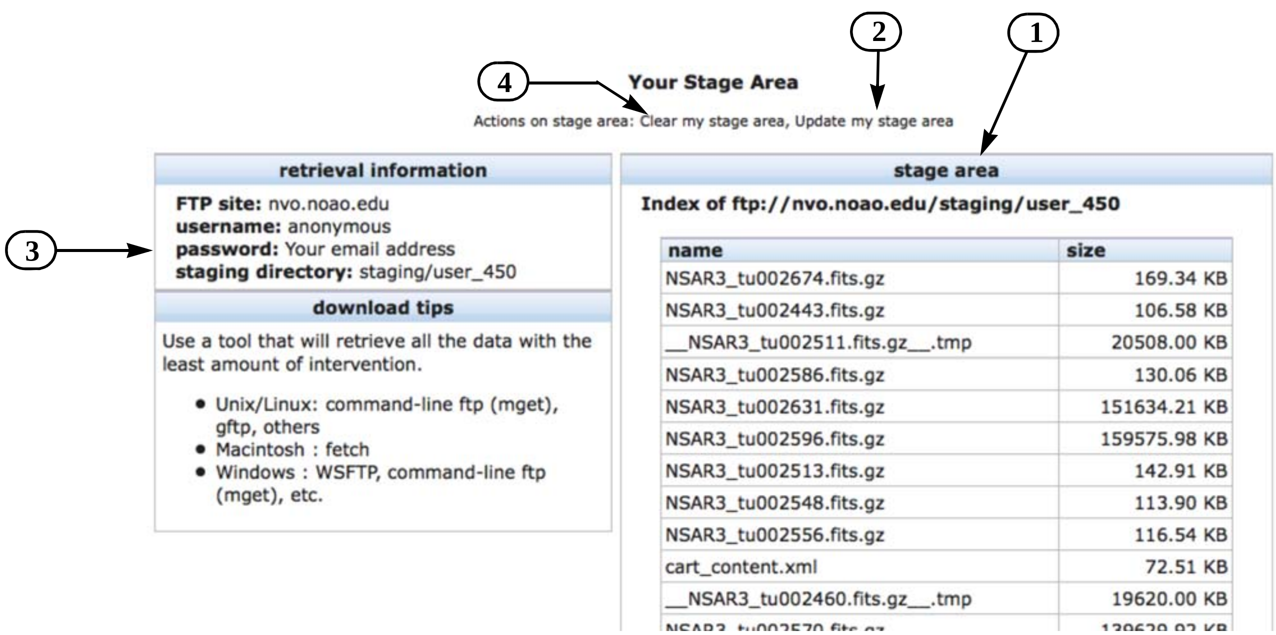


FIGURE 1.6: Portal Data Staging Area. Once the data files have all been staged, use a download client (e.g., ftp) to access the ftp server at nvo.noao.edu.

Download Data. Once all the data have been staged, they are ready to be downloaded from the Archive ftp site at ftp://nvo.noao.edu via a file transfer client of your choice (suggestions are offered in the **download tips** box). The login information for your staging area is given in the **retrieval information** box (3), including the path to your data. The staging area is a limited resource, and older files are deleted periodically to make room for new retrievals. Thus, prompt retrieval of your files is advised. To give yourself room for more downloads, you can delete your old data from the stage area by clicking the Clear my stage area text (4).

Data in the NOAO staging area have a limited shelf life, and may be deleted after one week if the staging area nears its maximum capacity. Therefore, prompt retrievals are advised.

Obtain Data from an Observing Run

In the following example, the raw data from the NOAO Survey program 2005B-0045 (PI: C. Miller) will be retrieved using a common type of query, as the PI might be expected to do. These data were obtained with the Mosaic-2 camera at CTIO during 24–30 April 2008.

Only PIs of observing programs are authorized to retrieve data during the proprietary period, and they must log in to the portal using their assigned password in order to stage their data. Queries for proprietary data from non-PIs will populate the Portal Data Grid, but these data cannot be selected and placed in the cart for retrieval. PIs may transfer authorization to co-investigators (Co-Is) using security certificates (see “Special Case: Co-I access to proprietary data” on page 1-21). To query for nonproprietary data only, check the **Public Data** checkbox.

