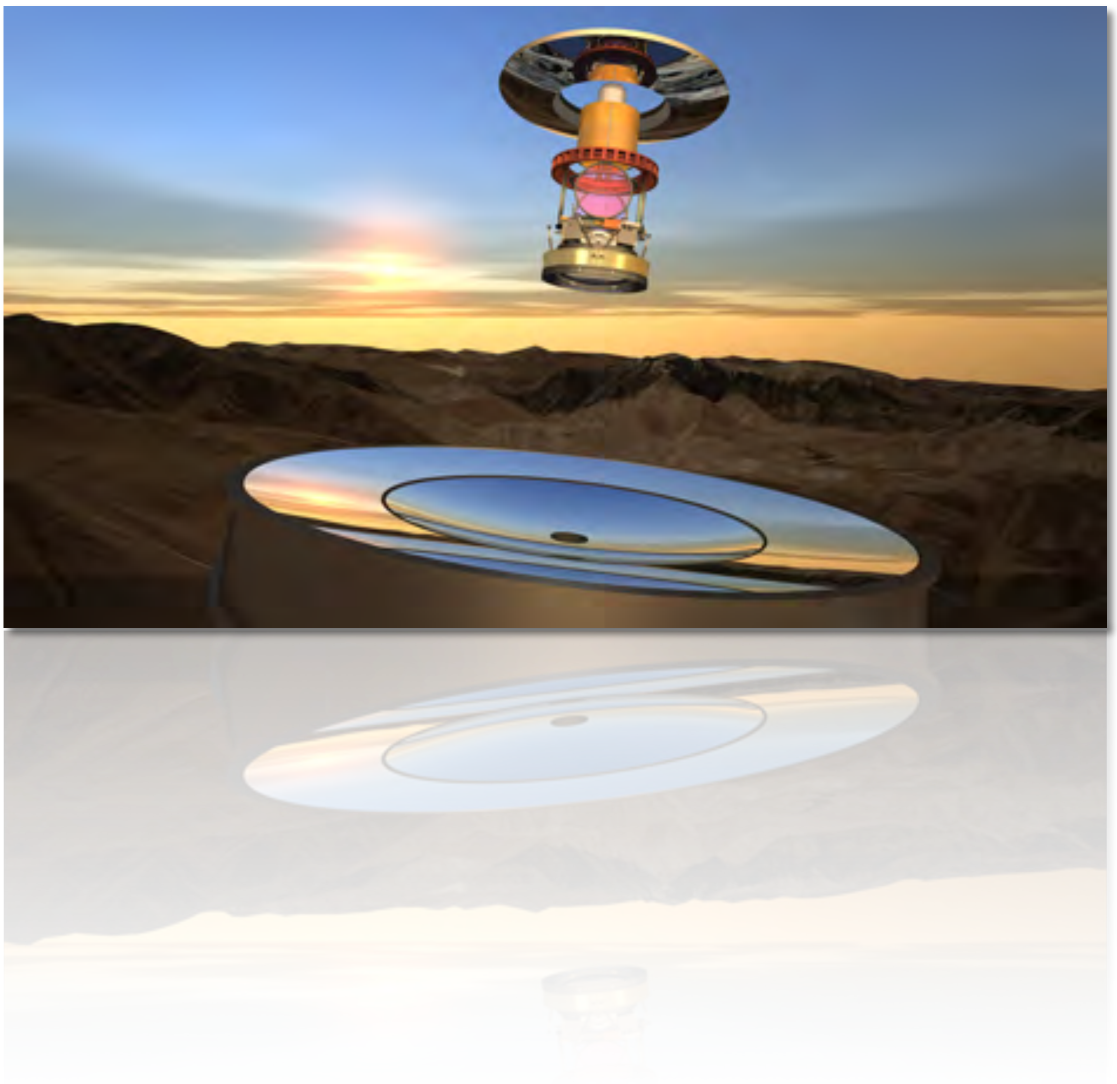
LSST DATA CHALLENGE HANDBOOK

WINTER 2013 EARLY DATA RELEASE



933 N. Cherry Avenue, Tucson, AZ 85726, USA · http://www.lsstcorp.org/

User Support

There are multiple sources of help for accessing or understanding LSST Data Challenge products and services. Please see *Chapter 1.4.* Obtaining Help for details. For help with understanding details of the processing algorithms, or with problems with data access, please e-mail the Data Management Help Desk:

E-mail: dm-help@lsst.org

Users may also find the Science Collaboration Wiki to be helpful.

Science Wiki: http://www.lsstcorp.org/sciencewiki/index.php?title=Main_Page

User Forum: https://www.lsstcorp.org/sciencewiki/index.php?title=Special:AWCforum

For web-based access to data, see:

Image access: http://lsst-web.ncsa.illinois.edu/lsstdata/dr-w2013/

Stripe 82 Image Viewer: http://moe.astro.washington.edu/sdss/

Catalog access: https://lsst-web.ncsa.illinois.edu/mydb/

To obtain a DB (catalog) access account: http://lsst-web.ncsa.illinois.edu/dbaccount

Winter 2013 Software release:

http://dev.lsstcorp.org/trac/wiki/Installing/Winter2013/

Revision History

Version	Date	Editors
1.0	2011 January	Richard A. Shaw, Michael A. Strauss
1.1	2011 August	Richard A. Shaw, Michael A. Strauss
2.0	2012 August	Richard A. Shaw
2.1	2013 January	Richard A. Shaw

This *LSST Data Challenge Handbook* is available on the Science Wiki (which requires a login):

Handbook: https://www.lsstcorp.org/sciencewiki/images/DC_Handbook_v2.1.pdf

Citations to this *Handbook* should read:

Shaw, R. A., ed. 2013, *LSST Data Challenge Handbook* (Version 2.1; Tucson, AZ: LSST Corp.)

The front cover features an image of the LSST mirrors, with people for scale, created from mechanical drawings by Todd Mason, Mason Productions, Inc.

LSST is a public-private partnership. Design and development activity is supported in part by the National Science Foundation. Additional funding comes from private foundation gifts, grants to universities, and in-kind support of Department of Energy laboratories and other LSST Member Institutions. The project is overseen by the LSST Corporation, a non-profit 501(c)3 corporation formed in 2003, with headquarters in Tucson, AZ.

Table of Contents

Preface

Scope of This Document

The purpose of this *Handbook* is to describe the LSST Data Management processing, and the data products it produces, in enough detail that a Science Collaboration member who is not familiar with them can evaluate their quality, scientific fidelity, and suitability to support their science goals with LSST. One of the key aims of the *Handbook* is to provide one-stop shopping for vital background material for the input data and the processing algorithms, with references to more detailed information where it exists. However, this *Handbook* is not intended to replace requirements documents, technical specifications for hardware, or software design documents. It is also not a user guide for data analysis, although advice is offered for some software that may be useful. An excellent overview of the LSST mission and technical capabilities can be found in the *LSST Science Book*, which is available at http://adsabs.harvard.edu/abs/2009arXiv0912.0201L. This document describes the data challenges in general and the accomplishments to date (see Chapter 1), the content and packaging of the data products (images, catalogs, and ancillary files) and how to access them (Chapter 2), the input raw data and the telescope/camera system that was used or simulated to obtain them (Chapter 3), the key algorithms and the processing flow that were used to produce the data products (Chapter 4), and an assessment of the scientific quality of the output data (Chapter 5).

Source Material & Attribution

The material for this *Handbook* was drawn from a large number of sources, including LSST technical documents, Wiki pages, external web pages, mail exploders, data file headers, software design documents, and informal conversations with experts. Often, figures, tables, and even text are excerpted from these sources. In order to keep the style and presentation relatively clean, the attribution to the source material is cited the first time it is used in the main body of each chapter. The last section of each chapter is devoted to a listing of references, contributing authors, and background resources that provide details that fall outside the scope of this *Handbook*.

Special Notes

Selected fonts are used to indicate special terminology, or to indicate user interaction with tools:

- ***Bold italic*** indicates technical terms of interest the first time they are used in the text, and are defined in the Glossary at the end of this *Handbook*.
- Names of software tools or packages are indicated with mixed-case **bold**.
- `Fixed-width type` indicates text that should be input to a software application or web tool.
- **Underlined Arial bold in teal** indicates text that appears on a “button” in an application.

The planned processing software for this data challenge has not been fully implemented, and will grow as the data challenge proceeds (see Chapter 1). Descriptions of processes or data products that are planned but not implemented in the current data challenge are denoted in tables and figures with a grey background.

Special notes appear throughout this *Handbook* to convey information of special interest or urgency. They are the following:

Informative notes are denoted with the light-bulb symbol, and generally contain tips and pointers that deserve special attention.

Cautionary notes are indicated like this, and indicate potential limitations of the data, the instrument, or the processing software that may affect the use or interpretation of the data products.

Warnings of serious consequences are indicated like this, and denote problems with the data that could lead to erroneous scientific or technical interpretations, or problems with software that could lead to errors that may not be apparent to users.

Chapter 1: Introduction to LSST Data Challenges

The LSST Data Management team has been carrying out a series of ***Data Challenges***, ever more sophisticated realizations of the LSST data and the pipelines and infrastructure needed to analyze them. This chapter describes the purpose and plan for the data challenges, and a rough plan for the functionality to be incorporated in future data challenges. The output data products, including images and catalogs, are described in Chapter 2. The final products are derived from a subset of the Sloan Digital Sky Survey calibrated images, the general properties of which are summarized in Chapter 3. The processing flow and applicable algorithms are described in some detail in Chapter 4.

This introductory chapter also includes a summary of applicable policies for use of these data and for scientific or technical publications that may be produced; it concludes with a description of how to obtain help in accessing or analyzing data from the current data challenge.

1.1. Winter 2013 Data Challenge in a Nutshell

The Data Challenges (Kantor et al. 2010) are being carried out as a series of semi-annual data releases, with the goal of prototyping critical portions of the full Data Release ***Production*** (DRP) system. Each Data Challenge release delivers various types of data and steadily improving levels of scientific quality. The development plan for the next data challenge is shaped on the one hand by the need to prototype critical, high-risk algorithms successfully prior to the Final Design Review, and on the other by having implemented predecessor stages in prior data challenges that are proven to be robust and scientifically viable.

The first release to the Science Collaborations, in January 2011, was largely intended for initial software integration and shakedown. The general goals were to remove instrumental signature from and calibrate single-visit images, perform PSF photometry on detected sources, and generate a catalog of objects from multiple detections of sources. The amount of data processed was relatively small, the processing stages were limited, and data quality goals were not uniformly achieved. The input data for the first release included the first large-scale runs of the LSST image simulation (***ImSim***) framework. These simulations were at an early stage of maturity and had not been validated against the expected performance of the LSST hardware, i.e., they were not intended be used to determine the final capabilities of the LSST.

The second Data Challenge release, in the summer of 2011, featured a much larger volume of ImSim data, a user interface based on Gator, VOInventory, and data quality assessment tools. Single frame measurements reliably met data quality requirements: the astrometric solutions and single-visit photometry were more accurate, as were time series of transient and variable objects. There were some issues with the fidelity of image quality as generated by the ImSim software (many of which were addressed shortly after the Summer 2011 release). The objective of the Winter 2012 production was to update the DM software stack (incorporate the latest stable releases of third-party software, migrate to more recent operating systems, etc.) and to streamline the production processing; no substantial algorithmic changes were introduced at that time. No large-scale data production runs were performed, although the stack was re-validated with smaller-scale runs.

The third data challenge, termed *Summer 2012*, used a small fraction (~5%) of the data from the *Sloan Digital Sky Survey* (***SDSS***)[1] Stripe 82, and had as its goal to exercise some of the latter stages

[1] The SDSS is described at http://www.sdss3.org/.

of the DR Production—i.e., building photometric catalogs from single-visit calibrated images, achieving rough star/galaxy separation, demonstrating high photometric fidelity, and producing deep, *r*-band image stacks. Basic instrument signature removal and image characterization operations were not necessary, as published SDSS images are already well calibrated in these respects. This data challenge also featured a release to the community of the DM processing software, in the hope that it will be of value to software-savvy, early adopters in the LSST Science Collaborations.

The primary goal for Winter 2013 Stripe 82 reprocessing effort was to add the capability to make background matched co-adds, and test it at the largest scale yet by processing the full extent of Stripe 82 data in three passbands. Background matched co-adds preserve the diffuse astrophysical backgrounds in the stacked image. This increases their scientific usability although there are some recognizable image artifacts. Also, the background as characterized in the co-adds was based on much higher S/N data products, thus making it easier to subtract it when needed. We plan to use this method to generate the co-adds in the full LSST DRP, and it was therefore important to have it built into the stack early. This will make upcoming tests of stackfit/multifit algorithms more realistic.

