

# **Summary of Data Management Principles**

## **Dark Energy Survey**

### **V2.3, 9/19/16**

This Summary of Data Management Principles (DMP) has been prepared at the request of the DOE Office of High Energy Physics, in support of the DOE Office of Science Statement on Digital Data Management and for reference by those submitting proposals in response to DOE Office of Science Funding Opportunity Announcements.

#### **Experiment description:**

The Dark Energy Survey (DES) comprises two imaging surveys of unprecedented scope and science reach: a wide-area survey and a narrower time-domain survey. The wide-area survey covers 5000 sq. deg. of the South Galactic Cap in 5 optical-near infrared filters (grizY) and will provide images of 300 million galaxies. The time-domain survey covers 30 sq. deg. in 4 filters (griz) and will discover and provide light curves for 3500 type Ia supernovae. These interleaved surveys are being carried out in 525 nights spread over 5 observing seasons, using the new Dark Energy Camera (DECam). DECam is a 3 sq. deg. field-of-view, 570-megapixel instrument mounted on the Blanco 4-meter telescope at the National Optical Astronomy Observatory's (NOAO) Cerro Tololo Inter-American Observatory (CTIO) in northern Chile. DECam was constructed by the DES collaboration to carry out the survey and serves as a facility instrument operated for the astronomy community by NOAO. DES survey operations started on Aug. 31, 2013; each 105-night observing season runs from August to the following February.

The DES collaboration designed and will use the survey to probe the origin of cosmic acceleration and the nature of dark energy through four complementary methods: galaxy clusters, gravitational lensing, large-scale galaxy clustering (including baryon acoustic oscillations), and supernovae. DES will make a significant advance in the precision and accuracy of dark energy parameter constraints. In so doing, it will also help lay the groundwork for the subsequent experiments LSST and DESI.

#### **DOE's roles in the experiment:**

Construction of the Dark Energy Camera (DECam) instrument and of a number of ancillary hardware systems was supported by the DOE Office of High Energy Physics and led by Fermilab. Maintenance of the Dark Energy Camera is shared between Fermilab/DES, supported by DOE DES operations funds, and CTIO/NOAO, through its base funding from the NSF Division of Astronomy (AST). DOE supports the operation of the experiment, including DES observing, and infrastructure for science operations, including science analysis computing, survey strategy, and calibrations, and contributes support to a number of other tasks, including supernova survey operations, weak lensing shear pipeline testing, and the DES Project Scientist. DOE also provides some support for DES Data Management (DESDM) operations, primarily at Fermilab and NERSC.

## **Partnerships:**

DOE and NSF are partners in supporting DES operations, with DOE providing support as described above and NSF providing the primary support for DESDM operations through a grant to the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign. The agencies provide oversight through a DOE-NSF Joint Oversight Group that meets approximately monthly with DES Management. In addition, the experiment is governed by the DES Memorandum of Understanding, an agreement between Fermilab, NCSA, and NOAO, the three institutions with overall responsibility for construction and operations. The DES Council of Directors of the three MOU institutions meets periodically. There is currently no MOU or statement of principles between NSF and DOE regarding data management.

## **Organization – Agency/Lab level**

As noted above, Fermilab is the lead lab for DOE's role in DES, providing leadership both for the DES collaboration's role in DECam maintenance and for overall operation of the experiment. NCSA is the lead institution for DESDM operations.

## **Organization – Experiment level**

The DES organization chart as of Sept. 1, 2016 is shown in Fig. 1. The DES Data Management PI leads the NSF-supported Data Management effort at NCSA, with responsibility for developing, operating, and maintaining the DES pipelines and serving internal and external data releases.