As in the previous example, an approximate query can be generated with the **Default Simple Query** panel, as shown in Figure 1.7 on page 1-20, and refined with the following steps:

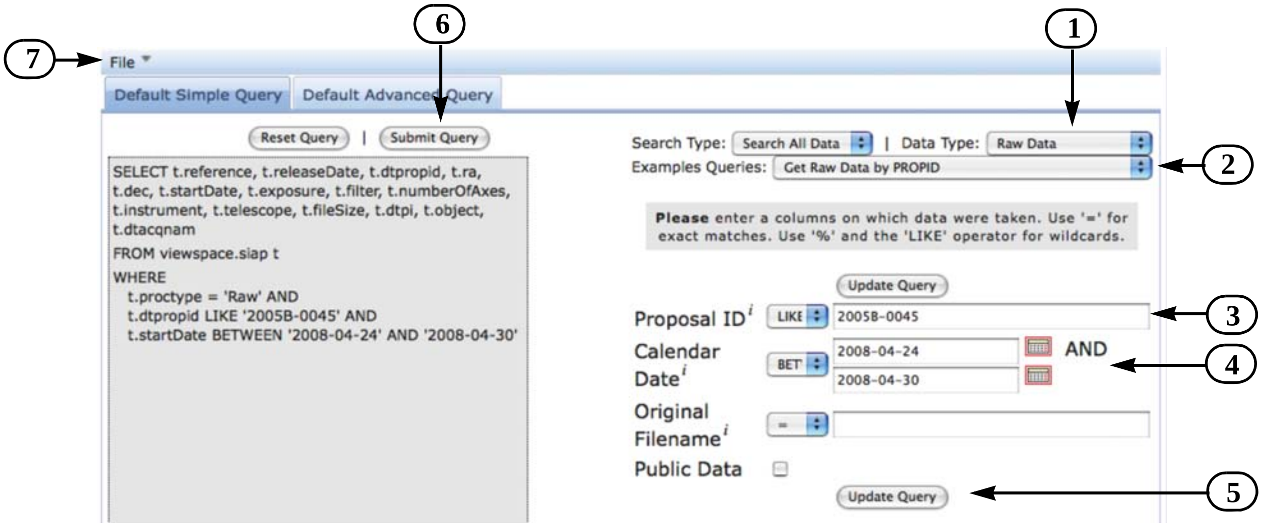


FIGURE 1.7: A Custom Query to obtain data from an observing run can be generated by modifying the **Get Raw Data by PROPID** example query.

1. Select **Raw Data** from the Data Type pull-down menu at the top-right.
2. Select **Get Raw Data by PROPID** from the Examples Queries pull-down menu at the top-right to build a query string that is similar to what will be needed in this example.
3. Refine the query string by entering the observing proposal number in the Proposal ID text box (in this case, "`2005B-0045`").
4. Select the **BETWEEN**[11] operator next to the **Calendar Date** label, and enter the first and last dates of the observations in the text boxes (`YYYY-MM-DD` format), or use the calendar widgets.
5. Press the **Update Query** button. The refined query will then be displayed in the left portion of the pane.
6. Either press the **Submit Query** button to execute the pending query, or alternatively...
7. Store the query text in a file on your machine by selecting **File**→**Save Query As...** from the pull-down menu at the upper left.

The results of the search will appear in the Data Grid panel, beneath the Query panel. Staging data from the cart and retrieving data proceeds exactly as in the prior example, "Retrieve Nonproprietary Science Data" on page 1-14. Recall that the file names may be restored to those assigned in the observing environment using a script: see "Data File Nomenclature" on page 1-7.

11. Note that the **BETWEEN** operator requires that the first argument be strictly less than or equal to the second argument.

The Portal currently limits the number of datasets that can match a single query to 1000 files. Although the number of files from even a moderate-length observing run with NEWFIRM will likely exceed this limit, it is necessary to keep the required disk space in the NOAO staging area to a reasonable level when servicing multiple users. Users with large numbers of datasets from lengthy observing runs will need to break their retrievals into individual requests that do not exceed the limit by, e.g., restricting queries to a time window of a few days.

Special Case: Co-I access to proprietary data. If you are a co-investigator on a program and wish to retrieve the data during the proprietary period, the PI can give you authorization to do so in a relatively secure way *without sharing his/her VO login information*.