LSST Data Management encourages the Science Collaboration members to explore data products from each data challenge to become familiar with the catalog database and access tools, and to perform quality checks on the data that may uncover problems that have not been documented. With this data release, collaborators may be able to derive new, useful science from these data products. In particular, the current production software does not include detection and orbit determination of solar system objects.

The next data release, in early CY 2013, will add the implementation of PSF matching, difference images, and object deblending. The Winter 2013 final release will continue the development of the full-fledged *Multifit* measurements of faint galaxies from single frames at a variety of epochs. The determination of astrometric models (proper motion and parallax), and asteroid orbits from the moving-object pipeline (MOPS), transient characterization, and a global photometric solution will be addressed in future data releases. Once the production system has met this level of maturity, a large fraction of the data quality requirements will have been met, the astrometric models derived from them should be valuable, as should time series of transient and variable objects. MOPS results will be preliminary, although shape measurements of faint galaxies may initially be relatively primitive.

1.2. Goals and Status for the Current Data Challenge

The plan for most recent round of processing, released in early Winter 2013, included several high-level goals for what data would be processed, and the science quality of the results. These goals and their status are summarized below. Goals that were not achieved for this release will likely be revisited in a later data challenge.

- Perform DRP processing on the SDSS Stripe 82 fpC images (as provided in DR8) from all epochs over 10 yr, which covers roughly 237 deg^2 in three passbands: *g, r,* and *i*. Produce calibrated, background-subtracted single frames (calexp's). **Status: achieved**; 2 million images processed covering $-40° <$ R.A. $< 55°$ and $-1.25° <$ Dec. $< 1.25°$.
- Produce deep co-add images in *r*-band, in the LSST SkyMap geometry, from which faint sources can be robustly detected. **Status: achieved**; 5126 images produced.
- Perform forced-source photometry on all calexp images, at the positions of the sources detected on the co-add images. **Status: achieved**; 1.5×10^7 sources detected, 3.8×10^9 forced-sources detected.
- Generate automated data quality reports that facilitate the assessment of the accuracy of the astrometric solutions and photometric zero-points. **Status: achieved (**re-formatting of presentation and navigation will be necessary prior to public release).

- Demonstrate that the depth and accuracy goals for photometric fidelity have been met:
  - Improved photometry of isolated, bright point-sources: <0.03 mag RMS. **Status: achieved.**
  - Photometry of extended sources: <0.05 mag RMS. **Status: achieved.**
  - Recovery of all sources in a reference catalog to the expected magnitude limit (~24.4 in *r*). **Status: achieved** (except for blended sources).
  - Accurate photometry in principal colors to the plate limit. **Status: achieved for point sources**.
- Produce good quality, model-fit based, extended-source photometry and galaxy shape measurements on single-frame reduced images. **Status: achieved for small galaxies.**
- Provide a stable release of the LSST software stack (the one used in the Winter 2013 DRP) to the community, along with instructions for installing, building and running the software. **Status: achieved.**

See Chapter 5 for a more detailed discussion of data quality. Generally speaking, this data release provides the fundamental source and forced-source catalogs, rough star/galaxy separation, and deep image stacks in the *r*-band. Left for future releases are pan-chromatic detection stacks, difference images, the identification of moving objects, multi-object fitting and measurement (for crowded fields and complicated targets), and global astrometric and photometric calibrations.

1.3. Publication and Data Use Policy

Results from the analysis of data products from this Data Challenge may appear in internal reports, as well as presentations, external technical publications, and (possibly) science papers. Authors of papers and reports that make use of data products produced by LSST software, or the LSST software components themselves, should be aware that publications should include an appropriate acknowledgement of the source of these resources.

Please note: there is a formal policy for publications of LSST Science and Technical material (LSST Board of Directors, 2011). This LSST Publication Policy provides direction on authorship, attribution, and acknowledgements, and requires an internal review of the content by the relevant Science Collaboration, Working Group, or other authority prior to publication.

1.4. Obtaining Help

There are multiple sources of help for accessing or understanding LSST Data Challenge products and services. The primary source of assistance, beyond this *Handbook*, is the Data Challenge User Forum. The intent of the forum is to create a place for LSST users to ask questions of other users. Although the forum will be monitored by LSST staff (who will also answer questions and participate in discussions), the goal is to provide an independent source of support for the LSST community, drawing from the collective experience of its own members. A forum format is well suited for questions and their subsequent discussion and resolution. An e-mail interface to the forum may be provided (for users who prefer to receive an email digest) depending upon demand. In addition, the forum will provide a searchable archive of previously asked questions, which users should consult prior to asking their own question. The ultimate goal is that the forum become a useful and self-sufficient resource for the LSST community, where users may gain insight into common (or rare!) problems and contribute to the growing understanding of LSST, its data characteristics and the data interfaces, while continuing to build a sense of community.

The User Forum can also facilitate the coordination of the scientific analysis and data quality assessment across Science Collaborations, and will include advice on tools, techniques, and avoiding pitfalls. To access the User Forum (which requires a login and password) go to:

https://www.lsstcorp.org/sciencewiki/index.php?title=Special:AWCforum

Users may also find the Science Collaboration Wiki (which requires the same login and password) to be helpful. You can access it via:

http://www.lsstcorp.org/sciencewiki/index.php?title=Main_Page

Finally, for help with understanding details of the processing algorithms, or with problems or technical issues related to data access, or for obtaining a login for the Science Collaboration web site, please e-mail the Data Management Help Desk. The Help Desk will address questions on a best-effort basis, with a goal of resolving issues within a few business days.

E-mail: dm-help@lsst.org

1.5. References and Further Information

Contributing Authors

Contributors to the technical content of this chapter include Mario Juric, Lynne Jones, Jeff Kantor, Dick Shaw, and Michael Strauss.

References

Kantor, J., Axelrod, T., Allsman, R., Freeman, M., Lim, K.-T. 2010, *Data Challenge 3b Overview*, LSST Document 9044 (Tucson: LSST Corp.), available at: http://www.lsstcorp.org/sciencewiki/images/DC3b_Scope.pdf

LSST Board of Directors 2011, *LSST Publication Policy*, LPM-53 (Tucson: LSST Corp.), available at: https://www.lsstcorp.org/docushare/dsweb/Get/LPM-53/lsstpublicationpolicy.pdf

Chapter 2: Accessing the Data Products

The model that has been adopted for user interaction with the current Data Challenge data is to provide users with the ability to search for, select, access, and retrieve data products of interest to them, with users analyzing the data on their personal compute platforms. Other models, such as providing compute resources on the LSST cluster, access to the LSST software stack and (easily) configurable reduction schemes, and the use of user-contributed software for bulk processing of LSST data are planned, beginning with this Data Challenge. This chapter describes the content and structure of the data products that have been produced for this Data Challenge, as well as the process for searching and retrieving them. Software that may be helpful for data retrieval and analysis is summarized at the end of this chapter.

2.1. Output Data Products

The data products that are produced by the production pipelines consist of images and catalogs, the contents and structure for which are described in the following subsection. The structure and other characteristics of the input raw images are described in Chapter 3.

2.1.1. Catalogs

The catalogs that are populated by the pipelines, listed in Table 2-1, are likely to be more extensively used than other kinds of data products, both for science and data quality evaluation. The output catalogs are stored as tables in the ***Science Database***[2]; the reference catalog of objects is stored in a separate database. It is easy for users to query and analyze portions of the catalogs using the **Gator** interface, which is described in Section 2.2.1.

Table 2-1: Science Catalogs

Catalog Type	Description
Exposure[3]	Describes each exposure, including the date/time of exposure start, the filter used, the position and orientation of the FoV on the sky, and other environmental information.
Source	Describes each detected source on each Calibrated Image (see Sect. 2.1.2.), including its location (x,y) on the detector, world coordinates (RA, Decl), brightness, size, and shape. The ForcedSource catalog is a variation on this theme, where locations from sources detected on a deep image Co-Add are fed to the measurement pipeline.
Object	Describes attributes of each astrophysical object, including the world coordinates, brightness in each color with time, etc.
Moving Object[4]	Attributes of moving (solar system) objects, including orbital elements, brightness, albedo, rotation period.
Reference Catalog	Catalogs of objects from a reference source. Includes object sky coordinates, type, size, shape, brightness, and orientation. For SDSS, the CFHT DEEP2 catalog was used.

[2] The Science Database schema for the Summer 2012 Data Release may be browsed at http://lsst-web.ncsa.illinois.edu/schema/index.php?sVer=S12_sdss
[3] Of the multiple tables listed in the Science Database schema browser that contain exposure metadata, the content of **Science_Ccd_Exposure** most closely matches the fields described in the public interface.
[4] The Moving Object catalog is planned for DC3b, but not yet available.

It is worth emphasizing the distinction between the terms ***source*** and ***object***. A ***source*** is a detection of an astrophysical ***object*** in a single image (i.e., an exposure), in a single passband, the characteristics for which are stored in the Source Catalog of the science database. The Data Management System attempts to associate multiple source detections in all passbands to single astronomical ***objects***, such as a star, galaxy, asteroid, or other physical entity, which can be static or change brightness or position with time. Usually an ***object*** will be associated with more than one instance of a ***source*** detection, with the exception of certain classes of transient objects. The fidelity of the objects depends upon the ability to de-blend overlapping sources in crowded fields. Finally, a ***forced source*** derives from an object that is detected on a deep Co-Add image, but that falls below the detection threshold on a single image. The Co-Add images provide a robust list of sources that are measured photometrically on single images (see Chapter 4).

2.1.2. Image Products

The content of the image products that can be produced by the pipelines (see Chapter 4) are listed in Table 2-2. The files are all in FITS (Pence et al. 2010) format, with very similar but not quite identical internal organization. **For the Winter 2013 Early release, only the deep co-addition images are being published.**

Table 2-2: Types of Science Images[5]

Type	Extension Contents	Size	Units	Description
Calibrated Image	1: Science	CCD	Electron	Images are corrected for instrument signature; paired exposures are combined with CR-rejection, and background-subtracted; calibrations are determined for WCS and photometric zero-point.
	2: Mask	CCD	[None]	Bit-encoded data quality mask: see Table 2-5 for definitions.
	3: Variance	CCD	Electron^2	Variance of Science image, which includes shot noise, read noise, contributions from the noise in calibration reference images, and (for ImSim only) co-addition of the paired visit exposures.
Template Image	MEF: 3	Sky Tile	TBD linear	Result of combining multiple Calibrated Exposures per passband, and removing moving objects and transients
Difference Image	MEF: 3	CCD	TBD linear	Difference between a Calibrated Exposure and warped, scaled Template Image
Deep Co-addition	MEF: 3	Sky Tile	TBD linear	Stacked calibrated science images, one per bandpass, with moving objects removed
Deep Detection Co-addition	MEF: 3	Sky Tile	TBD linear	Stacked, pan-chromatic[6] science image, with moving objects removed

The size of the co-add images is a function both of the LSST sky tesselization scheme, and of the organization of the SDSS survey data (see Chapter 3 for details); see Table 2-3 below. Also given are the sizes of key output tables of sources from the Stripe 82 processing.

[5] Rows with a grey background indicate data products that are normally part of the full Data Release Production, but are not provided or are not relevant for this data release.
[6] Pan-chromatic science images will be combined using the algorithm of Szalay, et al. (1999).

Table 2-3: Sizes of Stripe 82 Input and Output Products

Product	Number	Size of Image
fpC (input)	2×10^6	2048 × 1489 pix, 0.396 "/pix, 128 pix overlap in R.A.
Sky Patch	5126	2060 x 1937 pix, 0.396"/pix, 100 pix overlap in R.A.
Sky Tract	2	90° in RA, 0.2° in Dec; up to 410 patches long
Sources	14.7×10^6	N/A
Forced sources	3.87×10^9	N/A

Structure of the FITS Images

The LSST image files differ somewhat in their internal organization, depending on the type of information they contain. They all contain a science array from a single CCD detector, and most products also include pixel-level concomitant data as well: a variance array, and a data quality mask. The basic organization is shown in Figure 2-1 below. The *input* raw images (described in Chapter 3) are organized as simple FITS images—i.e., a header plus science data array in the primary ***Header Data Unit*** (HDU). The *output* images are stored as a primary header plus three ***image extensions***: one each for the science, mask, and variance arrays. In all cases the metadata (i.e., the keyword-value pairs) found in the primary HDU are applicable to all extensions in the file; metadata found in extension headers apply only to that extension.

