Wide-field Infrared Survey Explorer (WISE)

DBPREP Subsystem Design Document

Version 2.0

23 June 2009

Prepared by: Tracey L. Evans

Infrared Processing and Analysis Center
California Institute of Technology

WSDC D-D022

Approved By:

Roc Cutri, WISE Science Data Center Manager

Lin Yan, WISE Science Data Center Archive Scientist

Tim Conrow, WISE Science Data Center System Architect

Tracey Evans, WISE Science Data Center DBPREP Cognizant Engineer

Revision History

Date	Version	Author	Description
2 April 2009	1.0	Tracey L. Evans	Initial Draft
23 June 2009	2.0	Tracey L. Evans	Modified for DBPREP v.2.0

Table of Contents

1 INTRODUCTION

1.1 Subsystem Overview

This document presents the requirements, design, algorithms, and state of implementation of the *DBPREP* subsystem of the WSDC data processing system. *DBPREP* finds and reads in non-image data and metadata files output from various subsystems in the scan-frame and multi-frame pipelines, adds a few columns in the data structures that no subsystem provides (such as galactic and ecliptic coordinates), and outputs the combined data in files that are readable by the database loading system. *DBPREP* is run independent of the pipelines.

1.1.1 Requirements

The *DBPREP* subsystem is required to collate the non-image pipeline data to be entered into the database, and output them in a format that is readable by the database loading system.

The Level 4 requirements supported by this subsystem are as follows:

L4WSDC-002: The WSDC shall produce a Source Catalog derived from the images used to generate the WISE digital Image Atlas.

L4WSDC-006: The WSDC shall generate a preliminary Source Catalog derived from the WISE preliminary digital Image Atlas that contains sources detected in unconfused regions in the first 50% of the sky that is surveyed.

L4WSDC-015: The WISE Source Catalog shall contain the measured in-band fluxes or flux upper-limits in the four WISE bands for objects detected in at least one band in the WISE Atlas Images.

L4WSDC-016: The WISE Source Catalog shall contain uncertainties in the flux measurements (one sigma) in all bands for which a source is detected.

L4WSDC-085: The WISE Source Catalog shall as a goal contain flux estimates for sources in any band in which the object has saturated the WISE image data.

L4WSDC-017: The WISE Source Catalog shall contain equatorial (J2000) coordinates for objects detected in at least one band.

L4WSDC-018: The WISE Source Catalog shall contain uncertainties in the coordinates measurements for each object.

L4WSDC-019: The WISE Source Catalog shall contain one or more quality flags for each object entry that indicate if a flux measurement may have been contaminated due to the proximity of the source to an image artifact or another nearby source.

L4WSDC-020: The WISE Source Catalog shall contain one or more quality flags for each object entry that indicate if the detection of that object may be a spurious detection of an image artifact or transient event.

L4WSDC-025: The WSDC shall generate and maintain a database of source information extracted from the calibrated, single-epoch images for the duration of the project for use by the Project Team. The purposes of this archive are quality assurance, transient analysis and moving object identification.

L4WSDC-027: The WSDC shall identify and compile a listing of known solar system objects that are positionally associated with source extractions in the WISE single-epoch image frames.

L4WSDC-039: Within 3 days from receipt of a given data set at the WSDC all data shall be processed through the WSDS Scan/Frame pipeline which performs basic image calibration and source extraction from on images from individual orbits. The results of this processing step shall be Level 1 source extractions and image data, which are loaded into the WISE Level 1 extracted Source Working Database (L1WDB) and Image Archive allowing access by the WISE Science Team for external quality assessment.

L4WSDC-041: As a goal, the WSDC shall combine image data from multiple orbits and extract sources from the combined images at intervals of no shorter than 3 days and no longer than 30 days to generate a temporary, intermediate combined image archive and source database for the purpose of science data quality assessment by the WISE Science Team and WSDC.

L4WSDC-051: The WSDC shall make the WISE catalog and image products available to the community via the internet through appropriate web-based tools.

