Dark Energy Survey Data Management Project

Dark Energy Survey Data Management

Data Products

Version 3

Joe Mohr

U Illinois Astronomy, Physics and NCSA

for the DESDM Team

Feedback on this document should be directed to jmohr@uiuc.edu

Revision History Log

Revision	Date	Initiated By	Changes/Comments
1	12/02/07	Mohr	First Release
2	12/15/07	Mohr	Added discussion of data volumes
3	8/17/08	Mohr	Revised to reflect changes in Archive structure

Table of Contents

DESDM Data Products

1 Introduction

This document describes the data products that are made available to the DES Collaboration and the Community by the DESDM project. These data products are available through the DES Archive. As described in the DESDM Technical Design Report, the DES Archive consists of multiple filesystem nodes, a central Oracle database for tracking the data, a data access framework consisting of grid-enabled data transfer tools, and a web based DES Archive Portal that supports user retrieval of the data.

We categorize the data products as (1) image data and metadata, (2) catalog data, (3) survey metadata (4) additional collaboration data and (5) external collections. In the sections below we describe each of these categories.

2 Image Data and Metadata

Image data come in a variety of forms, which we describe below. Following an overview of image data types is a detailed discussion of the characteristics of each data type. We then describe the image metadata that are ingested into the DES Archive database to support the processing and selection of data.

2.1 Image Data Format

All image data are stored as FITS files. These image files (*img*) contain entire exposure when they arrive from DECam, but we split them into individual CCD images to enable efficient data parallel processing. In addition to the image data itself we use FITS files to store associated information such as the pixel variances and the pixel masks. The pixel variances or rms variations are stored as weight maps (*wmp* or *rms*). The *wmp or rms* is stored as a 32 bit float FITS image. The pixel masks are stored as as 16 bit unsigned integers, and we refer to them as bad pixel masks (*bpm*). As described in the Technical Design Report, this bpm uses different bits to encode the root case of a problem with a pixel. The typical DESDM processed image then has three extensions: (1) the image, (2) the weight map and (3) the bad pixel mask. The uncompressed size of a processed DESDM image for a single CCD is therefore 2048x4096x[4Bx2+2B]=80MB.

With CFITSIO for image I/O, DESDM operates equally well with compressed or uncompressed images or catalogs. The compression modes currently supported by CFITSIO are gzip, Rice and PLIO (see http://heasarc.gsfc.nasa.gov/docs/software/fitsio/compression.html for full description), and others may be added in the future. Gzip compression is lossless, but is a lower performance solution than Rice compression. We have run the DESDM system using gzipped inputs and outputs, but the CPU costs are considerable. We have not yet tested Rice compression, because it is not lossless for floating point data. Using Rice compression will involve first representing the floating point images as integers in the standard CFTISIO manner (parameters BZERO and BSCALE). We have scheduled tests of Rice compression, but at this time we are running without any compression. The PLIO compression is a special mode developed to support IRAF image masks, and it is not clear that it offers any advantage over Rice compression.

One advantage of archiving gzipped data is that the compression algorithm does its own checksumming when gunzipping. Our planned solution to guard against FITS image corruption is to employ CFITSIO checksumming in our image I/O subroutines; corrupted FITS files will be caught immediately during the image read step, and the DESDM system will extract the backup copy from local tape resources. At NCSA every file archived to tape appears on two separate tapes, and one of these tapes is stored off site. In the case of DESDM there will also be a copy of each file at Fermilab, the secondary DES Archive site.

2.2 Image Classes

There are several types of images produced and archived by DESDM. We divide these into five broad classes of files, as detailed in Table 2-1. Simply stated, the cal data are those images describing characteristics of the instrument needed for the nightly processing, raw data are those images ready for processing, red images are those produced during nightly processing runs, coadd images are those produced in the process of coadding single images of a band into deeper images of the sky and cutout images are images of particular portions of the sky.

Table 2-1: File Classes

Image Class	Description
cal	Calibration data describing the instrument that is needed for processing
src	Source data from DECam that still contains instrumental signatures
red	Processed data that has the instrumental signatures removed
coadd	Image data derived from the coaddition of processed images
cutout	Subimages created for the viewing of particular objects or regions of the sky

Within each of these classes there are several types of images. We list these file types for images in Table 2-2. These image types allow us to keep track of the scientific uses of each of the image data products.

Table 2-2: File Types defined for images

Image Type	Image Class	Description	Data	# Ext	Size
src	src	DECam Exposure	16b int	63(62 img)	2104x4096
raw_obj	red	Crosstalk corrected science image	32b flt	1(img)	2104x4096
raw_bias	-	Crosstalk corrected bias image	32b flt	1(img)	2104x4096
raw_dflat	-	Crosstalk corrected dome flat image	32b flt	1(img)	2104x4096
raw_sflat	-	Crosstalk corrected sky flat image	32b flt	1(img)	2104x4096
raw_dark	-	Crosstalk corrected dark exposure	32b flt	1(img)	2104x4096
red	-	detrended image	32b flt	3(img, wmp, bpm)	2048x4096
remap	-	reduced image remapped to coadd tile	32b flt	2(img, var)	Variable
diff	-	reduced image with template subtracted	32b flt	2(img, var)	2048x4096
biascor	-	bias correction image	32b flt	2(img, var)	2048x4096
flatcor	-	flat field correction image	32b flt	2(img, var)	2048x4096
illumcor	-	illumination correction from supersky	32b flt	2(img, var)	2048x4096
supersky	-	supersky produced from red images	32b flt	3(img, wmp, bpm)	2048x4096

fringecor	-	fringe correction from supersky	32b flt	2 (img, var)	2048x4096
coadd	coadd	combined remap images	32b flt	2(img, var)	14300^2
pupilcor	cal	pupil ghost correction	32b flt	2(img, var)	2048x4096
bpm	-	bad pixel image	16b int	1(img)	2048x4096
all cor images	-	corrections from night stored for comparison	32b flt	2(img, var)	2048x4096
reduced	cutout	Subimage for multiple epochs derived from reduced images	32b flt	2N(img,var)	Variable
remap	-	Subimage for multiple epochs derived from remap images	32b flt	2N(img,var)	Variable
coadd	-	Subimage derived from the coadd images	32b flt	2(imag,var)	Variable
diff	-	Subimage for multiple epochs derived from difference images	32b flt	2N(img,var)	Variable