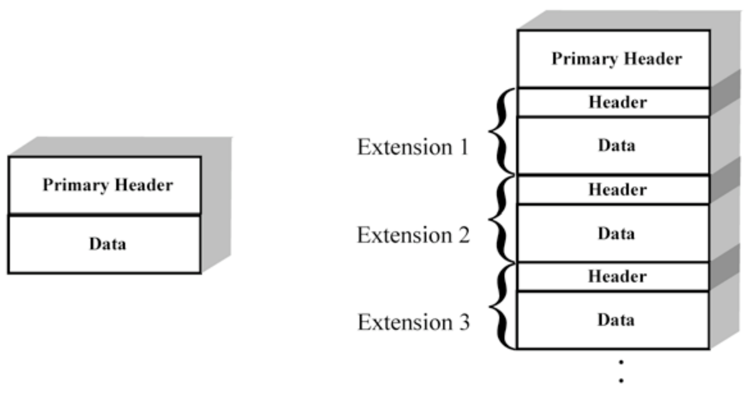


Figure 2-1: Schematic of the structure of a simple FITS file that stores a single image array (*left*) and a Multi-Extension Format (MEF) file that stores multiple components of an image (*right*).

Packaging of the Images

The packaging of image data products for LSST is a compromise among the competing needs of data-parallel processing (which requires small- to medium-sized files), efficient storage (where larger files are optimal), and rapid and reliable transport to users over the internet (for which modest sized, compressed files are optimal). End- users have access to images both individually, and as aggregated into Unix tar files.

The strategy for packaging images is also affected somewhat by the way in which the raw data were obtained in observing environment, and the scheme for tagging observations. Understanding the tree-based organization and file nomenclature is key to understanding how one image relates to another.

The file/path naming convention for the image products is given in Table 2-4. Image files are accessed through the URL: **http://lsst-web.ncsa.illinois.edu/** with the base path for the Winter 2013 SDSS data: **/lsstdata/dr-w2013/deepCoadd/**. The fully qualified name is the concatenation of the URL, the base path, path and filename.

Table 2-4: Image File Naming Convention

Product Type	Path/Filename
Co-add	**/**[gri]**/**<tract>**/**<patch>**/coadd-**[gri]**-**<*tract*>**-**<*patch*>**.fits**
Co-Add tarball	**/**[gri]**/**<tract>**.tar**

In the above Table, words in angle-brackets (italic) are identifiers, characters in square brackets represent a choice (of broad-band filters: *g, r,* or *i*) and characters in boldface black are literal. The *<patch>* is an identifier for the ***patch*** of sky within a ***tract***, with an origin at 0,0 and which can be decomposed into *<raID>* and *<decID>*. There are somewhat more than 400 patches in R.A. at the celestial equator; the partitioning in Dec borrows somewhat from the SDSS survey nomenclature[7]: 12 sectors of 0.2° that define the height of a patch are close in size to that of an SDSS *cam*era *col*umn (***camcol***). See Chapter 3 for details.

Finally, while modest file sizes are more practical for transport and real-time science analysis than an entire stripe, they can be inefficient when analyzing extended objects. For this reason, most data can also be retrieved bundled in a Unix *tar* file, if retrieved via **wget**. Here are examples of fully qualified path/filenames. The first is for a deep co-add image from camcol 4, strip S, in the *r* filter, followed by tar files for all files in a scanline in the *r* filter, and all files in any filter within the scanline:

```
/lsstdata/dr-w2013/r/3/calexp-r-3-0,0.fits
/lsstdata/dr-w2013/r/3.tar
```

Pixel-Level Concomitant Data

Masks

The mask image (extension 2) flags the various pathologies and other attributes of pixels in the science image; their meanings are given in Table 2-5. Each bit has a true (set) or false (unset) state. Flagged conditions correspond to specific bits in a 16-bit integer word. For a single pixel, this allows for up to 15 data quality conditions to be flagged simultaneously (thus far only 7 bits are defined), using a bitwise logical OR operation. Setting none of the bits, or a value of zero in the mask, indicates the pixel is suitable for science use and that no other special conditions apply. (But note that bits 5 and 6, when set, merely indicate the detection of a source, rather than compromised science quality). Note that the data quality flags cannot be interpreted simply as integers but must be converted to base-2 and interpreted as flags. These flags are set and used during the course of processing, and may likewise be interpreted and used by downstream pipeline stages or analysis applications.

Table 2-5: Meanings of Image Data Quality Mask Bits[8]

Decimal Value	Hex Value	Quality Condition Indicated
1	0x**1**	Static bad pixel (e.g., bad column, charge trap)
2	0x**2**	Saturated bad pixel

[7] SDSS survey nomenclature is described in a glossary of terms at: http://www.sdss.org/dr7/glossary/

[8] The assignment of named conditions to particular bits is subject to change, but probably not during DC3b.

4	0x**4**	Pixel flagged for interpolation in the science array
8	0x**8**	Pixel compromised by cosmic ray
16	0x**10**	Pixel in the edge region of a detector array, which is the half-width of the smoothing filter used for source detection, typically ~10 pixels
32	0x**20**	Pixel lies within the ***footprint*** of a detected astrophysical source
64	0x**40**	Pixel lies within the ***footprint*** of a detected source in a Difference Image; in this case the source is dimmer than its counterpart in the Template Image, resulting in a negative brightness profile.

Variance Arrays

The variance array describes the statistical uncertainty of the Science array at the pixel level. This is necessary because the processing for any given pixel involves many factors, including data quality bits that may be set, pixel-level operations with other images that themselves have variance arrays, and the creation of image stacks whose component images likely do not align perfectly.

2.2. Browsing, Queries, and Retrieval

2.2.1. Catalog Data

Public access to the output catalogs is provided through phpMyAdmin at https://lsst-web.ncsa.illinois.edu/mydb/. To access this database you will need an account, which can be requested through a web form at **TBD**. This database is meant to provide a powerful and complete access mechanism to the catalog data, the interface allows users to store and retrieve results from their queries in a variety of formats. This interface is oriented toward users who are conversant with the standard query language, SQL; examples of basic usage are given below.

Support for direct SQL queries of the Science database, and for programmatic access and storage of intermediate results, is not available outside of DM. In the mean time, experienced database users with advanced query needs, meaning those queries that are not supported with the **Gator** tool, should contact the DM Help Desk at dc-help@lsstcorp.org.

2.2.2. Images

Method 1: Web Interface

Public access to the raw and processed images is provided through a web service, the interface for which can be found at https://osiris.ipac.caltech.edu/cgi-bin/LSST/nph-lsst . The interface consists of a web page with direct links to images.

Method 2: Web Service and wget

An alternative method for downloading images involves creating a list of desired files, and using the **wget**[9] software to download them. This software will (by default) preserve the directory structure of

[9] The **wget** software is available at: http://www.gnu.org/software/wget/. The user manual is available at http://www.gnu.org/software/wget/manual/wget.html.

the originating file system. The list can be created by accessing the Science Exposure catalog through the **Gator** interface (see section 2.2.1.), constructing a query that selects the image of interest, and downloading the list of URLs to your local machine. The process is described in detail below.

1. Use the following command to retrieve this list of files:

```
unix% wget –ri myList.txt -nH --content-disposition --cut-dirs=2
```

Using **wget** will also preserve the directory structure through the path (unless explicitly disabled with a command-line argument). Users who wish to download large numbers of images may benefit from consulting with DM staff for alternative methods of mass-replicating data.

2.3. Software Resources

2.3.1. Data Management Software Stack

The version of the software stack used to produce the Summer 2012 data release is available to the community for their scientific use. See http://dev.lsstcorp.org/trac/wiki/Installing/Summer2012 for download and build instructions. The software stack includes a number of third-party software packages, configuration files, and requires specific compilers for the software to build on various platforms. The software is written in a combination of Python v2.7 and C++ (for computationally intensive modules) and is known to run on a number of Unix-based platforms, including RHEL 6 (DM's official development platform), Ubuntu 10.04, and MacOS X 10.7 (Lion). Ports to other Unix systems should be straightforward but are not supported.

While this software has been applied successfully to data from a number of cameras and surveys (The Subaru Hyper Suprime-Cam, SDSS, and LSST ImSim), adapting the software for other data may require significant effort. A far easier task would be for users to re-run SDSS Stripe 82 data, with the processing parameters tuned to suit their own scientific goals. User modifications, customizations, and innovations for their own analyses should be straightforward, but will require effort and moderate programming expertise in the Unix environment. While this software must still be considered prototype code, the geometric and photometric applications in particular are reaching a level of maturity approaching the state of the art, and may be very useful for the broader LSST community.

The Winter 2013 software stack consists of prototype code and is not for the faint of heart. While user feedback is welcome, support for installing and running the stack is minimal.

Recently it has become possible to distribute binaries for several choices of operating system. This mechanism provides a much easier path for using the DM Software Stack, though development in this case is limited to python scripts.

2.3.2. Other Software Tools

The astronomical community is fortunate to have access to a wide variety of software applications, tools, services, and software languages with which to discover, access, and analyze data. There is nothing so special about LSST data products that would preclude the use of most software that astronomers and engineers routinely use for analysis. However, the DM team has experience with some packages that have proved especially useful, which is summarized in Table 2-6 below. All of the listed software is free. We also include some sub-packages or software libraries under the parent package, if applicable, which may need to be installed separately on your computer. This list is by no means exhaustive, and it is mainly focused on data retrieval and analysis, with less emphasis on supporting software development by users. You may find additional suggestions for useful software on the User Forum or on the Science Wiki.

Table 2-6: Applicable Software Packages

Package/ Sub-Package	Version	Description
Aladdin	7.015b	Image display and analysis tool. Available at: http://aladin.u-strasbg.fr/
SAOimage/DS9	7.0.2	Image display and analysis tool. Available at: http://hea-www.harvard.edu/RD/ds9/site/Home.html/
pyds9	1.5	Python interface to XPA to communicate with DS9 .
IRAF	2.16	Widely used astronomical image analysis software. Available at: http://iraf.net/
TABLES	3.14	IRAF package for construction and analysis of tabular data. Available at: http://www.stsci.edu/resources/software_hardware/tables
python	2.6.6	Programming language useful for general scientific analysis. Also used in LSST DM software stack. Available at: http://python.org/
atpy	0.9.6	Python package for manipulating tabular astronomical data in a variety of formats, including FITS, VOTable, ASCII, as well as accessing SQLite, MySQL, and PostgreSQL databases. Available at: http://atpy.github.com/
matplotlib	1.1.1	Currently the most generally capable plotting library for python. Available at: http://matplotlib.sourceforge.net/
NumPy	1.6.2	Numerical operations using arrays. Available at http://pypi.python.org/pypi/numpy/
PyFITS	3.1	Python package for creating and updating FITS images and tables. Available at: http://www.stsci.edu/resources/software_hardware/pyfits
TopCat	3.9	Tabular data display, editor, and analysis application. Available at: http://www.star.bris.ac.uk/~mbt/topcat/
wget	1.13	GNU package for retrieving files using HTTP, HTTPS, and FTP. Available at: http://www.gnu.org/software/wget/

2.4. References and Further Information

Contributing Authors

Dick Shaw, Serge Monkewitz, Mike Freemon, Mario Juric, and K-T Lim contributed to the content of this chapter.

References

Annis, J., et al. 2012, *The SDSS CoAdd: 275 Deg² of Deep SDSS Imaging on Stripe 82*, ApJ, submitted. Available at: http://arxiv.org/abs/1111.6619

Pence, W. D., Chiappetti, L., Page, C. G., Shaw, R. A., & Stobie, E. 2010, *Definition of the Flexible Image Transport System (FITS), Version 3.0*, A&A, 524, A42; also available at: http://fits.gsfc.nasa.gov/fits_standard.html

Szalay, A. S., Connolly, A. J., & Szokoly, G. P. 1999, AJ, 117, 68

Chapter 3: Input Data

The LSST data processing software for the current data challenge was applied to data obtained from the Sloan Digital Sky Survey (SDSS). The SDSS, a major photometric survey conducted from 2003—2008, is one of the best known and most widely used/cited surveys in astronomy. A complete description of the latest imaging data release (DR8, see Aihara, et al. 2011) data products may be found at http://www.sdss3.org/dr8/. This chapter describes some basic attributes of the input data, how these data were selected, and how they map to the output data products.

3.1. SDSS Stripe 82 Data

The main objectives for the current Data Challenge were to perfect software that could construct deep, per-band co-additions of images and to identify sources on them, and to perform accurate photometric measurements on point- and extended-sources. Data from the SDSS Stripe 82 were in most respects ideal for this purpose. This portion of the SDSS is in many ways similar to the planned LSST survey in that it includes images from many epochs over a substantial span of time, with a nearly identical filter set, and it contains all the foibles of a real data including a range of seeing, sky background and atmospheric transparency.