1. Download a shell script for command-line file retrieval. Go to http://nvo.noao.edu/noaonvo/contrib.shtml or click the Portal contributed software text near the top of the Portal **cart** view. Download the secure_get.sh file to your machine, and ensure that you have permission to run it as an executable.
2. Using the Portal, build and submit a query that selects the data products of interest, bearing in mind the limit on the maximum number of files that can be staged in a single request.
3. In the Data Grid (see Figure 1.4 on page 1-16), right-click the See Data Grid as VOTable text, and save the file to your local disk (choose XML format, which is the default). This file contains information in VOTable format that is necessary to access the files currently in the Data Grid. You will need to generate more than one of these tables if you require more than one query to specify all the data files of interest.
4. Obtain a VO security certificate file from the PI.
5. Run the `secure_get.sh` script to retrieve the files prior to the certificate's expiration date.

The instructions for running the script are given at the top of the file. The command-line syntax (with a certificate) is:

```
secure_get.sh infile --cert CERT --pass PASS [--dir DIR]
```

where *infile* is the name of the VOTable containing the search results, *CERT* is the name of the certificate file, *PASS* is the password, and (the optional) *DIR* is the path to the destination directory. If you have more than one VOTable of search results, you will need to run the script multiple times.

In order for the PI to provide a secure certificate to a collaborator, he/she must do the following:

1. Log on to the VO at https://nvologin1.ncsa.uiuc.edu/protected/welcome.
2. Choose a validity lifetime for the certificate.
3. Choose a Package Key (i.e., a password) for the certificate.
4. Choose PEM format.
5. Click the **Download Certificate** button. A file called "`geteec`," which is the certificate file, will appear in your download directory.
6. Send this file to your collaborator, and tell them (preferably in a separate email) the certificate password and expiration date.

Obtain Calibration Reference Files

On occasion, users may wish to examine in detail the quality of the pipeline instrument signature removal, to recuperate raw data using custom software, or to create custom calibration reference files from other observing runs. In these cases, it is helpful to retrieve either the master calibration reference files or the raw files from which they were created. Retrieving these files is similar to other nonproprietary data retrievals, but the search criteria are sufficiently different that efficient queries require the **Default Advanced Query** panel of the Portal **query** interface. This example illustrates a search for NEWFIRM raw dark files for potential use in calibrating data from another observing run, for which the selection criteria include integration times near 60 s duration and observing dates at KPNO in early May 2009.[12]

As in prior examples, an approximate query is generated and then refined. But unlike the **Default Simple Query** panel, essential attributes in the advanced query string will be edited by clicking highlighted text[13] and then using custom selection menus and text boxes to define or modify attributes of the query. The approximate initial query, shown in Figure 1.8 on page 1-23, is generated with the following steps:

1. Select the following from the pull-down menus at the upper-right:
   - Search Type: Search All Data
   - Data Type: Raw Data
   - Example Queries: Get All Master Calibration Data by 7-day Date Range
2. Click the blue startDate text (2), which will open the attribute menu.
3. Click the **Modify t.startDate** text (3), which will open widgets to edit the value.
4. Select the BETWEEN operator from the pull-down menu (4) and enter the start and end dates for the search range (`YYYY-MM-DD` format) or use the calendar widgets.
5. Press the **Modify Condition** button (5) to update the query text with the new date range.

---

12. Consult the KPNO telescope schedule at http://www.noao.edu/kpno/forms/tel_sched/ for dates when NEWFIRM was scheduled.