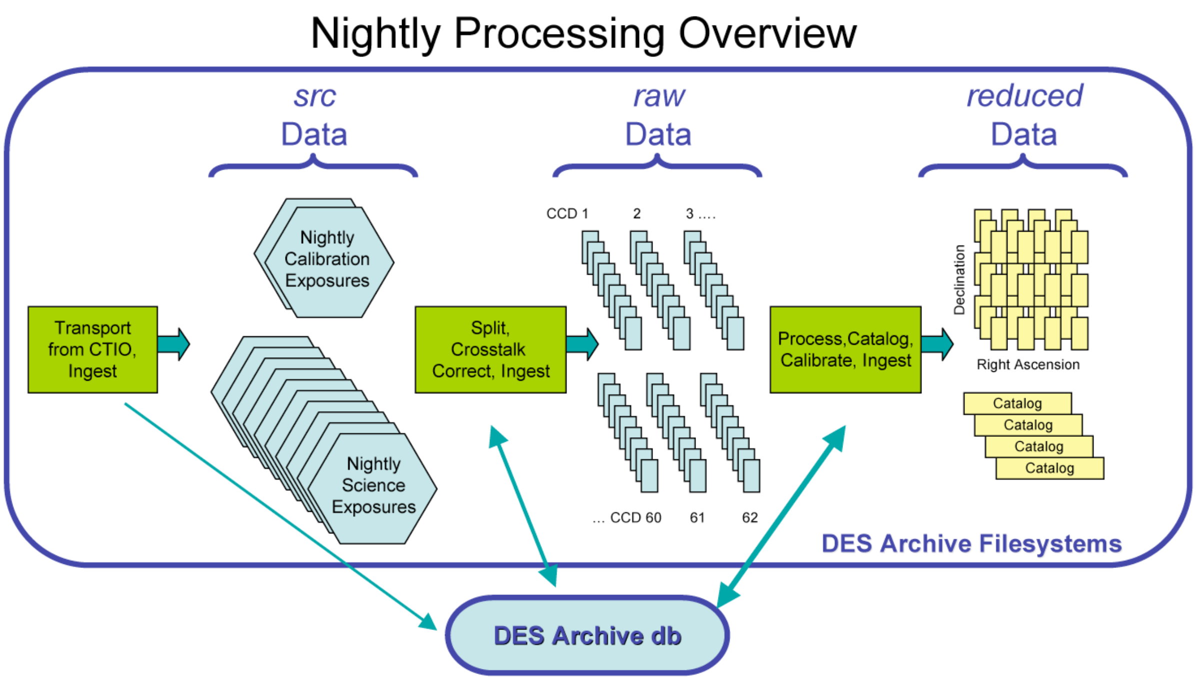

Figure 2-1: During the nightly processing, DECam exposures or src data are (1) transferred from CTIO for ingestion into the DES Archive and (2) split to CCD level raw data with crosstalk correction and ingestion. These raw data are then processed, calibrated and cataloged for ingestion into the DES Archive. Image metadata and catalog data are ingested in the DES Archive database to support further processing and scientific analyses.

2.2.1 FileClass *src*

There is only one type of fileclass src data. This is the DECam source data straight (*src*) from the camera and the collection of log data (*log*) stored for that night. We describe each of these in the subsections below.

2.2.1.1 DECam *src* Images

DECam exposures or source data are stored in single MEF files with 63 extensions; these exposures are approximately 1GB in length (uncompressed). The first extension or zero-header contains header information that is relevant to the exposure as a whole. The second through sixty-third extensions are image extensions with associated headers. These image extensions contain the readout from an entire DECam CCD (not an amplifier). The image data type is 16 bit unsigned integer, and the images are 2104x4096, allowing for overscan regions of 28 pixel width on both the left and right of each CCD. Two overscan regions are required for each CCD, because the CCD is read using two amplifiers. We designed the headers to make the DECam data standalone to the extent that is possible. We examined header structures for several different cameras, including the NOAO Mosaic2, the CFHT MegaCam and the SDSS camera. Ultimately we have chosen a Mosaic2 like header with the added information required to support reduction of image extensions that contain multiple amplifiers. The image extension headers contain information specific to the image contained in a particular extension.

Table 2-3 contains the zero-header keywords and description, while Table 2-4 contains the image extension headers. Along with the keyword, we have listed the data type, comment and an example of each.