3.1.1. SDSS Data Selection

All single-frame calibrated data from the SDSS Stripe 82 were targeted for processing during the current DRP, including most of the data from SDSS-II that were obtained to measure light curves for Type Ia supernovae (Frieman, et al. 2008). The ***Stripe*** area of interest for the current data challenge is centered on the Celestial Equator, spanning 2.5° in Declination between -50° ≤ R.A. ≤ +60°, which was imaged in five photometric bands (see Figure 3-1) over many epochs spanning 10 yr. Although a large number of images in the Stripe 82 area were eligible for processing, some ultimately proved not to be useful for quality reasons such as high grey extinction from clouds or poor seeing; processing was ultimately limited to a slightly smaller span in R.A. and the 3 most sensitive bands.

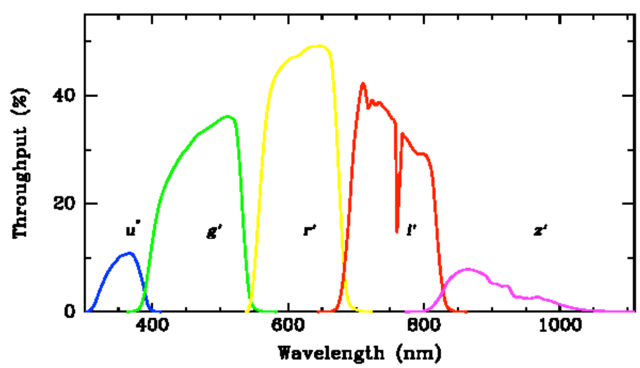


Figure 3-1: System throughput in each filter of the SDSS (*colored curves*), including the response of the telescope + camera, multiplied by the transmission through 1.3 atmospheres.

Attributes of the survey that were processed for this DRP are summarized in Table 3-1 below. See Annis et al. (2011) and references therein, as well as the SDSS DR8 Web site[10], for details of the observations.

[10] See http://www.sdss3.org/dr8/

Table 3-1: SDSS Stripe 82 Properties

Spatial Extent	$-40° \leq \alpha \leq 55°$
	$-1.25° \leq \delta \leq +1.25°$
Area imaged	237 deg^2
Astrometric Precision	<0.1 arcsec for stars with $r < 20.5$
Passbands	*g, r, i*
Single-frame depth (95% completeness)	*22.2, 22.2, 21.3*
Runs/field	Typically >50

Stripe 82 overlaps the DEEP2 survey (Davis, et al. 2007), where the *r*-band photometric depth is a little deeper than 25. This is fainter than the Annis et al. (2012) catalog, and deeper than is achieved in the Summer 2012 Data Challenge processing. Thus, the DEEP2 photometric catalog is extremely useful for evaluating the photometric catalogs produced in the current data challenge (see Chapter 5).

3.1.2. SDSS Camera and Survey Geometry

The SDSS was carried out with the wide field optical imaging camera, the properties of which were summarized by Stoughton et al. (2002). The camera features a focal plane array with 30 photometric CCDs that cover a roughly 2.5° wide swath in the scanning direction. The geometry of the focal plane array is shown in Figure 3-2, where the filters are arranged in 6 columns, with each row assigned to a common filter.

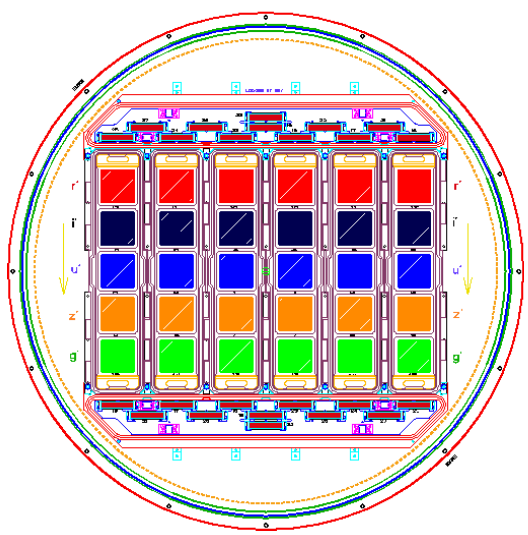


Figure 3-2: Geometry of the focal plane array for SDSS from the DR8 website. The 30 CCDs are arranged in 6 columns, each of which images sequentially through the 5 filters during a drift scan.

The FPA geometry is closely tied to the survey strategy, which in turn strongly influences the organization of the output SDSS image (fpC) files and, to a large extent, that of the co-add images in this data release. The following short summary of the SDSS survey geometry will aid the description to follow of the mapping to the terminology of this data release. See the SDSS-III Glossary[11] of terminology for details.

The SDSS observations were conducted in drift-scan mode, such that the sky was observed for ~54 s in each filter in sequence during a continuous scan, or ***run***. Data from each run consists of six continuous, non-overlapping, drift scans along camera columns, or ***camcols***, which are known collectively as ***strips***. Each strip is composed of six ***scanlines***, which are defined in the great-circle coordinates of the survey. Two interleaved strips, where the telescope is offset by half a CCD width N or S, are necessary to fill in the spatial gaps between the columns. Many such runs at various epochs constitute the input imaging data for ***Stripe*** 82. The data are partitioned along the scan direction into individual, overlapping images for processing purposes: an individual image in a single passband is a ***frame***; the set of frames from all passbands of the same piece of sky is a ***field***.

For co-add generation during DRP processing, the input frames from SDSS are re-projected to the LSST all-sky tesselization scheme called ***SkyMap*** which is partitioned into equal ***tracts*** (four of them, in this case) with overlapping boundaries along the celestial equator, starting at RA=0. Each tract is partitioned into overlapping rectangular images called ***patches***, which for SDSS data are bounded in Declination by scanlines, and sized in Right Ascension to be manageable for processing. Patches overlap one another by a small amount to assure that source near the boundaries are cataloged. Table 3-2 contains a summary of the geometry terminology for SDSS and this Data Challenge.

Table 3-2: Mapping of SDSS to LSST Terminology

Term	SDSS	Summer 2012 Co-Add
Survey area	Stripe (82)	Tract
Image partition	Frame	Patch
Image N-S boundary	Scanline	Scanline
Image sampling	0.396 arcsec/pixel	0.396 arcsec/pixel
Overlap	128 pix	100 pix

3.2. References and Further Information

Contributing Authors

Most of the information about the SDSS data was taken or derived from the project web site, and from the published papers that describe the data releases. Andy Becker, Simon Krughoff, K.-T. Lim, Russel Owen, and Dick Shaw contributed to the content of this chapter.

References

Abazajian, K. N., et al. 2009, *The Seventh Data Release of the Sloan Digital Sky Survey*, ApJS, 182, 543

Aihara, H., et al. 2011, *The Eighth Data Release of the Sloan Digital Sky Survey*, ApJS, 193, 29

Annis, J., et al. 2011, *The SDSS CoAdd: 275 Deg^2 of Deep SDSS Imaging on Stripe 82*, ApJ, submitted. Available at: http://arxiv.org/abs/1111.6619

Davis, M. et al. 2007, *The All-Wavelength Extended Groth Strip International Survey (AEGIS) Data Sets*, ApJ, 660, L1

[11] See the SDSS Glossary at: http://www.sdss3.org/dr8/glossary.php.

Frieman, J. A., et al. 2008, *Sloan Digital Sky Survey-II Supernova Survey: Technical Summary*, AJ, 135, 338

Ivezic, Z., et al. 2007, *Sloan Digital Sky Survey Standard Star Catalog for Stripe 82: The Dawn of Industrial 1% Optical Photometry*, AJ, 134, 973

For Further Reading

Additional details of the LSST system design, expected performance, the planned operations model, and the expectations for scientific discovery may be found in the LSST Science Book, which is available at http://adsabs.harvard.edu/abs/2009arXiv0912.0201L. The science requirements for the LSST hardware and survey are described by formal LSST documents; a detailed discussion of these requirements and their realization in system requirements is presented in:

Ivezic, Z., et al. 2008, *LSST: From Science Drivers to Reference Design and Anticipated Data Products* (astro-ph/0805.2366), available at: http://lsst.org/files/docs/overview_v1.0.pdf

An excellent, early summary of the SDSS data products is presented in:

Stoughton, C., et al. 2002, *Sloan Digital Sky Survey: Early Data Release*, AJ, 123 1485

The survey telescope design was described by:

Gunn, J.E., et al. 2006, AJ, 131, 2332

Chapter 4: Data Processing and Calibration

Science data processing for LSST is organized into a series of ***productions***, which are episodes of processing that are organized to achieve a particular purpose, such as issuing event alerts during a night's observing; generating calibrated images and object catalogs on an annual basis; or constructing calibration reference products. The types of productions that have been identified so far are given in Table 4.1, and in general they operate over a particular timescale of relevance. The focus of the current data challenge, and of this *Handbook*, is the *data release production* (i.e., the annual effort). Each production is a fairly involved activity that takes multiple data inputs and executes software on a massively parallel computing platform in order to generate a variety of science data products, related metadata, and data quality information. The software is organized into a series of ***pipelines***, or independently executable codes, each of which consists of one or more logical ***stages*** that perform discrete algorithmic operations.

The computations are optimized for high throughput on highly parallel computing platforms, but algorithmic software is organized to be largely independent of the execution environment. This chapter will focus on *what* scientific operations are performed in the Data Release Production for this Data Challenge, with relatively little discussion about *how* they are performed except to the extent that they affect the organization of the output data.

4.1. Pipeline Processing

4.1.1. Overview

The flow of the science data through the pipeline processing is shown in Figure 4-1 for this Data Release Production. Each step of the processing, indicated by the boxes in the center of the figure, is described in detail in the following subsections. As explained below, most steps related to instrument signature removal and image characterization are not performed for SDSS data since the input images are in most respects well calibrated. Inputs to the processing include the individual science frames and concomitant data, configuration files, and catalogs. Outputs include the deep co-added science images, catalogs, and data quality metadata. Most intermediate products that are produced during the course of pipeline processing, but that are not archived, are not discussed here.

The first phase of the processing, illustrated in Figure 4-1, transforms SDSS science images and concomitant data to the format of standard LSST calibrated images, performs approximate background characterization, identifies sources on all input images from all observation epochs, and performs photometry on the sources. This phase concludes by matching the detected sources against the reference catalog, which is the Stripe 82 Co-Add catalog of Annis, et al. (2011).

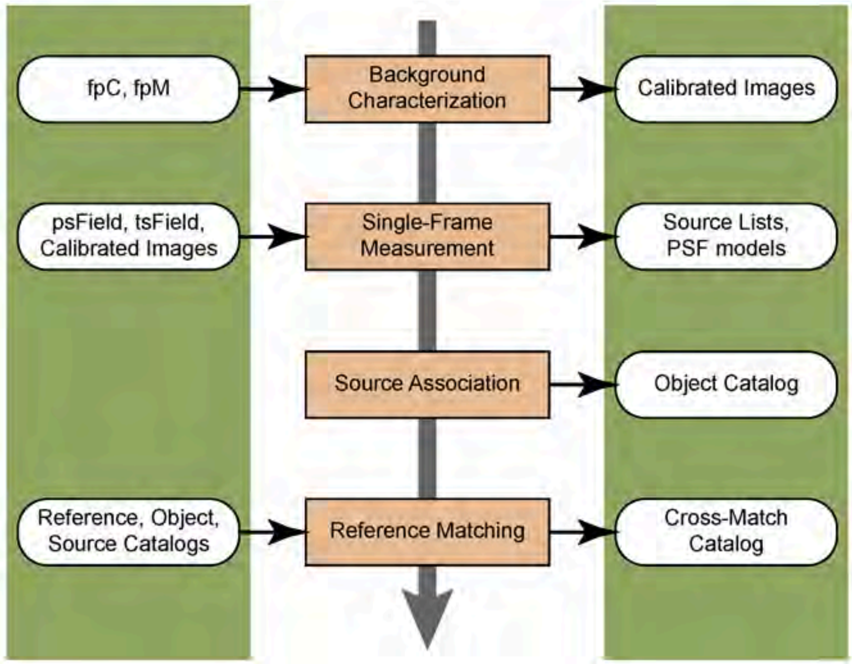


Figure 4-1: Flow of data taken with a common filter through single-frame processing of the Data Release Production pipeline to produce the Object catalog. Images and catalogs are shown as inputs to (*left column*) or outputs of (*right column*) the processing. See Figure 4-2 for the next processing stage.