L4WSDC-052: As a goal, the WSDC will maintain the data products in a way that distribution of the complete WISE Source Catalog to users via portable media would be possible.

L4WSDC-053: The WSDC shall make the Image Atlas and Catalog products accessible to the astronomical community in collaboration with the NASA/IPAC Infrared Science Archive (IRSA) to ensure long-term availability beyond the end WISE missions operations and data processing phase, and to insure interoperability with other NASA mission archives.

L4WSDC-056: The WSDC shall maintain an archive of metadata derived from data processing for the individual science images for the duration of the project for the purpose of analysis and support of image access tools.

L4WSDC-061: The WSDC archive shall provide a web-based interface to enable selection of sources extracted from single-epoch frames and/or combined Atlas Images based on position, flux, or combinations of any parameter maintained in the extracted source databases or Source Catalog.

1.1.2 Current Description

DBPREP finds and reads in non-image data and metadata files output from various subsystems in the scan-frame and multi-frame pipelines, and adds a few columns in the data structures that were not found in the input files:

- Galactic and ecliptic coordinates
- Source, frame, and scan identifiers
- Source detection, photometric quality, reliability, and morphology indicators
- Information on associations to other catalogs
- Observation date and time
- Position and counter data required by internal IRSA routines for data access

It then outputs the combined data in files that are readable by the database loading system. *DBPREP* is run independent of the pipelines, and must be run after Quality Assurance has been completed and 2MASS catalog associations have been found.

1.2 Applicable Documents

This subsystem conforms to the specifications in the following project documents:

- WISE Science Data Center Functional Requirements Document, WSDC D-R001
- WISE Science Data System Functional Design, WSDC D-D001
- Software Management Plan, WSDC D-M003
- WSDC/IRSA Archive Design, WSDC D-D007
- Various Interface documents, WSDC D-I*

1.3 Acronyms

2MASS	Two-Micron All-Sky Survey
DB	Database
FRD	Functional Requirements Document
IPAC	Infrared Processing and Analysis Center
IRSA	Infrared Science Archive project at IPAC
L4WSDC	Level 4 requirements for the WISE Science Data Center

QA	Quality Assurance
RTB	Regression-test baseline
SDS	Subsystem Design Specification
SIS	Software Interface Specification
SSO	Solar System Object
W1	WISE wavelength channel 1, 3.3 microns
W2	WISE wavelength channel 2, 4.7 microns
W3	WISE wavelength channel 3, 12 microns
W4	WISE wavelength channel 4, 24 microns
WISE	Wide-field Infrared Survey Explorer
WSDC	WISE Science Data Center
WSDS	WISE Science Data System

2 INPUT

2.1 Control Input: Command Line Parameters

DBPREP accepts the user's input parameters and uses them to select the files to read as input, create the output file names, and select the output file format to use.

Typing "dbprep -help2" will provide program help, listing all of the possible input control parameters for *DBPREP*. The first large set of parameters in the help, up to (but not including) "`frame_dirs,frdirs`", is part of the "basic" parameter set for the WISE pipeline, and I've only included here the ones that are used by *DBPREP*; for more information on this set of parameters see their documentation or source code. The *DBPREP* program's use of the input parameters it receives is described in the next section.

```
Command => DBPrep

Prepare and write IPAC or bar-delimited format source and metadata files for DB
loading

out_base,outbase,outb        : string              = %in_base%
Output file(s) base name

out_dir,outdir,outd              : file              = .
Input directory

qa_dir,qad                          : file              = %out_dir%/%qa_sub_dir%
Directory for QA output

qa_sub_dir,qasubdir,qasubd,qasub            : file quiet         = qa
Directory for QA output
```

DBPREP uses the out_base and out_dir parameters to create the output file full pathnames. The qa_dir and qa_sub_dir parameters indicate where certain input files can be found in the file structure.

```
verbose,v                    : switch
Increase verbosity of informatory output
```

```
log_file,logfile,log         : string
Write (or inherit) standard log file
'1' => do the usual thing.
'+' prefix means append.
'=' prefix means overwrite (the default).
Otherwise assume a literal file name.
```

```
test                         : switch
Do not actually run but display what would be run
```

```
debug,dbg                    : string private        = 0
Add targetted debug output
```

