



Data Management Operations (DMO)

Sean McManus

DMO Program Head



Data Management Operations

About

Who we are:

- DMO current workforce of 7.1 FTE consists of mostly technical staff; some shared with Data Lab

What we do:

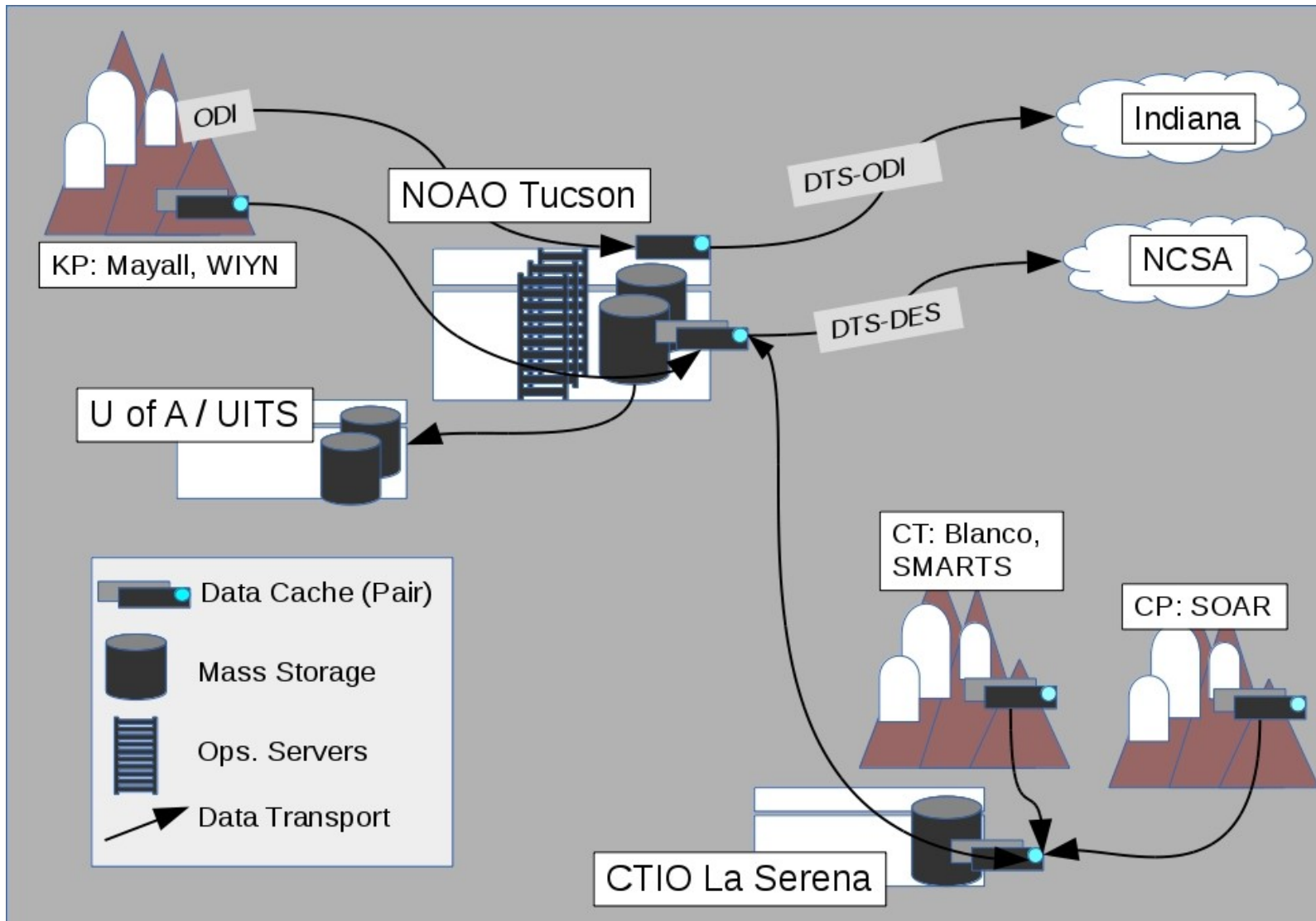
- Build tools and infrastructure to:
- Capture and transport observations off the telescope
- Perform data/metadata validation and reduction
- Serve data products via the Science Data Archive (SDA)