The next phase of the processing, shown in Figure 4-2 below, constructs co-added images for one or more bands. Configuration files allow quality thresholds to be imposed to select the input images. Pixel values are rejected if they are flagged in the input mask, or are outliers based on a configurable rejection threshold. The image background is estimated from a reference image for patch of sky (one with low background) and the background for the other contributing images is matched to the reference. Once created, sources are located on the co-added images and saved for further processing.

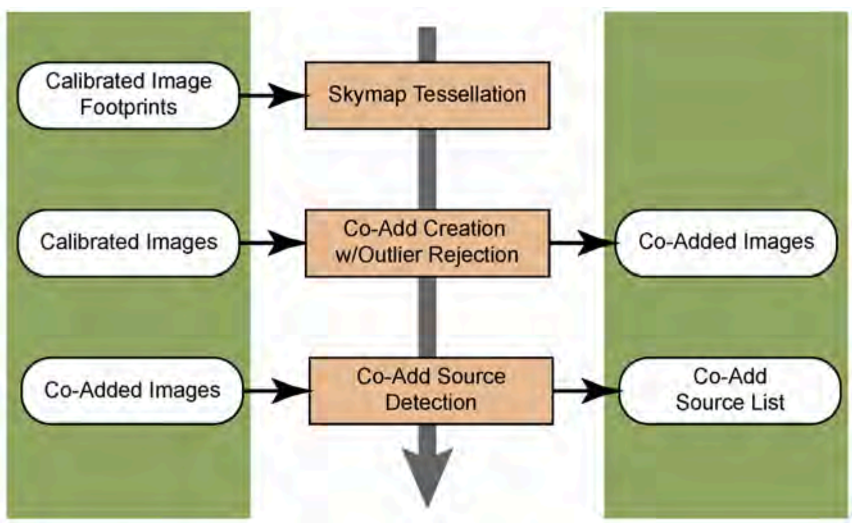


Figure 4-2: Similar to Figure 4-1 for the production of the per-band co-added images and the Co-Add Object catalog. Note that the co-add images have been re-projected and re-partitioned for the SkyMap reference frame. See Figure 4-3 for the final processing stage.

The final phase of the processing, shown in Figure 4-3 below performs photometry on all input images of the sources in the Co-Add Source list, which includes sources fainter than a 3-sigma rejection threshold on a single exposure would allow. The photometric measurements on these ***forced sources*** are matched to the input reference catalog. Data quality outputs are then generated for the production.

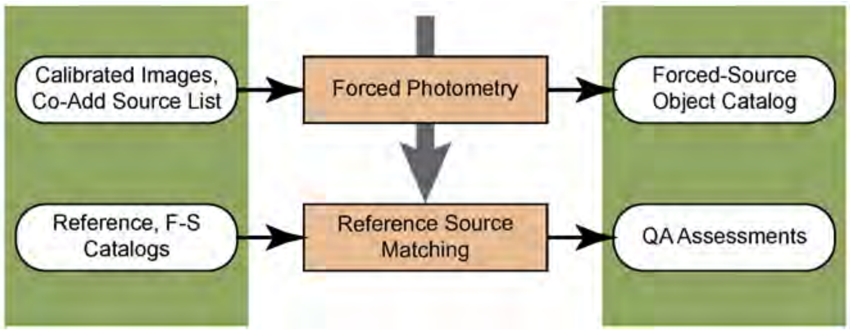


Figure 4-3: Similar to Figure 4-1 for the production of the Forced-Source (F-S) Object catalog. F-S objects are matched against the reference catalog to assess the catalog completeness and the fidelity of the photometry.

4.1.2. Single-Image Processing

Having described the overall data flow, this section describes the main processing steps in more detail. The SDSS image products are themselves the products of a survey-quality data reduction pipeline. Basic calibrations (instrument signature removal) and characterizations (bad pixel and cosmic-ray masking, WCS solution, PSF modeling) have already been performed; re-doing those steps with the LSST pipelines would not add significant value. Therefore, starting with SDSS fpC files, the algorithms described below are applied to generate the output products.

Background Subtraction

The background in science images is removed. The background level includes multiple sources, including twilight, airglow, scattered light (from the moon), ghosts and glints from the optical surfaces of the telescope and instrument optics, and the extended wings of very bright stars. The low-spatial frequency scattered light is determined in multiple, non-overlapping regions across each input field. Within each region a statistic is computed to estimate the local background, and the resulting values are interpolated over the extent of the field. The statistic, the interpolation scheme, and the region sizes are all configurable, and are currently: clipped mean, a natural spline, and 512 × 512 pixels, respectively. After the single-frame calibrations are determined, the background is added back to the image for use in downstream processing (including a superior background estimate that is generated during co-addition: see 4.1.3.).

Image Characterization

The background-subtracted images are now ready for single-frame calibrations and basic measurements. Note that all calibrations and measurements at this stage of the processing are performed at the ***field*** level. Therefore measurements of sources that fall on or near detector boundaries will be affected.

While a World Coordinate System (WCS) calibration is performed on single frames, it is not used for SDSS images in this data release. The WCS characterization from the Sloan Survey is very good, and also more than sufficient to satisfy the goals of the current production.

Likewise, the SDSS processing includes a high-quality characterization of the PSF within each field, which is used for downstream processing in the DM pipeline.

Photometric Calibration An estimate is made of the magnitude zero-point of each CCD in each visit image by comparing the published magnitudes in the reference photometric catalog to their instrumental magnitudes, applying color transformations as necessary. Currently, the reference photometric catalog is that of Annis, et al. (2011). Note that the result of the photometric calibration is to populate the exposure table in the science database with the photometric zero-point for each field. The pixel values remain unchanged, and have units of count s^{-1}.

Source Detection

Following the production of a calibrated, background-subtracted image, it is examined for ***sources***, or astrophysical targets in the field of view. A copy of the image is first convolved with a circularly symmetric Gaussian brightness profile that has the same width as the PSF for that image. Pixels above a configurable threshold in this smoothed image are flagged, and groups of contiguous pixels are measured to determine the centroid locations, fluxes, and shape parameters of (possibly overlapping) sources. Note that this step will ignore sources within a small margin (15 pix) near the edge of the field; this effect is ameliorated by the intentional overlap of pixels along a camcol. The source shapes are derived from an algorithm that is similar to that used for the SDSS. The resulting measurements (position, brightness, shape, orientation, and errors on those parameters) for the list of all detected sources are recorded in the ***source catalog*** of the science database. Various conditions that may compromise the quality of the source measurements are encoded in the data quality field; their meanings are summarized in Table 4-1 below. The working definition of pathologies that would render a source scientifically useless for downstream analysis is that one or more of the following bits are set: (0x1, 0x200, 0x800)—i.e., the source includes edge pixels, or that the source center is close to interpolated or saturated pixels.

Table 4-1: Meanings of Source Data Quality Flags

Decimal Value	Hex Value	Text Code	Quality Condition Indicated
1	0x**1**	`EDGE`	Source includes pixels within the edge region of a detector—i.e., the half-width of the smoothing filter used for detection, which is typically ~10 pixels.
2	0x**2**	`SHAPE_SHIFT`	While estimating the best-fit Gaussian filter, the derived centroid varied significantly from the initial guess
4	0x**4**	`SHAPE_MAXITER`	The adaptive moments solution required more than the maximum allowed iterations.
8	0x**8**	`SHAPE_UNWEIGHTED`	The adaptive scheme failed to converge, so the moments are unweighted and therefore noisy and unreliable.
16	0x**10**	`SHAPE_UNWEIGHTED_PSF`	The PSF's "adaptive" moments are unweighted. This flag is currently not used.

Decimal Value	Hex Value	Text Code	Quality Condition Indicated
32	0x**20**	`SHAPE_UNWEIGHTED_BAD`	The source is so noisy that no shape could be determined. The `SHAPE_UNWEIGHTED` flag will also be set.
64	0x**40**	`PEAKCENTER`	Centroid determination failed: derived center is set to peak pixel.
128	0x**80**	`BINNED1`	Source was found in 1×1 binned image. (Larger binning factors will eventually be used to detect extended, low surface brightness sources.)
256	0x**100**	`INTERP`	Source's footprint includes interpolated pixels.
512	0x**200**	`INTERP_CENTER`	Source's centre is close to interpolated pixels.
1024	0x**400**	`SATUR`	Source's footprint includes saturated pixels.
2048	0x**800**	`SATUR_CENTER`	Source's center is close to saturated pixels.
4096	0x**1000**	`DETECT_NEGATIVE`	Source was detected as having negative flux (in a difference image), at a significance of at least 5-sigma.
8192	0x**2000**	`STAR`	Source size and shape is consistent with being point-like.

While saturated sources, and sources found near the edge of the detectors, are included in the source catalog, large sources that fall near the N-S edge of a camcol may not be properly counted in completeness statistics.

Note that source detection step is performed twice: first on a per-field basis, and subsequently on the co-added images.

4.1.3. Image Co-Add Generation

Background Matching

Exposures taken at different times exhibit temporal and spatial variation in background levels. A goal of photometric pipelines is to remove the non-astrophysical background (e.g. airglow, scattered moonlight), while preserving the astrophysical background (e.g. unresolved sources, wings of galaxies). In practice the two are difficult to separate. For example, a diffuse astrophysical structure, such as a galaxy or nebulosity, can span an area wider than a single chip and be compromised by temporal effects like cloud cover. The pipeline uses a relatively robust algorithm for minimizing the non-astrophysical background in generating co-added images.

For each ***skymap patch***:

1. Select all exposures that overlap the patch (with possible rejection based on PSF quality or atmospheric conditions).
2. Choose a reference exposure based on maximal overlap with the patch, that has low background and that was obtained in very good atmospheric conditions.
3. Warp each exposure to the patch, and calibrate it to a common photometric zero-point. (Matching to a common PSF would be necessary for accurate photometry of the co-add, but this is not yet implemented.)
4. Match the background of each input exposure to that of the reference image by fitting a spatial model to the difference of the input and reference exposures, with bad pixels masked, and subtracting it from the input image. In the case of SDSS, the spatial model is a low-order, two-dimensional spline.

5. Co-add the background-matched images.

The result yields co-adds with high signal/noise ratio (S/N), with a minimal background consisting of both diffuse sources and the temporal background of the reference exposure. The final step is to estimate the background on this high-S/N, wide-area mosaic.

4.1.4. Source Photometry

Once sources have been identified, the flux is measured using multiple techniques, including aperture photometry and adaptive Gaussian moments. For stars (or any angularly compact source with an approximately stellar profile), the photometric calibration step described above assures that aperture and PSF fluxes will agree to good accuracy. For extended sources such as galaxies, the story is more complicated. A robust model fitting code (*Multi-fit*) is being developed for galaxy photometry, which is being designed to fit multiple components at once to complicated sources. Not all of the functionality has been implemented, but for the current release such sources are fit with a linear combination of multiple components: a delta function, an exponential profile, a de Vaucouleur profile, and a second-order shapelet basis. All components are convolved with the PSF model, and the flux is computed as the integral of the model. Note that the radius and ellipticity of the components are not currently fit: these are fixed to a small number of test points by applying a naïve (Gaussian) correction for the PSF to the adaptive moments of the source. However, the inclusion of the shapelet basis effectively allows for small perturbations in ellipticity and radius from the fiducial values.

There is currently no deblending of overlapping sources in the Source and Object catalogs, and no identification of moving objects. The star-galaxy separation with the current software stack is known to be problematic for angularly small galaxies. Photometric accuracy is likely to be poor for sources that are comparable to or larger than the size of the image section used to determine the local sky background.

Source photometry is repeated on individual input images after fainter sources have been detected on the co-added images. The quality of the resulting photometry is described in Chapter 5.

Source photometry is only performed on the individual input images, **not** on the co-add images. This avoids a number of problems, including imperfections in the creation of science-grade co-adds and the resulting correlation of errors in adjacent pixels. It is also most similar to the technique that will eventually be used in production LSST processing, using the *MultiFit* algorithm.

4.1.5. Catalog-level Processing

Source Association

Once the source catalog has been generated from all processed images, the source association pipeline identifies the (large) subset that corresponds to multiple detections of individual astrophysical targets.