13. See http://cas.sdss.org/dr7/en/help/howto/search/simplequery.asp for a tutorial on the similar SDSS Query Builder.

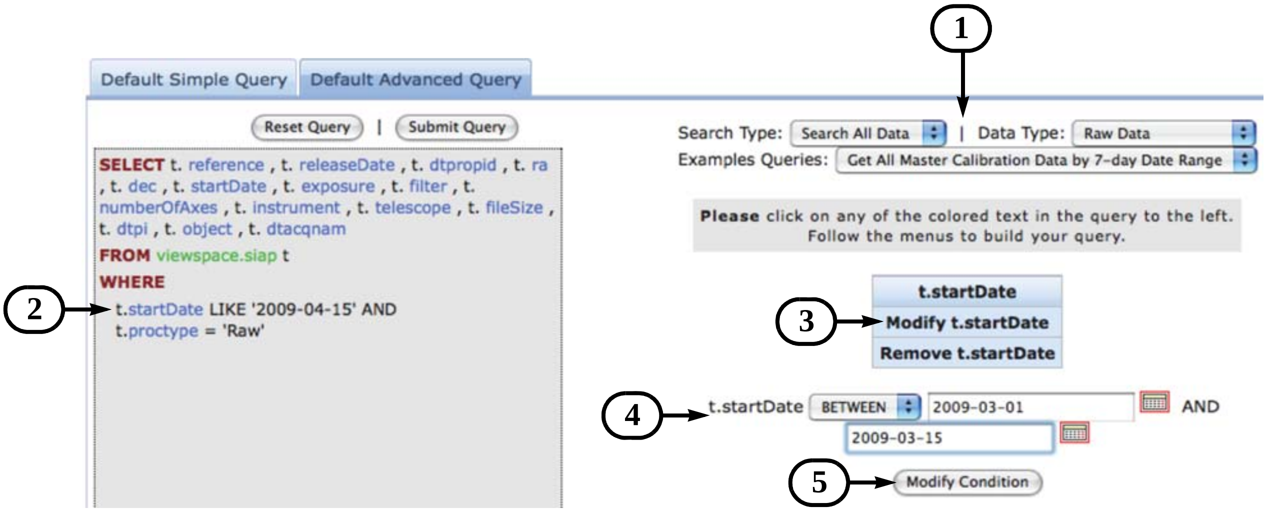


FIGURE 1.8: Portal **Default Advanced Query** panel, showing the editable query string (*left*) and input boxes (*right*). After clicking the blue *startDate* text in the query string, it can be updated with values in the boxes by pressing the **Modify Condition** button (*bottom*). Numbered callouts are explained in the text.

The next step is to add a new condition to restrict the query to data from the NEWFIRM instrument. This process is illustrated in Figure 1.9 on page 1-24, and consists of the following steps:

1. Click the **WHERE** text.
2. Click **Add Condition**.
3. Scroll down the Meta Columns selection box, and select instrument.
4. Type "`newfirm`" into the text box.
5. Press the **Add Condition** button to update the query string.

Now follow the above steps to add a new condition, this time for obstype with the value "dark." The result of adding this condition is illustrated in Figure 1.10 on page 1-24.

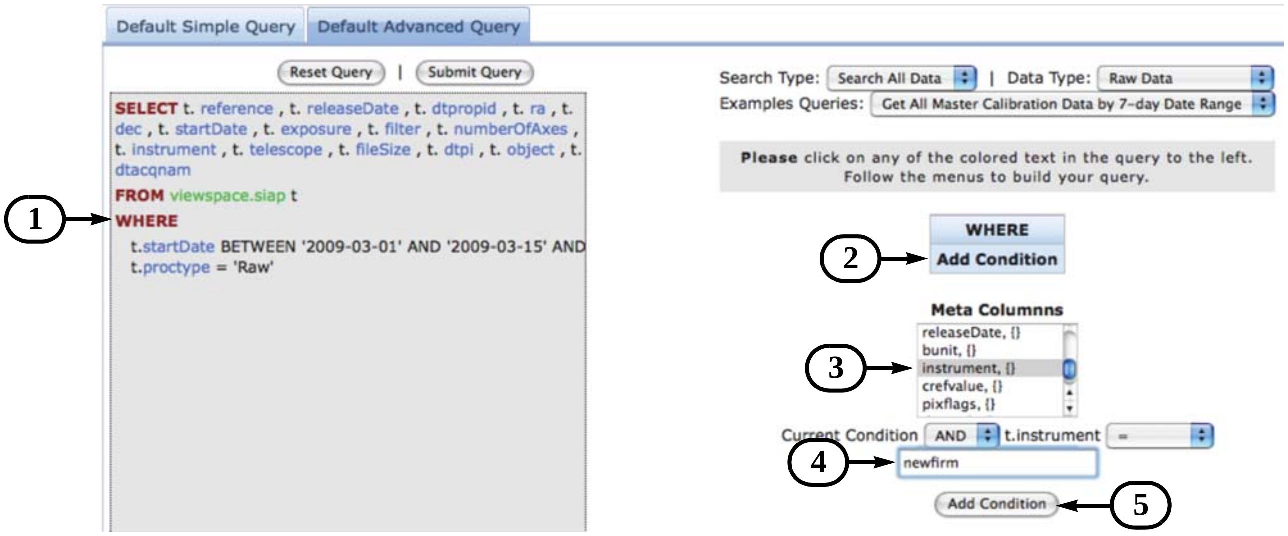


FIGURE 1.9: Portal **Default Advanced Query** panel, showing the process for adding a new condition to the query: the instrument name. See text for details.