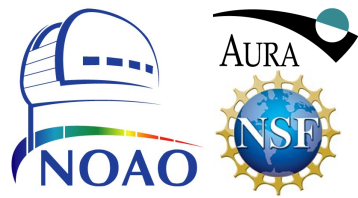
- The SDA is available at <http://archive.noao.edu>



Data Management Operations

Telescope to Archive Data Flow





Data Management Operations

Current Activities (FY2017)

- Telescope Automatic Data Archiver (TADA) implementation at Tucson/CTIO/KPNO
- DECam CP & MzLS pipeline processing and science verification
- IT Infrastructure improvements to support continuous Science Data Archive growth and Data Lab ramp up
- Planning for modernization of user and programmatic interfaces to the Science Data Archive

MARS Administration WELCOME, MARS · VIEW SITE / DOCUMENTATION / CHANGE PASSWORD / LOG OUT

Home > Audit > Audit records ADD AUDIT RECORD +

Select audit record to change

Search 534798 results (570561 total)

FILTER

2014 2015 2016 2017

Action: -----

<input type="checkbox"/>	STAGED	HIDE	FSTOP	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		archive
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		archive
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		archive
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		archive
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		archive
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		archive
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		archive
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		archive

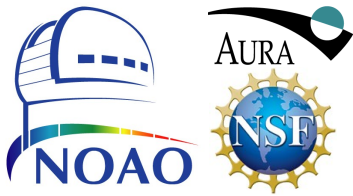
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	archive	May 3, 2017, 3:57 p.m.	<input checked="" type="checkbox"/>	04/27/17	ct4m	decam	/net/decdata1/deccp/NHPPS_DATA /DCP/Final/Submitted /c4d_170428_063607_oki_g_v1.fits.gz
--------------------------	-------------------------------------	-------------------------------------	---------	------------------------	-------------------------------------	----------	------	-------	---

DUPFITS

ETLKEY

HDREXIST

- TADA is a Python / Astropy-based system that extracts, transforms and loads variable FITS file metadata into the Science Data Archive
- Integrated with a suite of web services (MARS) for telescope schedule management, and observing night audit and remediation.
- TADA has been implemented for NOAO, SOAR, SMARTS, & WIYN instruments at KPNO and CTIO
- Automated services



Data Management Operations

TADA Status

MARS Administration WELCOME, MARS · [VIEW SITE](#) / [DOCUMENTATION](#) / [CHANGE PASSWORD](#) / [LOG OUT](#)

Home > Audit > Audit records ADD AUDIT RECORD +

Select audit record to change

Search Search 534798 results (570561 total) FILTER

2014 2015 2016 2017

Action: -----

STAGED HIDE FSTOP

archive

• TADA keeps an audit “paper-trail” that starts in the dome, passes through disk caches, and ends in successful archive ingest

<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	archive	May 3, 2017, 3:59 p.m.	<input checked="" type="checkbox"/>	04/30/17	ct4m	decam	/net/decdata1/deccp/NHPPS_DATA /DCP/Final/Submitted /c4d_170501_055757_opi_g_v1.fits.fz	Yes No
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	archive	May 3, 2017, 3:59 p.m.	<input checked="" type="checkbox"/>	04/27/17	ct4m	decam	/net/decdata1/deccp/NHPPS_DATA /DCP/Final/Submitted /c4d_170428_075307_ooi_g_v1.fits.fz	By fstop All archive mountain:cache valley:cache valley:direct valley:queue watch
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	archive	May 3, 2017, 3:59 p.m.	<input checked="" type="checkbox"/>	04/30/17	ct4m	decam	/net/decdata1/deccp/NHPPS_DATA /DCP/Final/Submitted /c4d_170501_051003_oki_g_v1.fits.fz	By errcode All DUPFITS ETLKEY HDREXIST
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	archive	May 3, 2017, 3:58 p.m.	<input checked="" type="checkbox"/>	04/27/17	ct4m	decam	/net/decdata1/deccp/NHPPS_DATA /DCP/Final/Submitted /c4d_170428_052408_ooi_g_v1.fits.fz	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	archive	May 3, 2017, 3:58 p.m.	<input checked="" type="checkbox"/>	04/27/17	ct4m	decam	/net/decdata1/deccp/NHPPS_DATA /DCP/Final/Submitted /c4d_170428_071300_opi_g_v1.fits.fz	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	archive	May 3, 2017, 3:57 p.m.	<input checked="" type="checkbox"/>	04/27/17	ct4m	decam	/net/decdata1/deccp/NHPPS_DATA /DCP/Final/Submitted /c4d_170428_063607_oki_g_v1.fits.fz	