## **Collaboration:**

The collaboration comprises roughly 400 scientists (including students and postdocs) at 25 DES institutions and consortia from the U.S., the United Kingdom, Spain, Brazil, Germany, Switzerland, and Australia. The organization of the collaboration is described in the previous section. Collaboration membership is governed by the DES Membership Policy. In addition to those at DES institutions, there are DES collaborators at non-DES institutions who have typically moved from a DES institution and arranged to maintain their participation in the project. In addition to those in the collaboration, there are a number of External Collaborators who have been granted limited data access for particular projects in exchange for providing resources or expertise deemed beneficial to the collaboration.

## **Data policy management:**

The DES Public Data Release Plan (which currently exists in advanced draft form) has been formulated and will be maintained by the DES Project Office (see Fig. 1 above). Some elements of the plan are contained in the DES Memorandum of Understanding and in the 2011 and 2015 proposals submitted to NSF AST to support DESDM operations. The Plan is described in the sections below. Responsibility for making the DES data public rests with the Project Office. Implementation of the data release plan has been assigned to the DESDM team, led by NCSA, and to the NOAO Science Archive (NSA).

# DES Organization Chart

9/1/2016

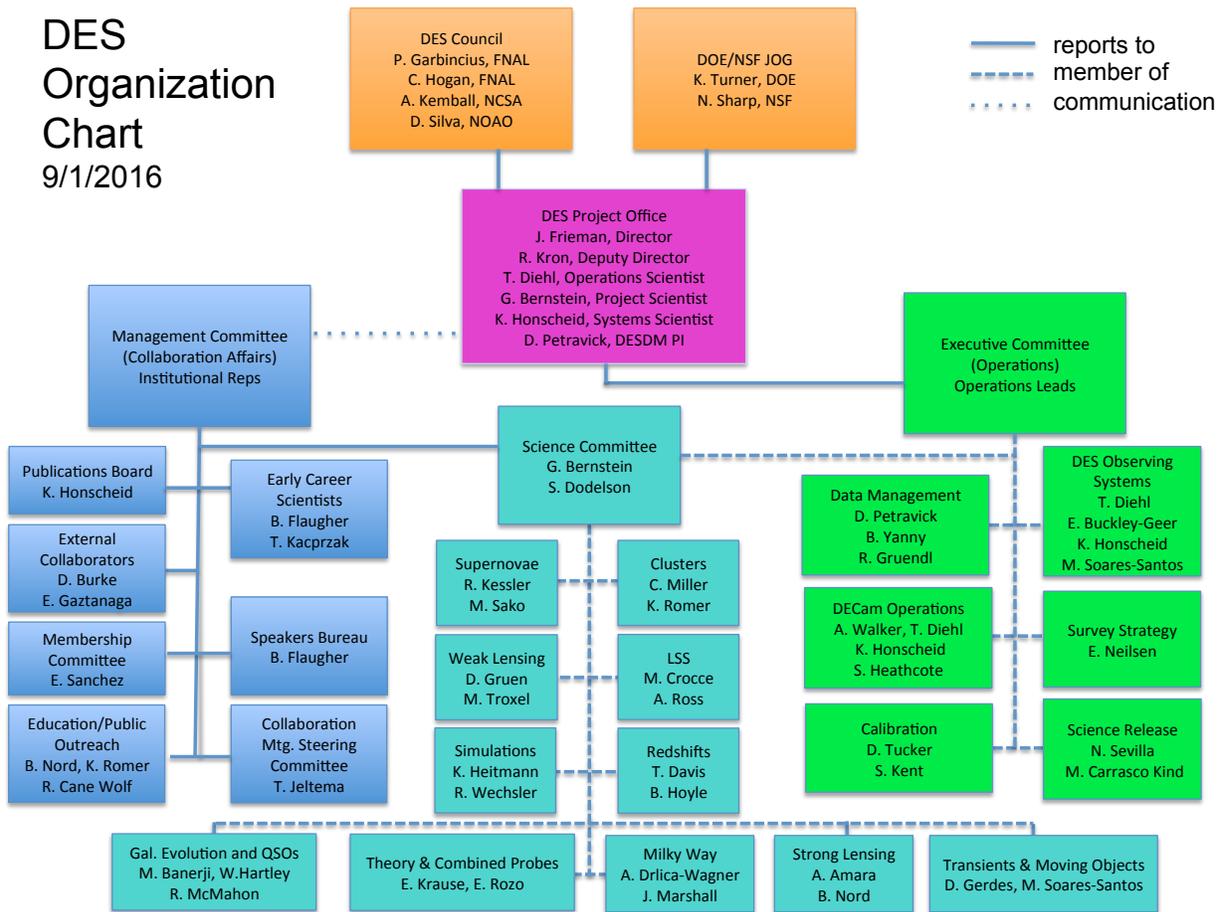


Fig. 1: DES Organization Chart.

## Data Description & Processing:

The DES Data Management System deals with and produces a number of types of data products. During the DES observing season, raw DECam images are sent via the NOAO Data Transport System to NCSA for nightly processing by the DESDM system. From these raw exposures, the DESDM system produces reduced (instrumental artifact-removed) single-epoch images and object catalogs derived from them, which are assessed to provide rapid feedback on data quality. The DES Supernova exposures are additionally processed through an image-subtraction pipeline (using previously generated template images) in order to identify transient events and in particular candidate supernovae for spectroscopic follow-up by other telescopes.

After each observing season, DESDM carries out an annual processing of the data taken to that point, which includes uniform reprocessing of the raw images, application of photometric calibration, generation of co-added images from multiple exposures of the same field, and derivation of object catalogs and masks from the co-added images. These science-ready object

catalogs include a number of measured quantities for each identified object, including different estimates of object flux, classification as a star or galaxy, and photometric redshift estimates (based on matching objects across filters). Additional derived object quantities, such as weak lensing shear estimates, are also ingested into the DESDM catalogs.