Table 2-3: DECam Exposure 0th Extension Header

Keyword	Data Type	Example	Comment
SIMPLE	Logical	T	File does conform to FITS standard
BITPIX	Integer	8	Number of bits per data pixel
NAXIS	Integer	0	Number of data axes
EXTEND	Logical	T	FITS dataset may contain extensions
NEXTEND	Integer	62	Number of extensions
FILENAME	String	decam-2006-10-06-40141298.fits	Filename
EXPNUM	Integer	452893	DECam exposure number
OBJECT	String	DES0550-2116A	Object name
OBSTYPE	String	object	Observation type
		Possible values: zero, dark, focus, object, dome flat, sky flat	
OBSERVER	String	John Peoples	Observer Name
PROPID	String	2006B-0569	Proposal ID
DETECTOR	String	decam	Detector
DETSIZE	String	[1:24576,1:28672]	Detector size
TELESCOP	String	ct4m	Telescope Name
OBSERVAT	String	CTIO	Observatory Name
LATITUDE	String	-30:09:54	Observatory Latitude (deg)
LONGITUD	String	70:49:00 W	Observatory Longitude (deg)
ALTITUDE	Float	2215.0	Observatory Altitude (m)
TELRADEC	String	FK5	Telescope coordinate system
TELEQUIN	Float	2000.0	Equinox of telescope coordinates
TELRA	String	00:21:16.92	Telescope RA (hr)
TELDEC	String	-48:39:43.6	Telescope Dec (deg)
HA	String	02:09:18.7	Telescope hour angle (hr)
ZD	Float	30.7	Telescope zenith distance (deg)
AIRMASS	Float	1.162	Airmass
FILTER	String	g SDSS c6017	Filter name
TELFOCUS	Integer	16688	Telescope focus
TIMESYS	String	UTC approximate	Time system
DATE-OBS	String	2006-09-05	Date of observation start
TIME-OBS	String	08:17:3.988	Time of observation start
MJD-OBS	Double	53983.34519	MJD of observation start
EXPTIME	Float	354.43	Exposure time (s)
EXPREQ	Float	360.00	Exposure time request (s)
DARKTIME	Float	359.34	Dark time (s)
EPOCH	Float	2007.054	Epoch of observation
WINDSPD	Float	5.4	Wind speed (mph)
WINDDIR	Float	24	Wind direction (deg)
AMBTEMP	Float	11.9	Ambient temperature (deg C)
HUMIDITY	Float	16	Ambient relative humidity (percent)
PRESSURE	Float	783	Barometric pressure (mbar)
SKYVAR	Float	2.54	Sky variation from IR camera (percent)
FLUXVAR	Float	0.51	Guide star flux variation (percent)
CHECKSUM	String	0WDA3T940TA90T99	Header Checksum for verifying file
CHECKVER	String	complement	Checksum software version
IMAGEHWV	String	SISPI hw 1.2.3.4	SISPI Hardware Version
IMAGESWV	String	SISPI sw 4.3.2.1	SISPI Software Version

Table 2-4: DECam Exposure Header for Image Extensions (1-62)

Keyword	Data Type	Example	Comment
XTENSION	String	IMAGE	Extension type
SIMPLE	Logical	T	File does conform to FITS standard
BITPIX	Long	16	Number of bits per data pixel
BZERO	Long	32768	Zero shift for unsigned short
BSCALE	Long	1	Scale factor for unsigned short
NAXIS	Long	2	Number of data axes
NAXIS1	Long	2104	Length of axis 1
NAXIS2	Long	4096	Length of axis 2
EXTEND	Logical	T	FITS dataset may contain extensions
INHERIT	Logical	T	Inherits global header
CCDNAME	String	FNAL #9817eFABR14-08	CCD Name
CCDNUM	Long	1	CCD Number
DETSEC	String	[1:2048,1:4096]	Detector Section
CCDSUM	String	1 1	CCD pixel summing
DATASEC	String	[51:2098,1:4096]	Data Section of CCD
TRIMSEC	String	[51:2098,1:4096]	Trim Section of CCD
AMPSECA	String	[1:1024,1:4096]	Amp A Section
BIASSECA	String	[1:50,1:4096]	BIAS Section for Amp A
GAINA	Float	2.7	Gain for Amp A
RDNOISEA	Float	7.8	Read Noise for Amp A
SATURATEA	Float	40000.0	Saturation for AmpA
AMPSECB	String	[1025:2048,1:4096]	Amp B Section
BIASSECB	String	[2098:2148,1:4096]	BIAS Section for Amp B
GAINB	Float	2.7	Gain for Amp A
RDNOISEB	Float	7.8	Read Noise for Amp A
SATURATEB	Float	40000.0	Saturation for AmpA
EQUINOX	Float	2000.0	Equinox of WCS
WCSDIM	Long	2	WCS Dimensionality
CTYPE1	String	RA---TAN	Coordinate Type
CTYPE2	String	DEC--TAN	Coordinate Type
CRVAL1	Double	356.759700000	Coordinate Reference Value
CRVAL2	Double	-43.749901000	Coordinate Reference Value
CRPIX1	Double	13240.2333710	Coordinate Reference Pixel
CRPIX2	Double	6274.06658936	Coordinate Reference Pixel
CD1_1	Float	7.500000E-05	Coordinate Matrix
CD1_2	Float	9.511111E-07	Coordinate Matrix
CD2_1	Float	9.511111E-07	Coordinate Matrix
CD2_2	Float	7.500000E-05	Coordinate Matrix
PV1_0	Float	0.000000	PV Distortion Coefficient
PV1_1	Float	1.000000	PV Distortion Coefficient
PV1_2	Float	0.000000	PV Distortion Coefficient
PV1_3	Float	0.000000	PV Distortion Coefficient
PV1_4	Float	0.000000	PV Distortion Coefficient
PV1_5	Float	0.000000	PV Distortion Coefficient
PV1_6	Float	0.000000	PV Distortion Coefficient
PV1_7	Float	0.002000	PV Distortion Coefficient
PV1_8	Float	0.000000	PV Distortion Coefficient
PV1_9	Float	0.002000	PV Distortion Coefficient
PV1_10	Float	0.000000	PV Distortion Coefficient
PV2_0	Float	0.000000	PV Distortion Coefficient
PV2_1	Float	1.000000	PV Distortion Coefficient
PV2_2	Float	0.000000	PV Distortion Coefficient
PV2_3	Float	0.000000	PV Distortion Coefficient
PV2_4	Float	0.000000	PV Distortion Coefficient
PV2_5	Float	0.000000	PV Distortion Coefficient
PV2_6	Float	0.000000	PV Distortion Coefficient
PV2_7	Float	0.002000	PV Distortion Coefficient
PV2_8	Float	0.000000	PV Distortion Coefficient
PV2_9	Float	0.002000	PV Distortion Coefficient
PV2_10	Float	0.000000	PV Distortion Coefficient
CHECKSUM	String	0WDA3T940TA90T99	Header checksum for validating file
DATASUM	String	0WDA3T940TA90T99	Data checksum for validation file
CHECKVER	String	complement	Version of checksum software

2.2.2 FileClass *red*

There are several types of data within imageclass red, including the crosstalk corrected raw images (*raw_obj, raw_bias, raw_dflat, raw_sflat, raw_dark*) ready for additional processing, and three types of images produced during further processing: (1) reduced images, (2) calibration images, (3) remap images and (4) difference images. In addition, there are a whole host of log files, parameter files, xml pipelines and properties files stored within this fileclass. We describe the image elements of these below.