Source association is carried out with the *OPTICS* algorithm (Ankerst, et al. 1999), although the current implementation is equivalent to the *DBScan* algorithm (Ester, et al. 1996). It is one of the most common clustering algorithms used in the science literature[12] and it is very efficient, with a runtime complexity of $O(n \cdot \log n)$. The idea is to examine each source in sequence and form ***clusters***, or candidate sources. Clusters can initially be individual sources, but clusters grow when the spatial separation between candidate members and the cluster is small enough. The algorithm is parameterized on the characteristic spatial separation (ε neighborhood) and the minimum number of points (MinPts) required to form a genuine cluster. These parameters are tuned to a given dataset so that the number of false associations is minimized. Preliminary experiments with the sources extracted from images lead to setting ε=0.5 arcsec, and MinPts=5.

The algorithm operates on the set of all sources falling into the image. These sources are taken from all the fields in all passbands within some padding distance P of the image; P is chosen based on an estimate of the maximum error in the raw WCSes, and is currently ~15 arcsec. The algorithm visits each source S in the sky-tile (in an arbitrary order). If the ε neighborhood of a source S contains at least MinPts other sources, and S has not already been placed into a cluster:

1. Create a new cluster *C*.
2. Add all the ε-neighbors of *S* that do not already belong to a cluster *C*.
3. Recursively perform step 2 for each ε-neighbor *S′* of *S* that has an ε-neighborhood containing at least MinPts other sources.

If the ε neighborhood of S contains less than MinPts other sources, it is called a *noise source* and is discarded. All clusters are stored in the ***object catalog***.

Objects composed of only one source (i.e., one detection in any passband) are in principle allowed, but have been disabled for the present in order to avoid corrupting the object catalog with garbage sources. When the required tuning of this algorithm is better understood, this restriction will be removed.

Source association is performed twice: on the individual image source detections, and on the detections in the co-add images. In both cases the results are cross-matched against the reference catalog for QA purposes.

4.2. References and Further Information

Contributing Authors

Yusra Al Sayyad, Steve Bickerton, Andy Connolly, Simon Krughoff, Dustin Lang, K.-T. Lim, Robert Lupton, Serge Monkewitz, Paul Price, and Dick Shaw contributed to the content of this chapter.

References

Annis, J., et al. 2011, *The SDSS CoAdd: 275 Deg^2 of Deep SDSS Imaging on Stripe 82*, ApJ, submitted. Available at: http://arxiv.org/abs/1111.6619

[12] See the Wikipedia entry for DBSCAN at: http://en.wikipedia.org/wiki/DBSCAN.

Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Jörg Sander, J. 1999, *OPTICS: Ordering Points To Identify the Clustering Structure*, in ACM SIGMOD International Conference on Management of Data, ACM Press, 49

Bernstein, G. M., & Jarvis, M. 2002, *Shapes and Shears, Stars and Smears: Optimal Measurements for Weak Lensing*, AJ, 123, 583

Bickerton, S. J. & Lupton, R. H. 2012, *An Algorithm for Precise Aperture Photometry of Critically Sampled Images*, MNRAS, submitted

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, *A density-based algorithm for discovering clusters in large spatial databases with noise*, in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, eds. E. Simoudis, J. Han, U. M. Fayyad, AAAI Press, 226

Lang, D., Hogg, D. W.; Mierle, K., Blanton, M., & Roweis, S. 2010, *Astrometry.net: Blind Astrometric Calibration of Arbitrary Astronomical Images*, AJ, 137, 1782

Shupe, D. L., & Hook, R. N. 2008, *The SIP Convention for Representing Distortion in FITS Image Headers*, available at: http://fits.gsfc.nasa.gov/registry/sip.html

For Further Reading

A detailed (and evolving) set of on-line notes on the pipeline stages, and their inputs and outputs, maybe found on the Data Management wiki at http://dev.lsstcorp.org/trac/wiki/DC3bProcessingStages.

Chapter 5: Data Quality Assessment

Data as processed by the DM production software are evaluated for scientific quality by computing various quantities of interest and comparing them, when possible, to established science quality metrics. A detailed set of science requirements for data quality is under development for DM data processing, many of which flow directly from high-level science requirements for LSST as a whole.

Quality is assessed for SDSS data in part by comparing measurements produced by the pipelines to one or more reference (i.e., “truth”) catalogs. This chapter describes how to access the automated quality assessment reports, and summarizes the initial assessment of the scientific quality of the generated data products. This chapter will also point out both known problems and specific questions that have yet to be addressed. It is anticipated that members of the Science Collaboration teams will contribute their own analyses to the assessment. This *Handbook* will be updated periodically as the understanding of the released data products matures. It is expected that the analyses themselves will be folded into a definitive Data Challenge report, to be published separately. This chapter concludes with examples of the types of feedback that the Data Management Team would find most useful.

The automated data quality assessment for SDSS data products presents an excellent opportunity for the Science Collaboration members to analyze the science data quality in even greater depth, and to explore new science questions with the catalogs.

5.1. Assessment of Processed Data

5.1.1. Pipeline Processing Diagnostics

The pipelines (or stages thereof) have the capability to report problems that may occur during processing. Issues not related to algorithmic flaws are generally resolved prior to archiving the data for release. Problematic data (which could result from poor observing conditions) are simply flagged. The flags that have been created to date are listed in Table 5-1, a list that will undoubtedly grow when the software recognizes more conditions. These flags reflect quality assessments at the level of a single patch: see the **`Science_Ccd_Exposure`** table in the database. It may be appropriate to exclude these effects from analysis, depending upon the objective.

Table 5-1: CCD Processing Diagnostics

Decimal Value	Hex Value	Text Code	Quality Condition Indicated
1	0x**1**	`PROCESSING_FAILED`	The pipeline failed to process this CCD.
2	0x**2**	`BAD_PSF_ZEROPOINT`	The PSF flux zero-point appears to be bad.
4	0x**4**	`BAD_PSF_SCATTER`	The PSF flux for stars shows scatter >0.03 mag.

5.1.2. Assessment of Forced Photometry

The algorithms and processes used in the processing pipelines have been described in Chapter 4 in enough detail to convey a basic understanding of how the input data have been reduced and calibrated. The production pipelines perform photometric calibration using the catalog of Ivezic (2007). It is possible to assess the photometric quality for this Data Challenge by examining the distribution of forced photometry *principal colors* of stellar sources, as defined in Ivezic (2004). Star-galaxy discrimination is derived from the Annis (2011) Stripe82 catalog to select point sources for the analysis; the current pipelines do not do any native star-galaxy separation. Figure XXX below illustrates the process of defining the principal color **w** (adopted from Ivezic (2004).

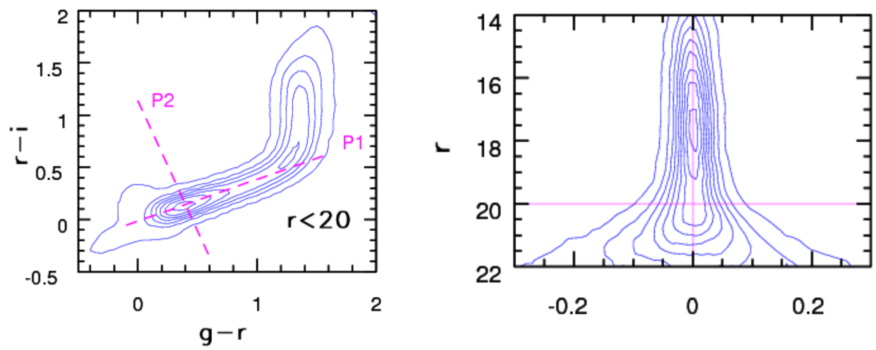


Figure 5-1: *Left*: The width of the principal color P1 is a function of the underlying stellar astrophysics of the sample, and errors in the photometry. *Right*: The width of the distribution (P2) increases as a function of magnitude as photometric uncertainties start to dominate.

In the analysis shown here, the width of the stellar locus distributions for principal colors **w** (shown above) and **x** (defined in the *r-i* vs. *g-r* diagram) were examined as a function of position (-10° < R.A. < 0°) in the Stripe[13]. The distribution of points around the principal colors w and x (the principal locus is at x=0) is plotted in Figure XXX as a function of magnitude. Separate panels show one epoch (i.e., all the data), and the median of all objects with 9 or greater measurements, showing the much tighter locus when photometric measurements are combined for non-variable sources.

[13] The full analysis for all ranges of R.A. may be found at http://dev.lsstcorp.org/trac/wiki/DC/Winter2013.

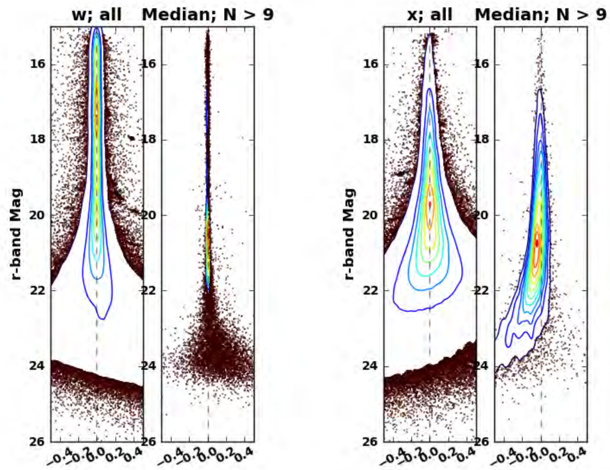


Figure 5-2: The width of the principal colors **w** (*left*) and **x** (*right*) as a function of magnitude. The distributions are shown both for single source measurements (*left panels*) and the median of all epochs with 9 or greater measurements (*right panels*). The locus for the median is much tighter than for single measurements, but increases at faint magnitudes as the photometric uncertainties start to dominate.

The width of the principal stellar locus (Figure XXX) steadily improves as more epochs)1, 2, 10, and 40) are used in the median. The plot also includes a histogram of the number of objects vs. *r*-band magnitude to show the improvement in the catalog depth with the large number of available epochs.

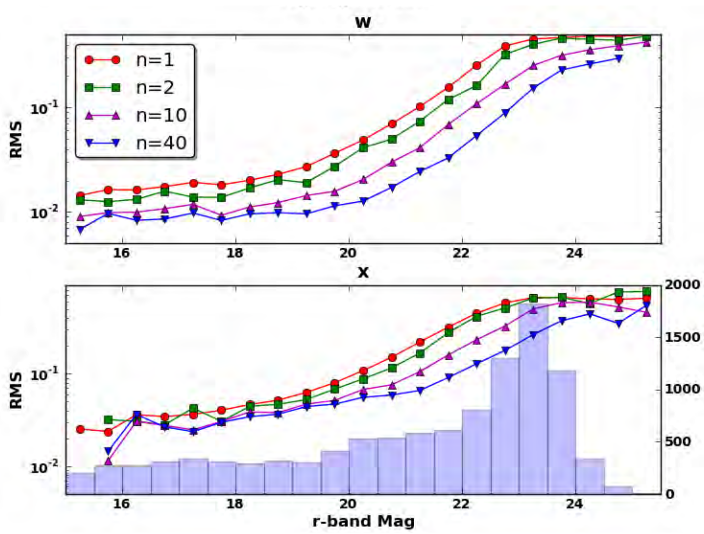


Figure 5-3: Comparison of detected sources from overlapping catalogs: Annis (2011) (*red*), DEEP2 (Davis et al. 2003) (*black open bars*), and the object catalog from this survey using a model PSF (*blue*).

The photometric depth and completeness of the forced-photometry on single frames is illustrated in Figure XXX. The loci show that the realized depth is considerably fainter than the 95% detection limits of the main SDSS survey, and reaches a slightly fainter limit than the Annis (2011) catalog. The technique of detecting faint sources on image co-adds, followed by accurate photometric measurements of these detections on single images, is now well established for LSST processing.

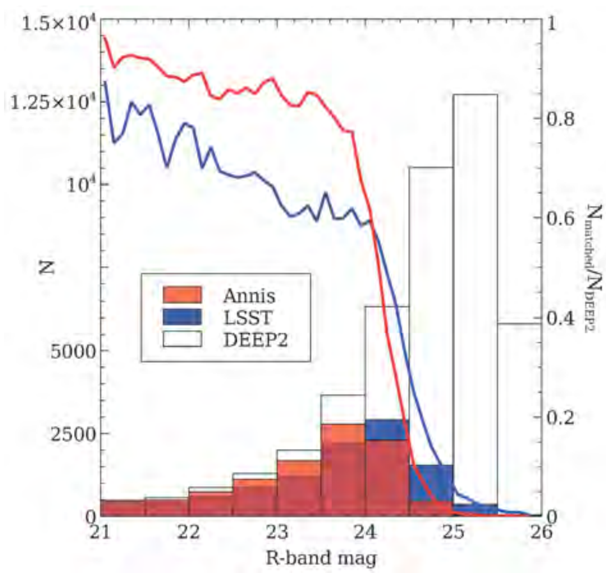


Figure 5-4: Comparison of detected sources from overlapping catalogs: Annis (2011) (*red*), DEEP2 (Davis et al. 2003) (*black open bars*), and the object catalog from this survey using a model PSF (*blue*).