Setting the parameter debug to '1' or similar will turn on the debugging option for *DBPREP*. This sends a whole bunch of output to STDOUT. Most people will never want to use this parameter.

```
frame_dirs,frdirs            : list of file (glob)
Frame directories (or fileglobs) to process
```

```
frame_level,level                    : string          = 1b
# Level of frames to process -- 1b (single-frame pipeline) or 3 (multiframe pipeline)
```

The frame_dirs parameter is a list (comma-separated) of input frame directories (containing input data files) or fileglobs of frame directories, *e. g.* '/wise/fops/scans/1d/*/fr/*'. The frame_level parameter indicates the data level to be used: 1b (scan-frame pipeline) or 3 (multi-frame pipeline).

```
out_format,outform,oform             : string          = ipac
# Output file format -- ipac (ipac ascii tbl format) or bar (bar-delimited DB input format)
```

The out_format parameter indicates which output format should be used. The IPAC ASCII format is more human-readable, but the bar-delimited format is used as input by the high-speed database loading processors.

```
tbl_version,tblvsn,vsn               : string          = 1
# Database table version number
```

The `tbl_version` number indicates the database table version to be used in the output file names. This version changes when the database table schemas change.

2.2 Data Input

DBPREP uses the `frame_dirs` input parameter to determine the names of each of the files it must read. These files are generated by various upstream modules, including *ICAL, WPHOT, SFPREX, SPCAL, FPCAL, WASSOC, QA,* and *SSOID*, and contain either source data (including source associations data) or metadata (including QA data). The file names and their applicable SIS documents for the Level 1b data are:

- Source data
  - <frame_dir>/<scan_id><frame_num>-mdex-1b.tbl (pht01 - WPHOT Output Photometry Table SIS, WSDC D-I107)
  - <frame_dir>/<scan_id><frame_num>-assoc-2mass_mch.tbl (arc09 - 2MASS PSC Association Matches File Interface)
- Metadata
  - <frame_dir>/<scan_id><frame_num>-w<band>-int-1b.fits (sfp01 - Lev1Bwcsfits SIS, WSDC D-I102)
  - <frame_dir>/<scan_id><frame_num>-meta-sfpipe.tbl
  - <frame_dir>/<scan_id><frame_num>-meta-sfprex.tbl (rex01 - SFPRex Metadata (SFMD) file SIS, WSDC D-I116)
  - <frame_dir>/<scan_id><frame_num>-meta-ical.tbl (icl04 - ICAL QA diagnostics/metrics/plots SIS, WSDC D-I105)
  - <frame_dir>/<scan_id><frame_num>-meta-mdex-1b.tbl (pht02 - WPHOT metadata table SIS, WSDC D-I113)
- Calibration source matches data
  - <scan_dir>/<qa_dir>/<scan_id>-spcalqa.tbl (pcl06 - Source matches to calibration stars in each scan, QA output SIS, WSDC D-I120)
- Calibration metadata
  - <frame_dir>/<scan_id><frame_num>-meta-fpcal-1b.tbl (pcl04 - FPCAL frame-level metadata table SIS, WSDC D-I114)
  - <scan_dir>/<scan_id>-spcal.tbl (pcl05 - Scan level photometric zero point values for four bands SIS, WSDC D-I115)
- Solar System Object source matches data
  - <frame_dir>/<scan_id><frame_num>-ssoid-1b.tbl
- Solar System Object metadata
  - <frame_dir>/<scan_id><frame_num>-meta-ssoid-1b.tbl

3 PROCESSING

3.1 Initialization

The *DBPREP* program initializes itself by reading and processing its control input parameters and prints out a program "banner" with version and date-time information. It uses the `frame_dirs` parameter to set the list of frames and scans it will access.