2.2.2.1 Crosstalk Corrected CCD Images

DECam *src* data are crosstalk corrected and split into 62 individual FITS files using the DESDM Crosstalk module. The image splitting is carried out by a program called *DECam_crosstalk*, and this splitting enables data parallel processing. We call these CCD images raw data. We store the raw data as 32 bit floating point images, and each is 2104x4096 in size (33MB uncompressed). The header of the raw data contains the union of the keywords in the 0^{th} extension header and the appropriate image extension header from the original src data.

In addition to these header parameters, the DESDM system adds several others. These are listed in Table 2-5 below.

Table 2-5: Raw data header keywords			
Keywords/Description	Data Type	Example	Comment
All 0^{th} header keywords			
All image extension header keywords			
FILENAME	string	Exposure_##.fits	Image name
PHOTFLAG	long	1	Night Photometric (1) or not (0)
DESDCVT	string	Date/time	DECam image conversion
COMMENT	string	DECam_crosstalk ${1} ${2} –crosstalk <file>	
DES_EXT	string	img	Extension type

For a description of the Crosstalk module, the *DECam_crosstalk* processing and the crosstalk correction algorithm, please see the DESDM Technical Design Report. This processing is carried out the same way for all types of src images (regardless of whether they are science or calibration exposures).

2.2.2.2 *red* Images

The DESDM system employs *imcorrect* to process raw object images (i.e. science targets) into reduced images. Table 2-6 describes the header. Keywords are added for each processing step that is carried out, and the actual *imcorrect* call is included as a long comment. Note that typically reduced images have associated weight maps and bad pixel maps. The headers for those extensions are quite empty, including the DES_EXT=bpm or wmp as appropriate.

Table 2-6: Reduced data header keywords for *img* extension			
Keywords/Description	Data Type	Example	Comment

All keywords are copied from the raw images into the reduced images			
Then the following keywords are deleted: CCDSUM, TRIMSEC, AMPSECA, AMPSECB, BIASSECA, BIASSECB, GAINA, GAINB, RDNOISEA, RDNOISEB, SATURATA, SATURATB			
The following keywords are then added			
OBSTYPE	string	reduced	Observation Type
SATURATE	float	43426.0	Global Saturate Value
SKYBRITE	float	2523.65	Sky Brightness (ADU)
FWHM	float	1.150	Stellar FWHM
Depending on what kind of processing, the following keywords are added:			
DESOSCN	string	Date/time	Overscan corrected
DESBIAS	string	Date/time	Bias subtracted
DESFLAT	string	Date/time	Flat fielded
DESILLUM	string	Date/time	Illumination Corrected
DESFRINGE	string	Date/time	Fringe Corrected
COMMENT	string	imcorrect <inlist> <outlist> -bias <image>	
DES_EXT	string	img, bpm, wmp	Extension type
SCAMPCHI	float	1.5037	Chi2 of SCAMP solution
SCAMPNUM	int	2548	Number of standards used by SCAMP
SCAMPFLG	int	0	Number of SCAMP warnings

2.2.2.3 Calibration Images

The calibration data products produced by *mkbiascor*, *mkflatcor*, *mkdarkcor*, *mksupersky* and *mkillumcor* have much more minimal headers. Table 2-7 contains a listing of the header keywords inserted by *mkflatcor*.

Table 2-7: Header Keywords for Calibration Products from *flatcombine*

Keywords/Description	Data Type	Example	Comment
OBSTYPE	string	dome flat	Observation Type
FILTER	string	g	Filter name(s)
CCDNUM	int	27	CCD Number
DESFCMB	string	date/time	flatcombine output
COMMENT	string	flatcombine <inlist> <outimage> -bias <image>	
DES_EXT	string	img, bpm, wmp	Extension type

The headers from the other programs are similar. For mkbiascor there is no FILTER keyword, the keyword DESMKBCR is added to note the image original, and OBSTYPE is set to *biascor*. For mksupersky the same header keywords are used as in Table 2.7 except that DESMKFCR is replaced with DESCOMB and OBSTYPE is set to *supersky*. Things are similar with mkillumcor; DESMKFCR is replaced by DESMKILC and OBSTYPE is set to *illumcor* for the illumination correction image. For the fringe correction image (useful for quality assurance purposes), DESMKFCR is replaced by DESMKFRG and OBSTYPE is set to *fringecor*.

2.2.2.4 *remap* Images

The *remap* images are produced from the *reduced* images to enable rapid coaddition at a later stage in the processing. The first step in remapping is a database query to determine which coadd tile(s) the reduced image overlaps. The astrometrically corrected WCS information for the image and a project specific set of coadd tiles are used in this process. Metadata descriptions of these coadd tiles lie within the DES Archive database *coadd_tiles* table (e.g. tangent plane, central Right Ascension and Declination, pixel size and number of pixels along each axis). For DES the coadd tiles are approximately 1 deg^2 and cover the entire survey region; neighboring tiles overlap by 2 arcminutes, which minimizes the cataloging problems that arise near the image boundary. Therefore, the total data volume of the *remap* images is somewhat larger than the data volume of the *reduced* images.

The remap images have two extensions: one for the image data and another for the weight map. Only a fraction of the header parameters within the *reduced* images are copied into the *remap* images. These parameters are listed in Table 2-8.