5.1.3. Automated Quality Reports

A quality assessment component of the DM processing system, known as *pipeQA*, was build to help validate the processing software. The automatically generated assessments for the current data release are provided through a web interface, which currently generates views for patches. The results, or data quality artifacts, consist of summary statistics, plots, and reports of specific tests against quality thresholds. The information is quite voluminous, and will only be briefly summarized here; DM team members use this tool to explore processing foibles in detail. The artifacts were generated for the Co-Add images and for only a small portion of the individual frames, owing to the required processing time and data volume (millions of input images in 3 colors). The results are not being made public at this time, mostly because we found the data volume and navigational complexity to be

The *pipeQA* results are not being made generally available at this time, mostly because we found that the data volume and navigational complexity make it hard for a non-expert to get a good, representative view the quality of the processing. The interface for this information will be re-factored to better serve the large datasets that will be processed in future Data Challenges.

Users who have a strong need to review the QA results should contact the Help Desk for access.

The following types of assessments were generated:

- Astrometric accuracy
- Completeness of object recovery
- Photometric fidelity (assessed with 6 comparisons among 4 measurements)
- PSF shape

Each assessment page provides summary plots and statistics; representative plots are shown here to acquaint end-users with some of the means used to assess the science data quality.

Astrometric Accuracy

The accuracy of the world coordinate system (***WCS***) solution can be inferred by comparing the WCS coordinates of bright stars with those of a reference astrometric catalog, which is shown in the summary plots in Figure 5-5. The left figure shows the offset between the measured centroids of matched objects and the catalog position of these objects, represented as a vector field. The top right panel provides the view of these vectors stacked at the position of the reference object, with the green circle representing the radius that contains 50% of the matches. The bottom panel provides a histogram of the offsets, with the median indicated. The offset[14], accumulated over the entire patch, is somewhat more than 0.1 arcsec in most cases. The detailed figures for each patch (not shown) shows the offsets for each matched star.

The RMS of the WCS solution must not be confused with astrometric fidelity in any sense. The RMS is merely a measure of how well the measured centroids of the detected stars match an imperfect coordinate representation: currently a cylindrical equal-area projection with a rough allowance for image distortion. The achievable astrometric accuracy requires a global astrometric solution, which has not been performed for this release.

[14] The 2-D distribution in this plot would be expected to vary as the Rayleigh distribution: $r \cdot \exp[-r^2 / (2 \cdot \sigma^2)]$, which peaks at $r = \sigma$.

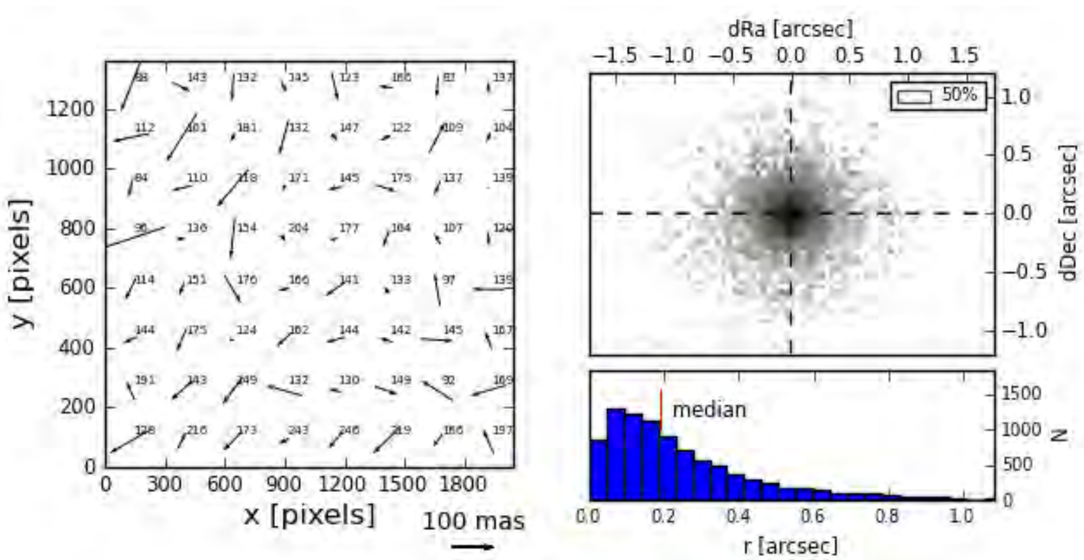


Figure 5-5: Astrometric data quality plots for set of exposures in the *r*-band over the selected patch. *Left:* Average distribution of offsets between matched star coordinates in the reference catalog and those recovered in pipeline processing. Measured coordinates are determined from the image world-coordinates of the stellar centroids. *Top right*: Stacked offsets from reference catalog position for all stars; *green circle* indicates the radius containing 50% of the matches. *Bottom right*: Histogram of the offsets, with median value indicated.

Object Completeness

For each patch, the detection completeness for stars is assessed based on the counts of four classes of objects, as a function of magnitude. *Matched* objects are detections that match 1-to-1 with reference catalog; *Blended* objects match N-to-1 with the reference catalog; *Orphan* objects do not match any entry in the reference catalog, and may include false positives, transients and asteroids; while *Unmatched* objects are those present in the reference catalog but not detected in the science image. The measure of completeness is computed as (matched + blended) / (matched + blended + unmatched), indicated with the blue line in the top-left panel of Figure 5-6. The magnitude below which the completeness is < 50% is indicated with the vertical line. The bottom-left panel shows the same breakdown for galaxies only. Orphans are included in both plots. The summary FPA figure provides a visual representation of the photometric depth, as shown in Figure 5-6.

The completeness plots likely paint too pessimistic a picture: they do not account for objects in the outermost few pixels of each patch (where edge effects compromise the results), which were not searched. Moving objects are also missing from the reference catalogs, which causes real detections to appear spurious.

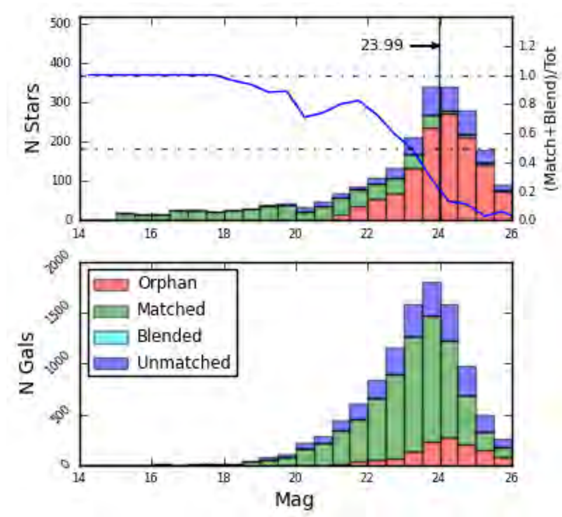


Figure 5-6: Histograms of the detected stars (*upper left*) and galaxies (*left*) in the reference catalog as a function of object magnitude. Objects in this field were *matched* uniquely, matched but *blended* with another object, not matched, or not detected (see text).

Photometric Fidelity

The fidelity and accuracy of the photometry on individual patches for this data release is determined by comparing the magnitudes as measured in a variety of ways. For each patch, the diagnostic diagrams show the difference in magnitudes $m_1 - m_2$ as a function of m_1 for the photometric measurements shown in Table 5-2 below (see also Chapter 4.1.4).

Table 5-2: Types of Source Brightness Measurements

Magnitude	Description
$m_{\mathbf{CAT}}$	Source magnitude from the reference catalog of Annis et al. (2012)
$m_{\mathbf{AP}}$	Source magnitude from counts within a circular aperture; zero-point set so that, for stars, the aperture flux equals the total flux
$m_{\mathbf{INST}}$	Source magnitude derived from Gaussian model
$m_{\mathbf{MOD}}$	Source magnitude derived from multi-fit (i.e., multiple component) model
$m_{\mathbf{PSF}}$	Source magnitude derived from PSF size/shape at the position of the source

The pipeQA web pages provide links to five inter-comparisons among pairs of these types. The agreement between cataloged and measured PSF magnitude for all objects (stars + galaxies) shows generally good agreement at brighter magnitudes, as shown in Figure 5-7; the scatter below r~21 is consistent with photon statistics.

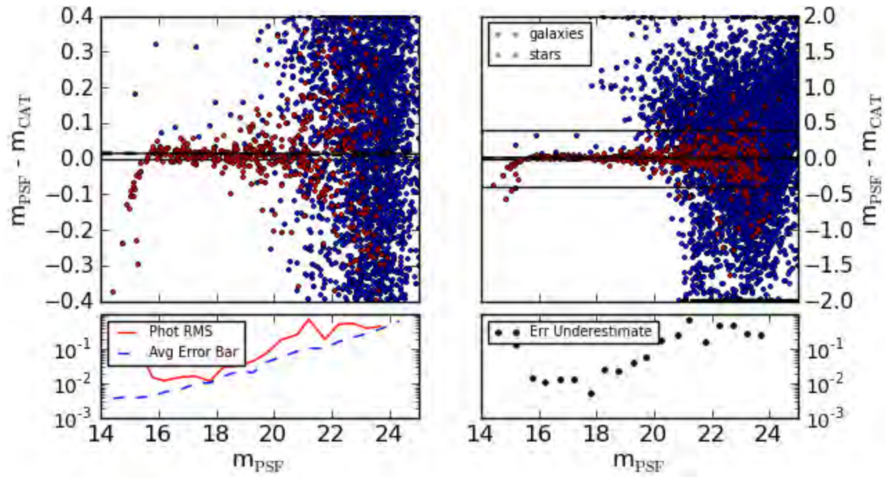


Figure 5-7: Comparison of the PSF magnitudes vs. those in the reference catalog (*right*) and zoomed in (*left*). Comparison is over the full patch, where stars (*red points*) are distinguished from galaxies (*blue*). The dispersion for m_{PSF}>21 to the limiting magnitude near 24 increases as expected for photon statistics. Plots of the photometric error bars appear below the comparison plots.

The width of the bright end of this distribution reflects the systematic floor in these measurement comparisons. For PSF magnitudes, this is typically 1-2%. The relationship of magnitude difference as a function of magnitude *for a single CCD* is shown in Figure 5-8 below.

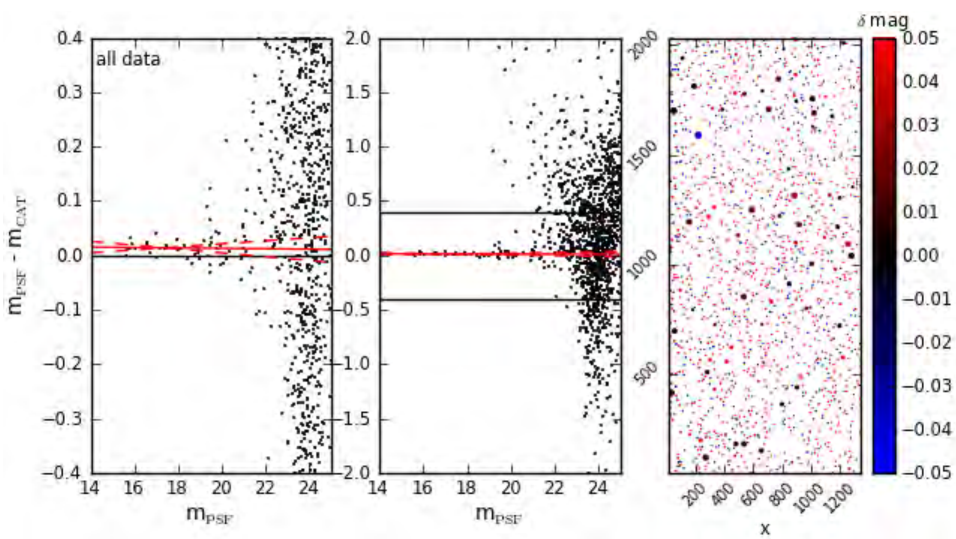


Figure 5-8: Comparison of the PSF magnitudes for all detctions vs. those in the reference catalog (*center*) and zoomed in (*left*) for a single patch. Stars (red) define the photometric calibration, and the linear trend is shown (*red dashed line*). Distribution of sources in the patch is also shown (*right*), color-coded by deviation in mag from the trend.