Data Management Operations

Pipeline Status

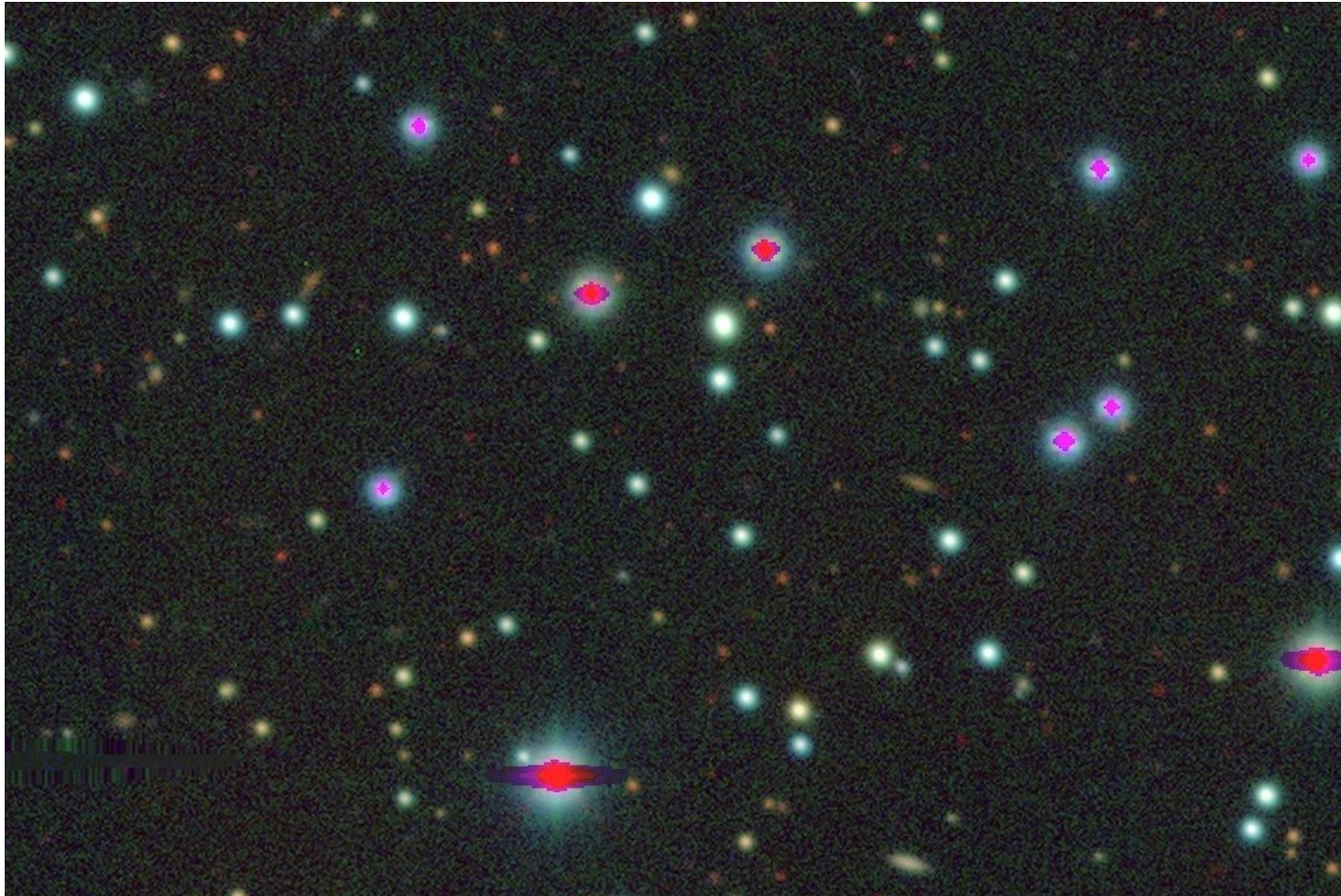
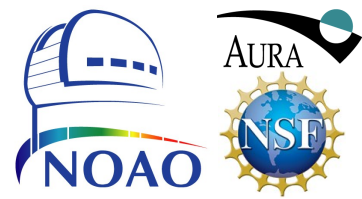