Table 2-8: Header Keywords for *remap* Images

Keywords/Description	Data Type	Example	Comment
OBJECT	string	CDFS	Object name
OBSERVER	String	John Smith	Observer(s)
FILTER	string	g	Filter name(s)
CCDNUM	int	27	CCD Number
TELRADEC	string	fk5	Telescope coordinate system
TELRA	string	05:23:45.1	Telescope RA (hours)
TELDEC	string	-45:27:13	Telescope DEC (deg)
TELQUIN	string	2000.0	Coordinate equinox
FILENAME	string	image.fits	Image filename
AIRMASS	float	1.150	Airmass
DATE_OBS	string	2007-05-23T05:53:40.2	Date and time of observation start
EXPTIME	float	100.0	Exposure time
DARKTIME	float	103.50	Dark time
SATURATE	float	61245.0	Global Saturate Value
DETECTOR	string	DECam	Detector
TELESCOP	string	Blanco 4m	Telescope name
OBSERVAT	string	CTIO	Observatory name
SKYBRITE	float	1043.32	Sky Brightness (ADU)
PHOTFLAG	int	1	Night Photometric (1) or not (0)
SCAMPCHI	float	1.5670	Chi2 of SCAMP solution
SCAMPFLG	int	0	Number of SCAMP warnings
SCAMPNUM	int	2527	Number of standards used by SCAMP
FWHM	float	1.10	Stellar FWHM
WCS parameters of the same form as in the *raw* and *reduced* data			

2.2.2.5 *diff* Images

Difference images are produced from *reduced* images by remapping a template (*coadd*) image to match the projection of the reduced image, scaling the template and reduced images to a common photometric scale and then subtracting the template image from the reduced image. These difference images and the associated weight map and bpm are then cataloged to select sources that have exhibited significant changes in flux.

At this time we do not produce difference images. We expect that the header structure of the difference images to be similar to that of the remap images.

2.2.3 ImageClass *coadd*

Combining the remapped data into deeper images of the sky produces the *coadd* data. The *remap* data are produced from the *reduced* data during the nightly processing, and during the coaddition process only an additional PSF homogenization followed by simple coaddition and filtering is required. The coadd images are large (1 deg^2) tangent plane projections with pixel scale set to be the same as the native DECam pixel scale.

In contrast to the *reduced* and *remap* images, the coadd data contain only two extensions: the image and an associated weight map. The pixel weight is turned to zero for the masked pixels.

Table 2-9 contains a listing of the header keywords for the coadd images.

Table 2-9: Header Keywords for *coadd* Images

Keywords/Description	Data Type	Example	Comment
OBJECT	string	CDFS	Object name
OBSERVER	String	John Smith	Observer(s)
FILTER	string	g	Filter name(s)
CCDNUM	int	27	CCD Number
TELRADEC	string	fk5	Telescope coordinate system
TELRA	string	05:23:45.1	Telescope RA (hours)
TELDEC	string	-45:27:13	Telescope DEC (deg)
TELQUIN	string	2000.0	Coordinate equinox
FILENAME	string	image.fits	Image filename
AIRMASS	float	1.150	Airmass
DATE_OBS	string	2007-05-23T05:53:40.2	Date and time of observation start
EXPTIME	float	100.0	Exposure time
DARKTIME	float	103.50	Dark time
SATURATE	float	61245.0	Global Saturate Value
DETECTOR	string	DECam	Detector
TELESCOP	string	Blanco 4m	Telescope name
OBSERVAT	string	CTIO	Observatory name
SKYBRITE	float	1043.32	Sky Brightness (ADU)
PHOTFLAG	int	1	Night Photometric (1) or not (0)
SCAMPCHI	float	1.5670	Chi2 of SCAMP solution
SCAMPFLG	int	0	Number of SCAMP warnings
SCAMPNUM	int	2527	Number of standards used by SCAMP
FWHM	float	1.10	Stellar FWHM

ELLIPTICITY	Float	0.11	Stellar ellipticity
WCS parameters of the same form as in the *raw* and *reduced* data			

2.2.4 FileClass *cal*

The cal data are required for the processing of the DECam images. These images are typically produced periodically over the observing season rather than nightly like the calibration images discussed above. Currently there are only a few kinds of cal images. One is the pupil ghost correction image, which is used to subtract off the scattered light contribution to the sky brightness distribution. The second is the bad pixel mask. The bad pixel mask is produced using the program *mkbpm* that takes as input a flat field correction (*flatcor*) and a bias correction (*biascor*). The header parameters for these correction images is currently quite minimal, following the structure of the nightly correction images (see Table 2-7). A third file is the ascii file containing the crosstalk coefficients for the detector. These are needed by DECam_crosstalk.

We also store template flat field, bias and illumination corrections in this portion of the archive. These are used in the Quality Assurance, providing a comparison set to which the nightly calibration images can be compared.

2.2.5 FileClass *cutout*

An image cutout service will provide collaboration access to image regions centered on objects or any specified sky position. We have not yet done any prototyping of this service. In our initial design we have four different types of cutout images. These types are named according to the source images from which they are derived: (1) reduced, (2) remap, (3) coadd and (4) difference. The coadd cutouts will be simple two extension MEFs containing the image and weight maps. The reduced, remap and difference coadds will be multi-extension, containing images and weight maps for each epoch requested by the Archive Portal user.

The contents of the cutout image headers have not yet been defined.

2.3 Image MetaData

Image metadata are stored in the DES Archive database. The core repository for image metadata is in the IMAGE, EXPOSURE, CATALOG and COADD tables, but metadata are also present in the DES_EXPOSURES and the ZEROPOINT table. The DESDM Technical Design Report contains a listing of the schema for our database. The guiding principle is that all files are tracked in the archive by storing metadata in the LOCATION table. Science metadata associated with FITS images and catalogs are stores in the filetype specific metadata tables listed above.

As is clear from the previous discussion in this section, DESDM also relies upon the FITS header as a simple repository for the metadata associated with particular images (and also catalogs). Typically, during processing image metadata are updated in the image headers and then during ingestion stages these revised metadata are captured and inserted into the DES Archive database. We have designed our system so that reading the actual FITS image files to obtain the needed metadata is never required—all critical information is stored in the database where it can be retrieved without any DES Archive filesystem access. Our approach offers important advantages in speed and efficiency.

3 Catalog Data

The catalog data are perhaps, of all data products described in this document, the closest to the science analyses. Because the DES Science Working Groups (SWGs) have just recently begun their development activities, the specific preferences of the SWGs for catalog parameters are not yet fully known. Thus, DESDM expects these details to change over the coming year. Below we first describe the catalog data formats, the locations where these data reside, and then the provide more detail about the cataloging of single epoch and coadd images.