3.2 Reading and Processing Data

3.2.1 Frame Directory Data

DBPREP next loops through the list of frames, using the current frame directory name to get the directory path, scan ID and frame number to be used for the data file names. It first checks for the existence of the image file <frame_dir>/<scan_id><frame_num>-w1-int-1b.fits; if it cannot be found it moves on to bands w2 through w4 in turn. When it finds one of the files, it reads that file's FITS header for the DATE_OBS keyword and stores it as the frame observation date-time.

3.2.1.1 Source Data

It then moves on to read and store the source data from the *WPHOT* output, file <frame_dir>/<scan_id><frame_num>-mdex-1b.tbl, including the added calibration data and contamination and confusion flags from *PCAL* and *ARTID*, respectively. It adds certain columns to this data to contain:

- Source, frame, and scan identifiers
- Galactic and ecliptic coordinates
- Contamination and confusion flag columns, if that data is missing from the source data file
- Detection, photometric quality, reliability, and extended source flags
- Solar system object and 2MASS PSC association cross-references along with 2MASS association position match and magnitude information
- The modified Julian date of the frame observation
- Source counter and unit sphere position and index data to be used by the IRSA database access routines, if the `out_format` parameter is set to the bar-delimited format.

Next it checks for the existence of the 2MASS PSC association cross-reference data file <frame_dir>/<scan_id><frame_num>-assoc-2mass_mch.tbl; if that file is not found in the expected location, *DBPREP* spawns a *WASSOC* process to create the file, containing only the data columns that *DBPREP* needs. It then reads and stores the 2MASS PSC association cross-reference data from the file and indexes the data by source ID.

It then fills the data for the added columns where possible: It constructs the frame and scan identifiers from the source data file name. The galactic and ecliptic coordinates are calculated using the source's equatorial position. If the source ID is found in the index to the 2MASS PSC associations, *DBPREP* stores the associated 2MASS source data in the proper columns.

DBPREP determines if the source is detected in a band and the photometric quality in that band by checking the profile fit photometry for that band.

- If the magnitude is less than 90.0 and the SNR is greater than or equal to 10.0, it is a detection with a photometric quality of 'A'.
- If the magnitude is less than 90.0 and the SNR is between 3.0 and 10.0, it is a detection with a photometric quality of 'B'.
- If the magnitude is less than 90.0 and the SNR is less than 3.0, it is a detection with a photometric quality of 'C'.
- However, if the magnitude sigma is greater than 9.0, it is an upper limit and a non-detection, with a photometric quality of 'U'.
- If the magnitude is above 90.0 it is a non-detection with a photometric quality of 'X'.

The detection and photometric quality flags are constructed using this information from each band. The modified Julian date is calculated from the value of the frame observation date-time in the header of the frame image FITS file.

If the `out_format` parameter indicates that the output will be in the bar-delimited format, *DBPREP* calls a Perl module that calculates the unit sphere position and index data for all of the source positions and sets the source counter to "0", which will be set to the row number when the data is loaded into the database.

(The reliability flag, extended source flag, and SSO association are TBA.)

After adding and filling the columns, *DBPREP* deletes certain columns not required for Level 1b output and appends this frame's source data to the output source file, to save memory (see section 4.1).

3.2.1.2 Source Metadata

DBPREP then reads the files containing the frame metadata, listed previously:

- <frame_dir>/<scan_id><frame_num>-w<band>-int-1b.fits
- <frame_dir>/<scan_id><frame_num>-meta-sfpipe.tbl
- <frame_dir>/<scan_id><frame_num>-meta-sfprex.tbl
- <frame_dir>/<scan_id><frame_num>-meta-ical.tbl
- <frame_dir>/<scan_id><frame_num>-meta-mdex-1b.tbl

It only reads certain metadata from each of these files, since not all of this data will be archived in the database. It saves all of the data from the above files for all frames in one frame metadata structure, since it will be output in a single frame metadata file, one row per frame. *DBPREP* also creates a new column for a row counter with a value of "0" if the `out_format` parameter is set to the bar-delimited format.