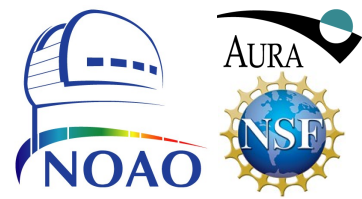
Figure: A sample patch of sky from the MzLS/BASS surveys (DR4)
[Courtesy of DESI Targeting Survey and Dustin Lang]



Data Management Operations

Pipeline Status

- DECam Community Pipeline (CP) producing timely calibrations in support of PI-based and major survey programs (e.g. DECaLS, BLISS, DECaPS)
- Mosaic3 pipeline primarily in support of MzLS - first and second years data (to date) is processed and publicly available
- Steward Bok/90Prime pipeline for BASS data in support of DESI Targeting Survey
- NEWFIRM pipeline retired and all data processed.



Data Management Operations

DECam Community Pipeline

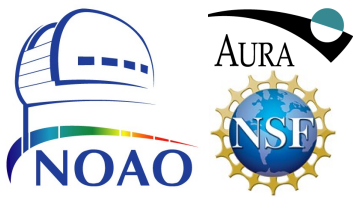
- DESDM orchestration framework eliminated and processing efficiency improved
- Astrometric calibration and photometric characterization changed to Gaia and PS1/Gaia respectively
- Brighter-fatter correction algorithm added but not the default - awaiting study panel evaluation
- DES is now treated as a community program for providing CP versions of calibrated exposures: Y1 and Y2 available via Science Data Archive