3.1 Catalog Data Formats and Repositories

Catalogs produced in the pipelines will be written to binary FITS tables. The SExtractor cataloger and other Terapix software like SCAMP use a specific type of FITS binary table for catalog storage: the LDAC catalog. These FITS binary tables store data efficiently and mesh nicely with our codes, which have adopted CFITSIO.

Typically speaking, these catalogs are small compared to the images, weight maps and bpms from which they are derived. Therefore, our strategy is to simply keep these binary FITS catalogs along with the image data products within the DES Archive. They are ingested along with the image data, and users can select and retrieve them through the DES Archive Portal in much the same manner that users can select images. While these FITS catalogs convenient repositories for the data, they do not enable efficient calibration or analysis of the catalog data. For that we turn to a relational database.

We ingest all binary FITS catalogs into tables within the DES Archive database. The object tables have indices built on a sky location and the identifier of source image so that they can be rapidly searched and filtered using parameters from the parent images (i.e. the FWHM or ellipticity of stars within the image). The DES Archive database will also support higher level science queries.

The DES Archive Portal will deliver catalog data in several different formats, including the binary FITS catalogs that were written in the cataloging process. When catalog data are selected from the object tables of DES Archive database queries, the user will be able to choose among several different formats. The include ASCII files, FITS binary tables and VO tables. Our system is modular, and so adding additional formats as they are requested by the collaboration will not be a challenge.

There are several different types of catalog data. Below we discuss these data according to whether they are derived from single epoch or coadd images.

3.2 Objects from Single Epoch Images

Reduced, remap and difference images are single epoch images derived from a single DECam exposure (through a process that include subtraction from a template image in the case of the difference image). These images are cataloged, and the catalogs are ingested into the DES Archive database. We discuss the catalogs from each of these images separately below.

3.2.1 *reduced* Images

3.2.1.1 Basic Cataloging and Model Fitting Photometry

During nightly processing the reduced or detrended images are cataloged using SExtractor and its extension that carries out model fitting to each object. Approximately 40 parameters are extracted for each object, including positional, morphological and photometric measurements. Photometry is calculated in instrumental magnitudes, because the photometric calibration occurs in the database after the nightly processing and ingestion are complete. The photometry extracted for each object includes (1) MAG_AUTO, (2) aperture photometry within six different apertures ranging from 2 to 12 arcsec diameters, (3) PSF fitting photometry for stars, and (4) model fitting photometry for galaxies that includes the effects of the locally measured PSF. The single epoch photometry extracted by the nightly processing from reduced images is carried out independently in each image (and therefore in each band).

These catalogs are ingested into the OBJECTS table of the DES Archive database. We do not include a complete list of the parameters extracted from each objects, because (1) this list appears in the database schema presented in the DES Technical Design Report and (2) we expect this parameter list to evolve in response to the needs expressed by the SWGs. Note that a particular object on the sky generates a new row in the OBJECTS table each time it is observed. These rows are later coupled to the appropriate "parent object" row in the COADD_OBJECTS table (see Section 3.3 below).

3.2.1.2 Shapelet Expansions

During the nightly processing, shapelet expansions are extracted for all objects. These expansions are carried out to higher order for isolated stellar objects, which are considered to be "PSF Standards" within the weak lensing framework. These shapelet expansions are then ingested into the DES Archive database. The shapelet expansions for the PSF Standards will be used at a later time to build a model for the PSF variation over each image. The shapelet expansions for the galaxies could be valuable for morphological studies or even estimates of galaxy photometry.

3.2.1.3 Multi-Epoch Cataloging

In collaboration with the Weak Lensing SWG, DESDM is developing the weak lensing pipeline. The strategy for the weak lensing pipeline involves extracting the PSF corrected shear from each galaxy by simultaneously fitting to all single epoch images in the band(s) of choice. This approach relies on a local model of the PSF that is derived using a principle component analysis of the PSF variation across and among all single epoch images of interest within the survey. The PSF Standards that are shapelet expanded and ingested during nightly processing (see previous section) are critical to this process.

There are a wide range of postprocessing options for cataloging that we will explore if needed to meet the photometric accuracy requirements in the DES Science Requirements document. For example, it is possible to create a postprocessing pipeline that would use all available single epoch images to extract the best photometric estimates for galaxies using all available constraints. One could model fit to each galaxy using all images in a particular band and perhaps even coupling some parameters (i.e. the galaxy position) across different bands. This approach could be quite effective in deblending the light in high density regions of the sky such as the cores of galaxy clusters and could well deliver the highest precision photometry.

3.2.2 *remap* Images

The remap images are cataloged using SExtractor during the nightly processing. These catalogs are then ingested into the DES Archive database for use in determining relative flux scales among pairs of overlapping images. To determine the flux scale ratio, we use aperture photometry of isolated stars within a narrow range of magnitudes. We ingest only the relevant objects in each remap catalog to avoid needless duplication (because the full catalogs for the reduced images from which the remap images were derived are ingested).

3.2.3 *diff* Images

The difference images are also cataloged and ingested into the DES Archive database. DESDM is working with the SNe SWG to deploy the components required for difference imaging into our nightly processing pipeline. We have not yet defined the parameters that will be cataloged, but in principle we could use SExtractor to carry out forced photometry of each difference image using a high signal to noise template image as a detection image. Object photometric measurements and the image of origin could then be stored in a difference imaging OBJECTS table.

3.3 Objects from Coadd Images

We use SExtractor and its extension to catalog coadd images. Coadd images are built in a processing step that occurs after nightly processing has been successfully executed on a significant number of nights (i.e. after the season). Coadd images in each band *grizY* are built for each tile. This process is described in greater detail in the DESDM Technical Requirements Document. The cataloging approach we currently use is to create a χ^2 image from the weighted sum of the coadd images (one for each band) within each tile. We then run SExtractor in dual image mode, using the χ^2 image for object detection and the coadd image in each band for measurements of photometric and morphological parameters. We collate these separate binary FITS catalogs during the ingestion process, resulting in a single row for each object in the COADD_OBJECTS table in the DES Archive database. Thus, the COADD_OBJECTS table is more closely related to the SDSS catalogs in the sense that photometry and morphological information is presented for each band within a single row of the table.