Single-epoch and coadded images are retained on disk in FITS format, which is the standard file format use in astronomy. The FITS format supports image meta-data, which DES populates in a manner that makes the image files useful. Catalogs, masks, and other data are maintained in an Oracle relational database, with access available to all DES collaborators. Additional access and QA tools are provided by the DES Science Portal.

## **Data Products and Releases:**

The publicly released DES data products will consist of: (I) raw DECam images taken during DES observing periods, (II) reduced (artifact-removed), calibrated single-epoch images processed by DESDM for both the wide-area survey and the supernova survey, (III) two releases of calibrated, co-added images and catalogs and associated data products output by DESDM, and (IV) time-sensitive alerts with information about transient objects.

### **I. Raw DES Data**

DES has agreed to a 12-month proprietary period for raw DES images. NOAO is managing access to DES raw data, making them automatically available to the general community on a rolling basis via the NSA 12 months after they are taken.

All raw DECam calibration data (flats, biases, etc.) taken by DES or by community observers have no proprietary period and will be made immediately available to the general community via the NSA, per NOAO policy. Moreover, any calibration files supplied by DES to NOAO are made immediately available to the general community via the NSA.

All raw DECam data taken during DECam commissioning (Sept.-Oct. 2012) and during Science Verification (which included both DES and community observations) have no proprietary period and were made immediately available to the general community via the NSA. This includes Science Verification data processed by NOAO through the Community Pipeline. The DES Science Verification data were taken in the period Nov. 2012-Feb. 2013.

### **II. DES Reduced Single-Epoch Data Products**

The annual processing after each observing season produces globally calibrated, reduced, single-epoch images. Those images that satisfy the DES criteria for inclusion in the annual internal release for the DES Collaboration, including both the wide-area survey and exposures in the supernova fields, will be included in the public release of these reduced images. These reduced images will be made available by NCSA to NOAO for public distribution via the NSA, once their data quality, particularly their photometric accuracy, has been validated by the collaboration. This validation process is expected to be completed 12 months after the images have been released to the collaboration as part of the internal co-add release (see below).