Data Management Operations

IT Infrastructure Developments

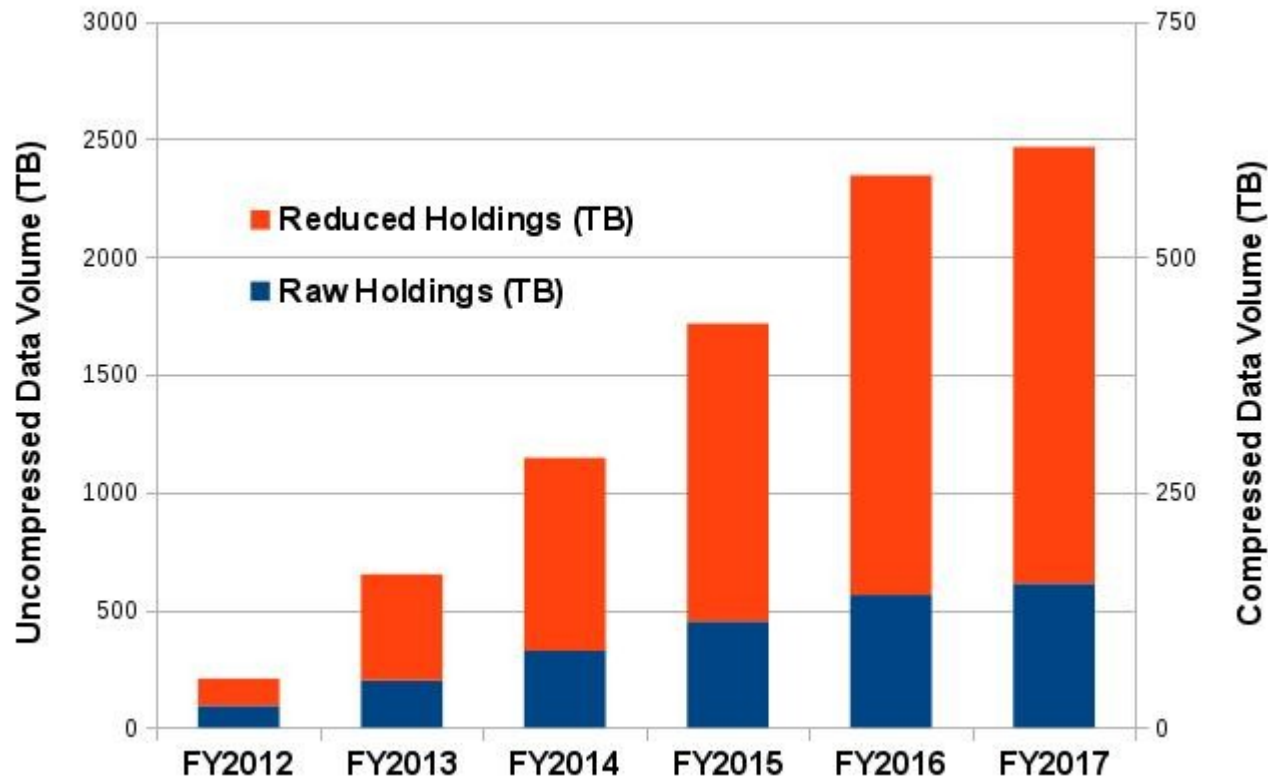
- 10 General Purpose Data Lab machines
 - 20 cores, 40 threads, 256G RAM each
 - Spinning and solid-state disk media
- Data Lab redundant mass storage system
 - IBM General Parallel File System
 - 900 TB installed, can augment as needed
- Science Data Archive mirror recently installed at University of Arizona Information Technology Services (UITS)
- ANTARES system also installed at UITS
- Mountain cache systems upgraded at CTIO/KPNO
- Mass storage upgrades pending to support hosting SDSS data

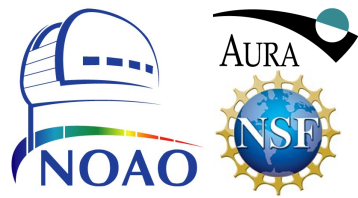


Data Management Operations

Science Data Archive Holdings & Usage

- Uncompressed holdings approach 2.5 Petabytes.
- Compression reduces disk utilization by ~4:1