The object parameters extracted in the coadd cataloging are currently the same as those extracted in the single epoch cataloging, but they are extracted for each band. We do not provide a full listing of the coadd objects parameters here, because the DESDM Technical Design Report contains the DES Archive database schema. Also, as mentioned previously, we expect this parameter list to evolve as the DES SWGs begin to build their analysis codes and use them to analyze the simulated data DESDM has been processing and cataloging in the data challenges.

4 Survey Metadata

The DES main survey has a solid angle of 5000 deg^2 and will be fully imaged multiple times in five different photometric bands: *grizY*. The main survey then consists of approximately 50,000 pointings (i.e. 5000 deg^2 / 3 deg^2 $pointing^{-1}$ * 5 bands x ~5 layers). Not every image taken with DECam during the DES survey time will meet the data quality requirements, which include ranges of acceptable seeing and extinction. Therefore, we will track the progress of the survey using a table in the DES Archive database. Each pointing of the DES survey will be listed in this DES_Survey table, and as we process survey images we will record the data quality parameters for those images. These data quality parameters will allow the collaboration track the status of the survey at any time and to simply select pointings with appropriate band and location for initial or repeat observation.

The fields of the DES_Survey table have not yet been finalized; Table 4-1 contains an tentative list.

Table 4-1: Survey Table

Field	Datatype	Description
PROJECT	VarChar(20)	Project name (DES)
POINTING_NAME	VarChar(20)	Name (DESA0520-5523)
RA	BinaryDouble	Pointing Right Ascension (degrees; 75.12345678)
DEC	BinaryDouble	Pointing Declination (degrees; -41.23456789)
EQUINOX	BinaryFloat	Equinox (2000.0)
LAYER	Integer	Layer of imaging
BAND	VarChar(10)	Observing band (grizY)
EXPOSURE	BinaryFloat	Exposure time (seconds; 100.0)
DATE_OBS	VarChar(20)	Date and time of observation
OBSERVER	VarChar(20)	Observer(s)
OPERATOR	VarChar(20)	Telescope Operator
PSF_FWHM_MN	BinaryFloat	Mean of PSF FWHM among 62 CCDs
PSF_FWHM_RMS	BinaryFloat	RMS of PSF FWHM among 62 CCDs
PSF_ELLIPTICITY_MN	BinaryFloat	Mean PSF ellipticity among 62 CCDs
PSF_ELLIPTICITY_RMS	BinaryFloat	RMS of PSF ellipticity among 62 CCDs
SKYBRITE_MN	BinaryFloat	Mean sky brightness among 62 CCDs
EXTINCTION	BinaryFloat	Extinction at center of field
PHOTOMETRIC	BinaryFloat	Photometric indicator (i.e. from IR camera)
DECAM_STATUS	Integer	Camera status red flags present or not

5 Additional Collaboration Data

The DES Simulation team and the DES SWGs will store data products in the DES Archive. Below we briefly describe these contributions.

5.1 DES Simulation Team

The DES Sim team (H. Lin, C. Stoughton, N. Kuropatkin, Fermilab) produces simulated DECam image datasets that the DESDM team uses in their data challenges. During survey operations a program of image simulation will likely be a very effective way of measuring the performance of the DESDM system and of estimating completeness and other aspects of the survey dataset as a function of position on the sky. The data products from the Sim team will be stored in the DES Archive (primarily at the Fermilab node, but also elsewhere as needed).

5.1.1 Simulated Catalogs

DESDM stores the object catalogs that serve as input to the image simulations within the DES Archive database. Specific tables are prepared for each generation of simulation data, such as the CATSIM3_TRUTH table for the catalogs used in Data Challenge 3. These tables serve as truth tables for comparison to the catalogs that arise from DESDM processing. These comparisons are the basis for the tests that determine whether our codes are meeting the requirements defined in the DES Science Requirements and DESDM Requirements and Technical Specifications document.

5.1.2 Simulated DECam Images

New sets of DECam images with ever increasing levels of realism are created for each of the DESDM Data Challenges (DCs). At present these image datasets are typically 500 deg^2 in sky coverage observing in all DES bands (*griz* for DC3) with 4 layers of imaging for a total of 8000 deg^2 of total imaging data. These data are divided into 10 observing nights and coupled with bias frame and dome flat calibration data to create simulated observing nights that are comparable to those we expect during the DES survey. The scale of these simulated datasets is vast—comparable to 20% of the entire SDSS imaging dataset in 10 nights of observing—and impressively deep, delivering photometry on objects 10 times fainter than those detected in SDSS. DESDM now processes these SDSS scale datasets routinely in our yearly data challenges. These simulated images are the *src* image data described in Section 2 above.

5.2 DES SWGs

No planning has been finalized, but it is clear that the DES SWGs will create cluster and SNe catalogs as well as many other science data products. The natural place to store these for collaboration access will be the DES Archive. DESDM resources will be made available to aid the SWGs in storing their data. As these plans are finalized, we will update this section of the document.

6 External Collections

External datasets are used for the image cataloging and calibration during the nightly processing, to characterize the operation of the pipeline, to estimate the completeness of images, and to enable scientific analyses of the DES data. Many external datasets are stored within the DES Archive database. Below we include a discussion of these datasets broken in to broad categories of (1) Astrometric Calibration, (2) Photometric Calibration, (3) Simulated Catalogs, and (4) Complementary Science Collections.

6.1 Astrometric Calibration

The DES Archive database contains tables holding several different astrometric catalogs. We have stored all the available information for each of these catalogs. Currently we have tables for the USNO-B2 survey, NOMAD and the UCAC survey. We routinely use the USNO-B2 survey for astrometric calibration in our data challenges and in processing the BCS science data.