### III. Co-added, Calibrated DES Data Products

The depth and photometric calibration precision of DES data will increase from season to season as more images covering a given sky region are accumulated. DES will make two major releases of co-added, calibrated data products, consistent with the 2011 DESDM Operations Proposal to NSF. The user interface to the data will be comparable in capabilities to other access sites for astronomical survey data. These two public data releases will be served by NCSA. In addition to these 2 releases of DES survey data, the collaboration also publicly released value-added catalogs based on the Science Verification data in January 2016; these are served by NCSA.

#### A. DES Public Data Release 1

The first public data release, DR1, will include products derived from DES data taken up through February 2016, namely DESDM-processed co-add images and catalogs and associated data products (masks, quality information, etc.) for data taken during the first three DES seasons and the DES Science Verification period that satisfy the criteria for inclusion in the internal DESDM release to the Collaboration. In addition, DR1 will include co-add images and catalogs of the DES Supernova fields.

The date for DES DR1 public release will be no earlier than one year after the internal release to the DES collaboration is complete. As of this writing (Sept. 7, 2016), processing for this internal release is on-going, so the public DR1 is currently expected to be no earlier than early 2018. The public release will follow in-depth vetting of the associated data products by the collaboration. If that vetting process uncovers the need for reprocessing of the data to meet scientific standards, then the DR1 release date will be correspondingly later.

#### B. DES Public Data Release 2

The final public data release, DR2, will include DESDM-processed co-add images and catalogs and associated data products for data taken during all DES observing seasons and the DES Science Verification period that satisfy the criteria for inclusion in the internal DESDM releases to the collaboration. Exposures included in DR1 will be reprocessed for DR2. DR2 data are expected to cover the wide-area survey footprint to a depth of approximately 8-10 tilings per filter. DR2 will be released no earlier than one year after the final internal release that is its foundation, in order to give the collaboration time to vet and characterize the data quality in detail.

DR2 will be served and maintained by NCSA for a period of approximately 2 years. Data access will be via http or other standards available at the time. DR2 data products to be released will include fully calibrated coadd image files, masks, weight maps, and catalogs of photometric, astrometric, and other image parameters. The data will be documented to the standards of a peer-reviewed data release document.

### IV. Transient Alerts

The Supernova Working Group will publicly announce supernova candidates detected in the supernova pipeline via periodic astronomer's telegrams (ATELs). In addition, a system called

DESAAlert responds to gamma-ray bursts by generating a VOEvent if the source is contained within the area of sky that DES has observed. Starting with DES Year 3, DES is publicly announcing via a web interface information for bright transient events as soon as they are processed.

### **Plan for Serving Data to the Collaboration and Community:**

Collaboration access to the raw and reduced images and catalogs is provided by NCSA, with additional access tools from the DES Science Portal. Public access to the raw and calibrated single-epoch images is provided by the NOAO Science Archive web/database interface, which is maintained by NOAO. Access to the DR1 and DR2 releases will be via database and Science Portal queries and will be provided by NCSA.

### **Plan for Archiving Data:**

In addition to the raw and processed single-epoch images, the NOAO Science Archive will provide long-term preservation and curation and public access to DES DR1 and DR2 following a to-be-agreed-upon data transfer from NCSA. DESDM will ensure that the files are delivered to NSA with full documentation. The date for the data transfer is to be determined but should be sufficiently before the end of the 2-year NCSA DR2-serving period to ensure a smooth transition with no interruption in public access to these DES data products.

### **Plan for Making Data Used in Publications Available:**

To the extent reasonably possible, we will ensure that the data points shown in simple published graphs (e.g., where two quantities are plotted against each other or histograms) are publicly available in machine readable form, once the corresponding publication has been accepted for publication. For reasons of practicality, this will not include the information in more complex plots such as images and maps. The underlying data products for DES science results will be made available on the timescales described above in the section on Data Products and Releases. In some cases, DES science publications will be based upon catalogs derived from the DES data products; the plan for release of such catalogs (or for release of instructions for generating such catalogs from DES data products where possible), is still under consideration.

### **Responsiveness to SC Statement on Digital Data Management**

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