The final step is to select for exposure times near 60 s, as illustrated in Figure 1.10. Adding this condition is exactly like adding the other conditions, except for choosing exposure in the selection box (3) and selecting a modest range of exposure times (4) **BETWEEN** 50 and 70 s (since an exact match against a floating-point number is unwise). After pressing the **Add Condition** button (5) and then the **Submit Query** button, which will match hundreds of data files, select files of interest in the Data Grid and load them into the cart for staging, as in prior examples.

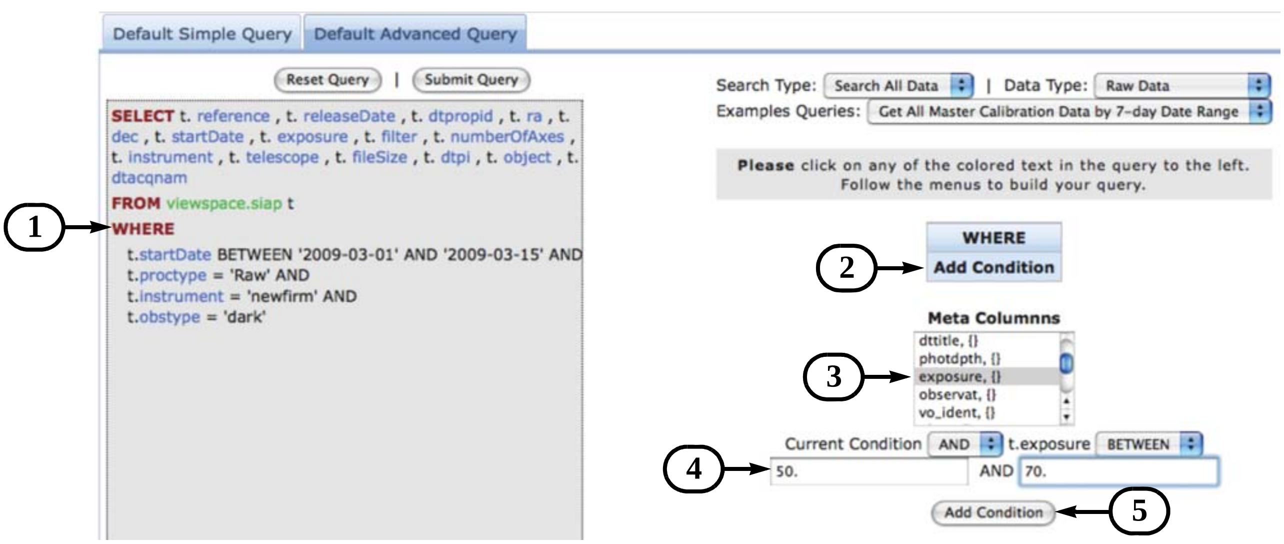


FIGURE 1.10: Portal **Default Advanced Query** panel, showing the process for adding an exposure time as a query condition. See text for details.

Searching for pipeline-produced *Master Calibration* reference files is very similar to the above example, except that the initial default query (Get All Master Calibration Data by 7-day Date Range) should instead search on the Data Type Master Calibrations.

1.4 DATA USE POLICY

There are a number of conditions that apply to the use of data from the NOAO Science Archive; see http://archive.noao.edu/nsa/condOfUse.html for details. In general, AURA (or one of the partner observatories), possibly in addition to the PI of the program, assert ownership of the intellectual property that includes the data and derived products (including catalogs, but not including scientific papers), subject to a royalty-free license to the U.S. federal government. AURA policy provides for free use of these data for scientific research, following a limited period during which the investigator of a program has exclusive (proprietary) rights to those science data. Use of the data and derived products also requires that users provide appropriate attribution (see below). Nonproprietary data include the following:

- all data obtained with any program to characterize the instrument health and performance, including CCD bias and dark frames, flat-fields (including dome- and twilight-flats), comparison arc-lamp exposures, and focus exposures;
- all contemporaneous weather data, all-sky camera images, and seeing monitor data;
- all calibration reference frames that are produced by NOAO calibration pipelines, whether generated from on-sky science frames or not (except that metadata that identify the host science program will be redacted);
- all metadata from all exposures as soon as they are archived, including those from science frames, unless specific authorization has been granted by the NOAO Director to withhold these metadata for the duration of the proprietary period; and
- all science data for which the proprietary period has expired.