6.2 Photometric Calibration

Currently photometric standards from both the SDSS southern stripe and the Smith, Tucker & Allam souther ugriz standard star calibration program are stored within a single table. We routined calibration the images for data challenges and from the BCS survey using this table.

6.3 Complementary Science Collections

At this time the DES Archive database contains no complementary science collections. It is expected that other datasets such as catalogs derived by the South Pole Telescope (SPT) team and the VISTA VHS-DES team will be present in the archive. The DES Archive will contain the available spectroscopic redshifts measured within the southern sky, because these will be key in the calibration and testing of the photometric redshift pipeline. In addition, published source catalogs in the radio, optical and X-ray will be particularly valuable. No catalog available today approaches the size of the DES catalog, and so data volume is not a central concern.

There will very likely also be complementary science collections in image form. These would enable optimal multi-wavelength analyses of sources, including multiwavelength cluster finding.

Appendix

A. DES Archive Data Volumes

The size of the initial survey and the variety of DESDM data products lead to a large collection of data in the DES Archive filesystems and in the DES Archive database. By the end of the survey the raw data and a complete set of processing on those data will lead to a 1.5PB dataset. The full dataset including the seasonal reprocessing will reach 4PB. In addition, the DES Archive database will exceed 100TB.

A.1 DES Archive Filesystem Collections

The yearly accumulation of data is tabulated below. Note that these data volumes are in TB and refer to *uncompressed* data products. The image data types Src, Reduced, Remap, Diff and Coadd are all described in Section 2.1. The **Src** data volume assumes that DES acquires 300GB each of 105 nights in each season. We estimate a long night will produce 360GB, corresponding to 60 calibration exposures (10 bias plus 10 dome flats in each filter grizY) plus 300 science exposures (10hrs observing with 1 exposure every 120s). Thus, our estimate of the seasonal volume assumes we lose are not operating 20% of the observing time. The volumes of the other data types follow directly from this. **Reduced** images are larger than the Src science images, because they are 32 bit instead of 16 bit. In addition, there may be associated weight and bad pixel masks. We examine the possible image configurations. The volume of **Remap** images is larger than the reduced dataset by a factor of $(62/60)^2=1.0678$ because neighboring remap tiles overlap by 2 arcminutes. Difference images (**Diff**) are the same size as the reduced images. **Coadd** volumes derive from having 5000 tiles, each 13800^2 pixels with images in five observing bands grizY and one detection image. The coadd images also have associated weight maps and bad pixel masks. **Catalogs** and all other files associated with the image processing are estimated at 10% of the processed image volume (does not include weight maps and bad pixel masks). The **Total** column contains the sum of images and catalogs. Reprocessing will produce a second set of reduced, remap and diff images and associated catalogs, and the volume is denoted in the

Reprocess column. Finally, the **Sum** of the seasonal and reprocessing data appears in the final column.

Table A-1: Seasonal Data Volumes

Data Type	Src	Reduced	Remap	Diff	Coadd	Catalog	Total	Reprocess	Sum
32bit	30.2	50.4	53.8	50.4	22.9	17.7	225.5	155.0	380.4
64bit	30.2	100.8	107.6	100.8	45.7	17.7	402.9	257.3	660.3
80bit	30.2	126.0	134.5	126.0	57.1	17.7	491.7	295.5	787.1

We show three sets of Seasonal Data Volumes. The first row (32 bit) corresponds to the case where all processed image data products consist of a single 32 bit image. The second row (64bit) is the case where each processed image has an associated 32 bit weight map. The third row is the case where each processed image has a 32 bit weight map and a 16 bit bad pixel mask. Again, none of these volumes include compression, which may lead to a factor of two reduction in image volumes.

Using the Seasonal Data Volumes from above we estimate the Cumulative Data Volume in Table A-2 below. To do this we assume that we keep (1) the source data, (2) the nightly processed data and associated coadds and (3) the seasonally reprocessed data. We show the cumulative volumes for the same three cases as above: (1) 32 bit data, (2) 64 bit data and (3) 80 bit data. Again, the volumes are in TB and are for uncompressed data products.

Table A-2: Cumulative Data Volumes

Season	32bit	64bit	80bit
1st	380.4	660.3	787.1
2nd	760.8	1320.5	1574.3
3rd	1141.2	1980.8	2361.4
4th	1521.6	2641.1	3148.6
5th	1902.1	3301.3	3935.7

We expect to operate with the 80bit data format. In addition, we will store all versions of the processed data products. Therefore, the DES Archive must be prepared to handle substantial data volumes of approximately 4PB. The bulk of these data will be stored on tape at the primary archive site at NCSA. We will grow our spinning disk over the course of the survey to ensure that the reduced and remap datasets are available for quick processing and analysis using the coadd and weak lensing pipelines. Table A-1 shows that these datasets grow at approximately 250TB per year. Archive users can access all DES data, but the latency for the data stored on tape will be much higher than for that stored on disk.

A.2 DES Archive Database

The Archive volume is dominated by the object tables. These include the OBJECTS table that holds all the objects extracted from the single epoch images, the COADD_OBJECTS table that holds objects extracted from the coadd data, and the DIFF_OBJECTS table that holds objects selected in the difference imaging. Of these tables, the OBJECTS table is by far the largest. We estimate that 1kB of information is stored for each object, and that each layer of imaging over the survey region produces approximately 1 billion objects (typically 50 objects/arcmin^2). We will image the survey region in five bands with a minimum of five imaging layers in each band, and as many as 20 layers of imaging in *i* band. This corresponds to approximately 40TB for the

OBJECTS table alone. If the cataloged objects from all seasonal reprocessing are stored, the DES Archive database will exceed 100TB. A variety of techniques will be used to improve performance of the database in the face of such a large dataset. These include table partitioning and perhaps even the generation of new tables during reprocessing. The separate tables could be joined for queries aimed at comparing two sets of reprocessing, but typically users will be seeking the latest reduction with the best available calibration. Nevertheless, all scientifically useful versions of the data will be stored.