3.2.1.3 Calibration Metadata

Next *DBPREP* reads certain metadata quantities from the file containing the calibration frame metadata for the frame, <frame_dir>/<scan_id><frame_num>-meta-fpcal-1b.tbl. It again saves all of the data from this file for all frames in one calibration frame metadata structure, since it will be output in a single calibration frame metadata file, one row per frame. *DBPREP* also creates a new column for a row counter with a value of "0" if the `out_format` parameter is set to the bar-delimited format.

3.2.1.4 Solar System Object Source Matches Data

It then reads the Solar System Object source matches data from the file <frame_dir>/<scan_id><frame_num>-ssoid-1b.tbl. It adds certain columns to this data to contain the source, frame, and scan identifiers, as well as the source counter and unit sphere position and index data to be used by the IRSA database access routines if the `out_format` parameter is set to the bar-delimited format.

As with the source data, *DBPREP* then constructs the source, frame, and scan identifiers from the scan ID, frame number, and WISE source number in the file. If the bar-delimited output format is selected, *DBPREP* calls the Perl module that calculates the unit sphere position and index data for all of the source positions and sets the source counter to "0". *DBPREP* saves all of the data from this file for all frames in one SSO source matches data structure.

3.2.1.5 Solar System Object Metadata

DBPREP next reads certain metadata quantities from the file containing the SSO frame metadata for the frame, <frame_dir>/<scan_id><frame_num>-meta-ssoid-1b.tbl. It adds the frame and scan identifiers to this data, as well as the row counter of "0" if the `out_format` parameter is set to the bar-delimited format, and saves all of the data from this file for all frames in one SSO frame metadata structure.

3.2.2 Scan Directory Data

DBPREP now loops through the list of scans to which all the input frames belong, using the current scan directory name to get the directory path and scan ID to be used for the data file names. The first file name also requires the value of the `qa_dir` parameter.

3.2.2.1 Calibration Source Matches Data

In this loop, *DBPREP* first reads certain columns from the file containing the calibration source matches data, <scan_dir>/<qa_dir>/<scan_id>-spcalqa.tbl. As before, if the bar-delimited output format is selected, it adds the source counter and unit sphere position and index columns to be used by the IRSA database access routines, calls the Perl module that calculates the unit sphere position and index data, and sets the source counter to "0". All of the data from this file for all scans is saved in one calibration source matches data structure.

3.2.2.2 Calibration Metadata

Next *DBPREP* reads the calibration scan metadata file, <scan_dir>/<scan_id>-spcal.tbl. It adds columns containing the zero-point type to be used, profile-fit photometry or aperture photometry, and flags from the calculations of both types of zero-points. These columns are filled using the values found in corresponding keywords in the header of the metadata file. Also, if the `out_format` parameter is set to the bar-delimited format, it adds the row counter column with values of "0". All of the calibration scan metadata for all scans is saved in one calibration scan metadata structure.

4 OUTPUT

After *DBPREP* has finished reading in all the required data from the frame and scan directories specified by the `frame_dirs` parameter, it constructs seven output file names using the `out_base` and `out_dir` parameters and writes the data to those files.

4.1 Source Data

To save memory, *DBPREP* writes out the source data for each frame separately, when it has finished reading and processing that frame's source data. The file name is constructed from the `out_base` and `out_dir` parameters and the file type "src_data", the version from the file level ("1b" or "3"), and the file format determined by the `out_format` parameter ("bartbl" or "tbl"). For example:
<out_dir>/<out_base>-src_data-1b.bartbl

For each frame *DBPREP* appends the source data to the output file. It uses the format described in the data dictionary/SIS document "arc01 - Level 1b Source Data Table Archive DB Interface Specification," WSDC D-I138.

4.2 Source Frame Metadata

After processing all the input data from the frame and scan directories, *DBPREP* writes out the collected source frame metadata. The output file name is constructed from the `out_base` and `out_dir` parameters and the file type "src_framedata", the version from the file level ("1b" or "3"), and the file format determined by the `out_format` parameter ("bartbl" or "tbl"). For example:
<out_dir>/<out_base>-src_framedata-1b.tbl

There is one data row per frame. The format is described in the data dictionary/SIS document "arc02 - Level 1b Frame Metadata Table Archive DB Interface Specification," WSDC D-I139.