Data Management Operations

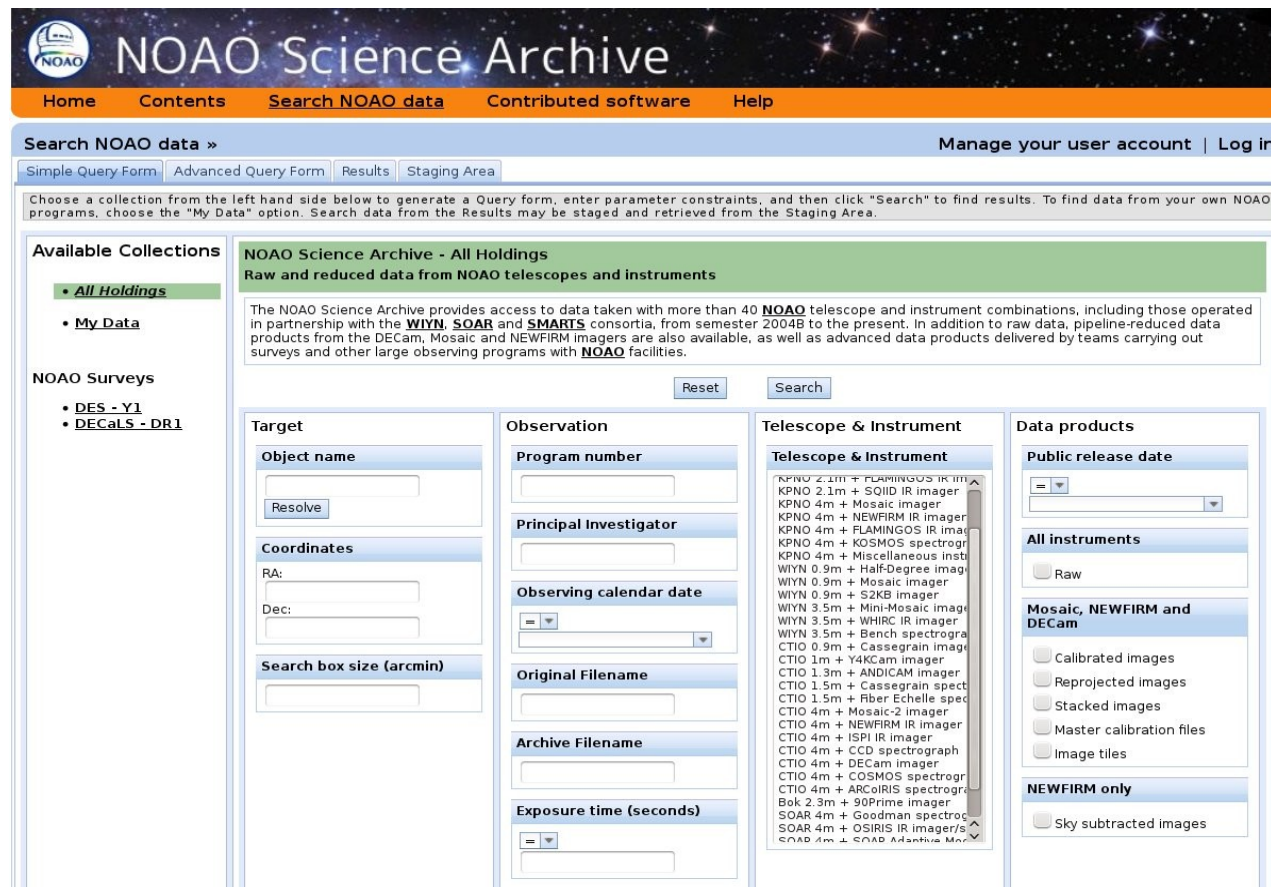
Science Archive Usage

- Archive utilization also on the rise.

	Avg. Monthly Bandwidth (GB)	Avg. Monthly Downloads (file count)	Avg. Monthly Web Pages (page views)	Avg. Monthly Unique Users
FY2012	511	11422	63775	2402
FY2013	2407	30038	183001	1495
FY2014	3641	40938	55875	1455
FY2015	6185	49523	217225	1641
FY2016	15838	86521	72815	1734

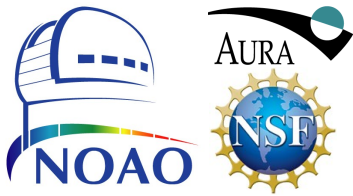
Data Management Operations Science Data Archive Software Status

- Software mostly frozen through FY16, but new hire in place to reboot new development.



The screenshot shows the NOAO Science Archive search interface. At the top, there is a navigation bar with links for Home, Contents, Search NOAO data, Contributed software, and Help. Below this is a search bar with options for Simple Query Form, Advanced Query Form, Results, and Staging Area. A message instructs users to choose a collection from the left-hand side to generate a Query form, enter parameter constraints, and click "Search" to find results. The main content area is divided into several sections:

- Available Collections:**
 - **All Holdings** (selected)
 - My Data
- NOAO Surveys:**
 - DES - Y1
 - DECaLS - DR1
- NOAO Science Archive - All Holdings:**
 - Raw and reduced data from NOAO telescopes and instruments
 - The NOAO Science Archive provides access to data taken with more than 40 NOAO telescope and instrument combinations, including those operated in partnership with the WIYN, SOAR and SMARTS consortia, from semester 2004B to the present. In addition to raw data, pipeline-reduced data products from the DECam, Mosaic and NEWFIRM imagers are also available, as well as advanced data products delivered by teams carrying out surveys and other large observing programs with NOAO facilities.
- Search Form:**
 - Target:** Object name (with a Resolve button), Coordinates (RA and Dec), Search box size (arcmin).
 - Observation:** Program number, Principal Investigator, Observing calendar date (with a dropdown menu), Original Filename, Archive Filename, Exposure time (seconds) (with a dropdown menu).
 - Telescope & Instrument:** A list of telescope and instrument combinations, including KPNO 2.1m + FLAMINGOS IR imager, KPNO 2.1m + SQUID IR imager, KPNO 4m + Mosaic imager, KPNO 4m + NEWFIRM IR imager, KPNO 4m + FLAMINGOS IR imager, KPNO 4m + KOSMOS spectrograph, KPNO 4m + Miscellaneous instruments, WIYN 0.9m + Half-Degree imager, WIYN 0.9m + Mosaic imager, WIYN 0.9m + S2KB imager, WIYN 3.5m + Mini-Mosaic imager, WIYN 3.5m + WHIRC IR imager, WIYN 3.5m + Bench spectrograph, CTIO 0.9m + Cassegrain imager, CTIO 1m + Y4KCam imager, CTIO 1.3m + ANDICAM imager, CTIO 1.5m + Cassegrain spectrograph, CTIO 1.5m + Fiber Echelle spectrograph, CTIO 4m + Mosaic-2 imager, CTIO 4m + NEWFIRM IR imager, CTIO 4m + ISPI IR imager, CTIO 4m + CCD spectrograph, CTIO 4m + DECam imager, CTIO 4m + COSMOS spectrograph, CTIO 4m + ARCIRIS spectrograph, Bok 2.3m + 90Prime imager, SOAR 4m + Goodman spectrograph, SOAR 4m + OSIRIS IR imager/spectrograph, SOAR 4m + SOAR Adaptive Mode imager.
 - Data products:** Public release date (with a dropdown menu), All instruments (Raw), Mosaic, NEWFIRM and DECam (Calibrated images, Reprojected images, Stacked images, Master calibration files, Image tiles), NEWFIRM only (Sky subtracted images).



Data Management Operations

Science Data Archive Software Status

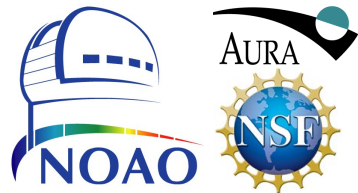
- Historically a PI data access tool but new developments will focus on serving of large public datasets
- Capabilities being added to ingest high level science products from past and present survey programs
- Bulk data transfers possible via FTP site:
<ftp://archive.noao.edu/public/hlsp/>
- Current software implementation (which pre-dates Data Lab) is being re-evaluated in parallel with ongoing operations
- Feature development must be balanced with sustainability.
- New hire will coordinate with Data Lab to determine roadmap for development going forward



Data Management Operations

Looking Ahead

- SDA will serve DESDM-reduced Y2/Y3 images as they go public
- DECaLS DR4 available soon
- SDSS Public data mirror
- Migration of legacy NOAO survey holdings in progress
- TADA development transition to maintenance mode
- Improvements to Science Data Archive back-end software and user interface



Data Management Operations

Supplementary Slides



Data Management Operations

Personnel

Position	FY17 FTE	FY17 Staff
Program Management	0.3	McManus (0.1), Lauer, (0.1), Dickinson (0.1)
Infrastructure Dev/Ops	4.3	McManus (0.9), Peterson, (1.0), Pothier (1.0), Norris (0.5), Fitzpatrick (0.1)
Pipeline Dev/Ops	1.8	Hererra (1.0), Valdes (0.8)
Science Support	0.7	Lauer (0.5), Valdes (0.2),
TOTAL	7.1	