PSF Shape

For each patch, the diagnostic diagram (see Figure 5-9) the ellipticity of star shapes used in the PSF model are plotted as a function of position in the focal plane. The summary tract figures show the median vector (offset and angle) of the ellipticity for each patch, as well as the effective FWHM in arcsec for the final PSF model.

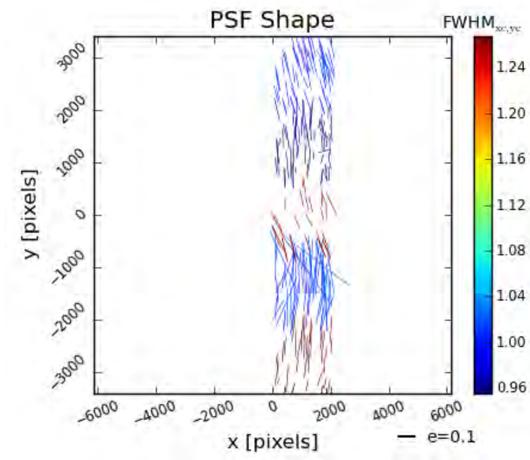


Figure 5-9: Magnitude and orientation of the PSF ellipticity for all stars in the patch. (The variation of ellipticity parameters across the tract is not shown here.)

5.2. Community Feedback

One of the major objectives of publishing data from these data challenges is for a wider, scientific audience to participate in the evaluation of the data, and to bring more expertise to bear on the technical challenges of data reduction and analysis. Feedback can take a number of forms, and will cover a wide variety of topics.

5.2.1. General Feedback

It is perhaps useful to summarize the types of feedback, what form it might usefully take, and the mechanisms that are in place for receiving it. The following are meant as guidelines.

This Handbook. If you find errors, incomplete information, or have requests and suggestions for improving this *Handbook*, please contact the HelpDesk: dm-help@lsst.org.

Science/Technical Issues. Please post to the Science User Forum (https://www.lsstcorp.org/sciencewiki/index.php?title=Special:AWCforum) any suggestions you have for new algorithms, approaches, and techniques for either analyzing the science quality of the data, or for improving the pipeline processing. This forum is also a good place to engage in discussions of what analyses might be most important or fruitful, and to see what other Collaboration members are doing.

Results of Analysis. If you have conducted some analysis of the data products from this data challenge, and have a written analysis in hand, these are likely best posted on the Science Wiki: http://www.lsstcorp.org/sciencewiki/index.php?title=Main_Page. Naturally the more thoroughly documented the analysis, the more useful it will be to others, including the DM team. In particular, it is important to describe which data were analyzed (i.e., selection criteria for catalogs).

Questions about Pipeline Processing: Detailed questions about how the pipelines work should be directed to the HelpDesk: dm-help@lsst.org.

Technical problems. If users have problems with accessing data products, the access tools (**Gator** and VOInventory), or if the related services do not respond or do not provide sensible results, please report this to the HelpDesk: dm-help@lsst.org. Please be as specific as possible about the date/time when the problem occurred, the input you provided, and any error messages that resulted.

5.2.2. Science/Technical Feedback

Evaluation of science data quality for LSST will be an involved and sometimes complex process, and it is clear that the Data Management design effort will for the near term be limited to relatively targeted, mostly basic assessments.

It would be helpful to receive evaluations that would help characterize the data quality, and to identify problems and shortcomings where they exist. Please post these ideas on the Science User Forum at https://www.lsstcorp.org/sciencewiki/index.php?title=Special:AWCforum.

5.3. References and Further Information

Contributing Authors

The following individuals contributed to the material presented in this chapter: Andy Becker, Steve Bickerton, Mario Juric, Simon Krughoff, Dustin Lang, Robert Lupton, and Dick Shaw.

References

Annis, J., et al. 2011, *The SDSS CoAdd: 275 Deg^2 of Deep SDSS Imaging on Stripe 82*, ApJ, submitted. Available at: http://arxiv.org/abs/1111.6619

Davis, M., et al. 2003, *Science Objectives and Early Results of the DEEP2 Redshift Survey*, SPIE, 4834, 161

Ivezic, Z., et al. 2007, *Sloan Digital Sky Survey Standard Star Catalog for Stripe 82: The Dawn of Industrial 1% Optical Photometry*, AJ, 134, 973

Ivezic, Z., et al. 2004, *SDSS Data Management and Photometric Quality Assessment*, AN, 325, 583

Shaw, R. A., Levine, D., Axelrod, T. S., Laher, R. R., & Mannings, V. G. 2010, *Science Data Quality Assessment for the LSST*, Proc. SPIE, 7740, 14

Glossary

The following glossary of technical terms that were used in this *Handbook* includes many of those that can be found on the Data Management Wiki.

Term	Definition
Adaptive (Gaussian) moments	The second moments of the source intensity distribution, measured using a scheme designed to have near-optimal signal-to-noise ratio.
Alert	Refers to the structured communication that is issued rapidly via the internet to the community that characterizes detection of one or more sources that are new, or have changed significantly in position or brightness, relative to the applicable ***image template***.
Amplifier	Electronic component of a ***CCD*** that is used to recover the signal during read-out. For LSST, multiple amplifiers on each CCD will enable simultaneous read out of adjacent regions of the detector. Often this term is used as a synonym for a read-out ***channel***.
Calibrated image	Also known as a ***calexp***: an image from a single visit to a region of sky that, in the case of LSST is the combination of two ***snap images***, each of which has been corrected for instrumental signature. The ***WCS*** determination, PSF characterization, photometric calibration, and various other characterizations have also been determined for calibrated images. Calibrated images have an effective exposure time equal to the sum of the components, which for LSST is expected to be 30 s.
Camera Column	Also abbreviated to ***camcol***, The output of one camera column of CCDs (each with a different filter) obtained during a ***run***. See SDSS glossary of terminology.
CCD	Charge-coupled device. This is the type of ***sensor*** that will be used in the LSST camera for detecting and recording radiation in the visible band. Contiguous portions of a CCD detector can be read out simultaneously through parallel output ***channels*** if the electronic design includes multiple ***amplifiers***.
Data Challenge	A structured set of activities that processes large volumes of astronomical data using the Data Management software stack over a massively parallel, high-throughput computing platform. The data challenges aim, over time, to demonstrate the ability to produce images and catalogs with the scientific fidelity, scalability, and throughput comparable to that expected from LSST.
DIA	Difference Image Analysis. Refers to the data products or catalog entries that are generated during the pipeline stage by this name, which include the detection new ***sources*** and brightness changes in known ***objects***.
Difference image	A pixel-by-pixel difference between the image being processed and an ***image template***, where the template has been warped to the same geometry, photometrically scaled, and background-matched. The resulting difference image is zero everywhere, apart from shot noise and objects that are new, or have changed in brightness or position, relative to the template.
Exposure	One of a pair of ***raw images*** in a single band, taken sequentially, of the same area of sky. Pairs of exposures facilitate the identification of cosmic ray artifacts. This term is a synonym for the less commonly used term ***snap image***.

Term	Definition
Field	A part of a ***camcol***, consisting of the images of the same part of the sky, in all 5 filters, taken contemporaneously. See SDSS glossary of terminology[15].
FITS	Flexible Image Transport System, which is the astronomical (IAU) standard for structured data in a file. Allowed contents include images, tables, and ***metadata*** stored in ASCII headers.
Footprint	The spatial extent of an image or of an astronomical object within an image. The footprint can be irregular in shape, such as that of a galaxy, or of a group of spatially overlapping but non-coincident images.
FPA	Focal Plane Array, or a regular grid of ***sensors*** placed at the focus of the imaging camera.
FoV	Field of view, used to describe the spatial extent of the sky observed with the ***FPA*** or with a portion of it (e.g., a single detector).
HDU	Header and Data Unit, or a data structure in a ***FITS*** file that consists of an ASCII header and the data that the header describes. Note that a (primary) HDU may consist only of a header, with no data blocks. FITS extensions are structured as HDUs that appear after the primary HDU in a FITS file.
Image Extension	An ***HDU*** within a ***FITS*** file consisting of a header plus a binary image array. The pixel values may be expressed in any form allowed by the FITS Standard (e.g., integer or floating-point).
Image Template	An image of a section of sky in a single band that is deep, of very high image quality, and where all transients, moving objects, and artifacts have been removed. Such images are used as templates to perform ***difference image*** analysis in order to detect variable, transient, and moving objects.
Metadata	Strictly speaking, information or data that describe other data, such as an image. Most metadata are stored in the ***science database***. Metadata also appear in the keyword/value pairs in the headers of FITS files.
MSS	Mass storage system, which stores large volumes of data on a variety of media (including spinning disk and tape), whose contents appear to a user to be a regular file system.
Object	Refers to an *astronomical object*, such as a star, galaxy, asteroid, or other physical entity. Objects can be static, or change brightness or position with time. Generally an object will be associated with more than one instance of a ***source*** detection.
Patch	A skymap ***patch*** is a subdivision of a ***tract***, or major section of sky in the ***SkyMap*** tesselization scheme. Patches are stored as individual FITS files; the size arranged to fit into memory for most end-user compute platforms.
Pipeline	A unit of data processing software that is independently executable within the Data Management System, and which performs a logically connected sequence of operations. Pipelines are composed of one or more processing ***stages***.
Production	Episodes of data processing that are organized to achieve a particular purpose, such as issuing event ***alerts*** during a night's observing; generating ***calibrated images***, ***source*** and ***object*** catalogs, and ***image templates*** on an annual basis; or constructing calibration reference products.

[15] See http://www.sdss3.org/dr8/glossary.php

Term	Definition
Provenance	The structured description of the origin of a data product, including its processing history, data dependencies, and software version identifier. All provenance information is stored in the science database, although some of it is replicated in the headers of image data products.
PSF	Point Spread Function, or the two-dimensional brightness profile of a point source (i.e., an unresolved astronomical object) as it is realized by the detector, including all effects of the atmosphere, telescope/camera optics, and detector sampling.
Raw Image	A regular array of pixel data, and associated ***metadata***, that were obtained from the observing environment in a single exposure from one or more detectors in the ***FPA***.
Scanline	A subdivision of a ***strip***, where each strip is composed of 6 scanlines. They are defined in great circle (i.e., survey) coordinates, and are bounded on the top and bottom by lines of constant ν, with no E or W boundaries. The term is often, but not quite correctly, used interchangeably with ***camcol***.
Science Database	The repository of all metadata that describe all observations, their provenance, and their quality attributes; and of the measurements that have been made on the images, such as source positions, brightnesses, and other attributes.
SDQA	Science Data Quality Analysis. The software and processes that measure quality items of interest (metrics), and compare them to thresholds that define nominal conditions.
Sensor	Generic engineering term used to refer to a single detector, which in the case of the LSST is a ***CCD***s.
SkyMap	LSST all-sky tesselization scheme. The imaging data are arranged as a sequence of large, rectangular, overlapping ***tracts***. Tracts contain an inner region described by a collection of verticies. The inner regions exactly tile the portion of sky covered by the SkyMap; all pixels beyond the inner region provide overlap with neighboring tracts and a rectangular outer border.
Sky Tile	A region of sky with an extent that depends upon position and sky tiling scheme. For many regions the extent is roughly $0°.5 \times 0°.5$. Full-sky images, such as templates, deep-detection co-adds, etc. will be partitioned in this way.
Source	A single detection of an astrophysical ***object*** in an image, the characteristics for which are stored in the Source Catalog of the science database. The Data Management System attempts to associate multiple source detections to single ***objects***, which may vary in brightness or position over time.
Stage	A portion of a ***pipeline*** that performs a discrete algorithmic operation, and which is not independently executable.
Strip	A scan along the line of constant survey latitude η, the center of a ***stripe*** is set at a given n; centers of strips have a boresight offset added. Because the columns of SDSS CCDs include gaps, two offset but overlapping observations are used to fill in a ***stripe***. These two scans are called strips, one N and one S. Note that while strips are centered on a given n, they are not bound by η lines; rather they are rectangular regions in the geometry of the camera.
Stripe	Stripes are the sum of two ***strips***, defined in survey coordinates (λ,η). A stripe is defined by a line of constant η, bounded on the N and S by the edges of the two ***strips*** that make up the stripe, and bounded on the E and W by lines of constant λ.

Term	Definition
Tract	Sub-division of the celestial sphere in the ***SkyMap*** tesselization scheme, with a size that is adjustable for the dataset to be processed and a configurable projection. For SDSS Stripe 82, tracts are only one patch high and use a cylindrical equal-area (CEA) projection.
WCS	World Coordinate System, which is the specification of a mapping between detector coordinates (i.e., pixels) to a reference system such as celestial coordinates. The formalism for the WCS mapping is defined in the ***FITS*** standard.