4.3 Calibration Source Matches Data

Next *DBPREP* writes out the collected calibration source matches data. The output file name is constructed from the `out_base` and `out_dir` parameters and the file type "cal_data", the version from the file level ("1b" or "3"), and the file format determined by the `out_format` parameter ("bartbl" or "tbl"). For example:
<out_dir>/<out_base>-cal_data-1b.bartbl

There is one data row per calibrator source match. The format is described in the data dictionary/SIS document "arc03 - Level 1b Calibration Match Data Table Archive DB Interface Specification," WSDC D-I140.

4.4 Calibration Frame Metadata

It then writes out the collected calibration frame metadata. The output file name is constructed from the `out_base` and `out_dir` parameters and the file type "cal_framedata", the version from the file level ("1b" or "3"), and the file format determined by the `out_format` parameter ("bartbl" or "tbl"). For example:
<out_dir>/<out_base>-cal_framedata-1b.tbl

There is one data row per frame, and the format is described in the data dictionary/SIS document "arc04 - Level 1b Calibration Frame Metadata Table Archive DB Interface Specification," WSDC D-I141.

4.5 Calibration Scan Metadata

DBPREP next writes out the collected calibration scan metadata; the output file name is constructed from the `out_base` and `out_dir` parameters and the file type "cal_scandata", the version from the file level ("1b" or "3"), and the file format determined by the `out_format` parameter ("bartbl" or "tbl"). For example:
<out_dir>/<out_base>-cal_scandata-1b.bartbl

There is one data row per scan, and the format is described in the data dictionary/SIS document "arc05 - Level 1b Calibration Scan Metadata Table Archive DB Interface Specification," WSDC D-I142.

4.6 Solar System Object Source Matches Data

The Solar System Object source matches data is written next; the output file name is constructed from the `out_base` and `out_dir` parameters and the file type "sso_data", the version from the file level ("1b" or "3"), and the file format determined by the `out_format` parameter ("bartbl" or "tbl"). For example:
<out_dir>/<out_base>-sso_data-1b.bartbl

There is one data row per SSO source match, and the format is described in the data dictionary/SIS document "arc06 - Level 1b Solar System Object Match Data Table Archive DB Interface Specification," WSDC D-I143.

4.7 Solar System Object Frame Metadata

And finally, *DBPREP* writes out the collected SSO frame metadata. The output file name is constructed from the `out_base` and `out_dir` parameters and the file type "sso_framedata", the version from the file level ("1b" or "3"), and the file format determined by the `out_format` parameter ("bartbl" or "tbl"). For example:
<out_dir>/<out_base>-sso_framedata-1b.bartbl

There is one data row per frame, and the format is described in the data dictionary/SIS document "arc07 - Level 1b Solar System Object Frame Metadata Table Archive DB Interface Specification," WSDC D-I144.

5 TESTING

Version 1.1 of *DBPREP*, identified as version "dbprep 3664 2009-03-18 23:04:48Z evans" in the logfile (indicating Subversion revision 3664, dated 2009-03-18 23:04:48Z, committed by user 'evans'), was run on the WSDS version 2.1 regression-test baseline on 1 Apr 2009. The output data and run log file are currently found in /wise/evans/dev/db/rtb/v2.1.

Version 2.0 of *DBPREP*, identified as version "dbprep 4692 2009-06-12 01:30:19Z evans" in the logfile (indicating Subversion revision 4692, dated 2009-06-12 01:30:19Z, committed by user 'evans'), was run on the WSDS version 2.4 set of data for scans 76543a, 76543b, and 76544a on 11 Jun 2009. The output data and run log file are currently found in /wise/evans/dev/db/rtb/v2.4_2.0. (TBD: add to the Subversion structure.)