The applicable proprietary period for science data depends upon the policy of the host observatory, and may be different under some circumstances. The default period for observing programs obtained through the NOAO time allocation process is 18 months from the calendar date of the exposure, except for programs granted “survey” status for which the proprietary period is usually less. Well-justified requests for extensions may be granted by the NOAO Director, although the request should generally be made when the observing proposal is first submitted for consideration.

1.4.1 Proper Citation for NOAO Data

Authors of papers that make use of data from the NOAO Science Archive should include one of the following credit statements in their publications:

> "National Optical Astronomy Observatory / Association of Universities for Research in Astronomy / National Science Foundation"

The above citation applies to data collected and processed by NOAO. If you use high-level science products that were generated by one of the Survey Project PI teams, the following credit statement should appear in your publication:

> "This research draws upon data provided by [*Survey PI*] as distributed by the NOAO Science Archive. NOAO is operated by the Association of Universities for Research in Astronomy (AURA), Inc. under a cooperative agreement with the National Science Foundation."

The identity of the Survey PI is given on the archive holdings[14] Web page.

1.4.2 Referencing Datasets in Journals

The professional journals of the American Astronomical Society have developed a means to reference specific datasets within the manuscripts of submitted papers. The mechanism has been described by Accomazzi, Eichhorn & Rots (2007). It is helpful to the various archive centers to provide this information so that archive usage statistics can be compiled, and so that links to valuable science results from the original datasets can be offered to future archive users. There is now an AASTeX macro that lets authors include Data Set Identifiers in their papers that will provide a link from the electronic edition of the manuscript to the NOAO Science Archive. One of the AASTeX macros is:

\dataset[***DatasetID***]{***text***}

Here, the *DatasetID* is the identifier for your NOAO dataset; this has not been implemented yet in NOAO data headers. Once these identifiers, and the supporting VO services, have been defined for data in the NOAO Science Archive, you will find this information in the value of the `DS_IDENT` header keyword.

1.5 References & Further Information

Contributing Authors

People who have contributed to the description of the NOAO Science Archive, the data formats, and the NOAO VO Portal, include Dick Shaw, Chris Miller, Mark Dickinson, Howard Lanning, and Rob Seaman.

14. http://archive.noao.edu/nsa/holdings.html

References

Accomazzi, A., Eichhorn, G., & Rots, A. 2007, in ASP Conf. Ser. 376, ADASS XVI, ed. R. A. Shaw, F. Hill, & D. Bell (San Francisco: ASP), 467

FITS 2008, *Definition of the Flexible Image Transport System*, IAU FITS Working Group (Version 3.0: Greenbelt, MD: FITS Support Office, NASA/GSFC); available at http://fits.gsfc.nasa.gov/fits_standard.html

Miller, C. J., Gasson, D., & Fuentes, E. 2007, in ASP Conf. Ser. 376, ADASS XVI, ed. R. A. Shaw, F. Hill, & D. Bell (San Francisco: ASP), 625

Pence, W. D., Seaman, R., & White, R. L. 2009, PASP, 121, 414

Shaw, R. A. 2007, *Essential Imaging Metadata*, DPP System Technical Report STR2007–01 (Tucson: NOAO)

Smith, R. C., Dickinson, M., Lowry, S., Miller, C. J.,Trueblood, M., & Valdes, F. 2007, in ASP Conf. Ser. 376, ADASS XVI, ed. R. A. Shaw, F. Hill, & D. Bell (San Francisco: ASP), 615

Zárate, N., Seaman, R., & Tody, D. 2007, in ASP Conf. Ser. 376, ADASS XVI, ed. R. A. Shaw, F. Hill, & D. Bell (San Francisco: ASP), 351

For Further Reading

Additional documentation on the NOAO VO Portal, including examples and tutorials, is available at http://nvo.noao.edu/noaonvo/help.shtml.