Wide-field Infrared Survey Explorer (WISE)

Archive Preparation and Transfer Procedures

Version 1.0

26-October-2009

Prepared by: Lin Yan

Infrared Processing and Analysis Center
California Institute of Technology

WSDC D-C003

Approved By:

Lin Yan, Archive Scientist, WISE Science Data Center

Davy Kirkpatrick, QA Scientist, WISE Science Data Center,

Roc Cutri, WISE Science Data Center Manager

Revision History

Date	Version	Author	Description
04-Sept-2009	Version 0.5	Lin Yan	Initial Draft
26-October-2009	Version 1.0	Lin Yan	

Table of Contents

1 INTRODUCTION

This document describes the procedure with which the WISE data product will be loaded into the archive database. The mission requirement is that within 1 week of receiving the raw data at IPAC, the WISE archive will deliver the data product, including the source catalogs and reduced images to the science team. Therefore, the WISE archive will need to load new datasets every 3 days, i.e., archive data loading frequency of twice a week.

2 FLOW CHART FOR THE ARCHIVE LOAD AND TRANSFER PROCEDURE

In order to ensure the WISE data being archived in a timely fashion, we plan to have the IRSA release newly loaded on Mondays and Thursdays every week. These two dates are specified only for the data being accessible by the WISE science team. The dates and time for the actual data loading into the archive can be flexible and not necessarily tied to these two fixed dates. For example, the actual archive loading could happen on Friday, then the release to the users on Monday. With this fixed schedule, we will plan the required activities prior the IRSA database loading. This will also give enough time for the WISE archive and QA team to validate the newly loaded data in the IRSA database.

We visually illustrate the archive data loading and transferring procedure with the flow chart below. We summarize the data loading procedures as follows.

1. When the raw data comes down WSDC, it will be ingested and processed through the WISE pipeline.

2. The WISE QA team will exam and validate the processed data and inform the WISE operation team when the data is ready for loading.

3. The WISE OPS team will then prepare the source tables and stage the data for IRSA loading. Fits images will be copied from the WISE server to the IRSA machine. Specifically:

    a. The WISE OPS team will make a master schedule of various activities with a logical sequence. For example, DBPREP and IMPREP are two codes which need to be executed in order to prepare the source tables and fits images for the IRSA loading.

b. DBPREP will prepare the tables, attach the loading ID to the data, and stage the tables in a specific disk area from which the IRSA database loading script will fetch the data.
c. IMPREP will prepare the image fits data, and generate a list of files which can be used as an input to an ftp script which will copy the specific set of fits images from the WISE disks to the IRSA data disks.
d. DBPRPE and IMPREP will also make data file manifests which will keep records of files which have been transferred from WISE to IRSA.
e. Besides executing DBPREP and IMPREP, the WISE OPS team will also physically execute the script for the fits image transfer and also the script for the IRSA database loading.
f. DBPREP will stage the prepared data tables for loading on a WISE disk. We plan to keep these files for 1 week, i.e. (2-3) data loads, before deleting the files.

4. The WISE data will be loaded into the IRSA database. The current plan is following:

    a. The full source catalog database will be organized into individual segments, and the data from each loading forms one segment. Each segment database will be tagged by a load ID so that in the event of failure, the specific set of data will be traced easily and replaced without interruption of the full database.
    b. The IRSA team will provides scripts for both loading the source tables into the IRSA database as well as transferring the fits images.
    c. The script for transferring the fits images will be limited to a reasonable speed, for example, 18MB/sec, in order not to impose any interruption to the WISE pipeline processing.

5. After the new dataset is loaded, the WISE QA team and the external (outside WSDC) WISE science team will be able to view the newly loaded immediately. If no problems have been identified, the newly loaded dataset will be automatically a part of the full archive. If problems are found, the IRSA team will be alerted and the newly loaded dataset will be removed from the access. Both WISE archive and QA team will work together with the IRSA team to identify the nature of the problems and re-load the archive with the correct dataset. Figure 2 shows the flow chart how the WISE QA team will interact with various groups to validate the newly loaded WISE data and keep track of any anomalies. As stated before, DBPREP and IMPREP will generate two manifests listing all of the files transferred from the WISE to IRSA disks. The WISE QA will perform checksum on the new data to be sure that the correct quantity of data being loaded into the archive.

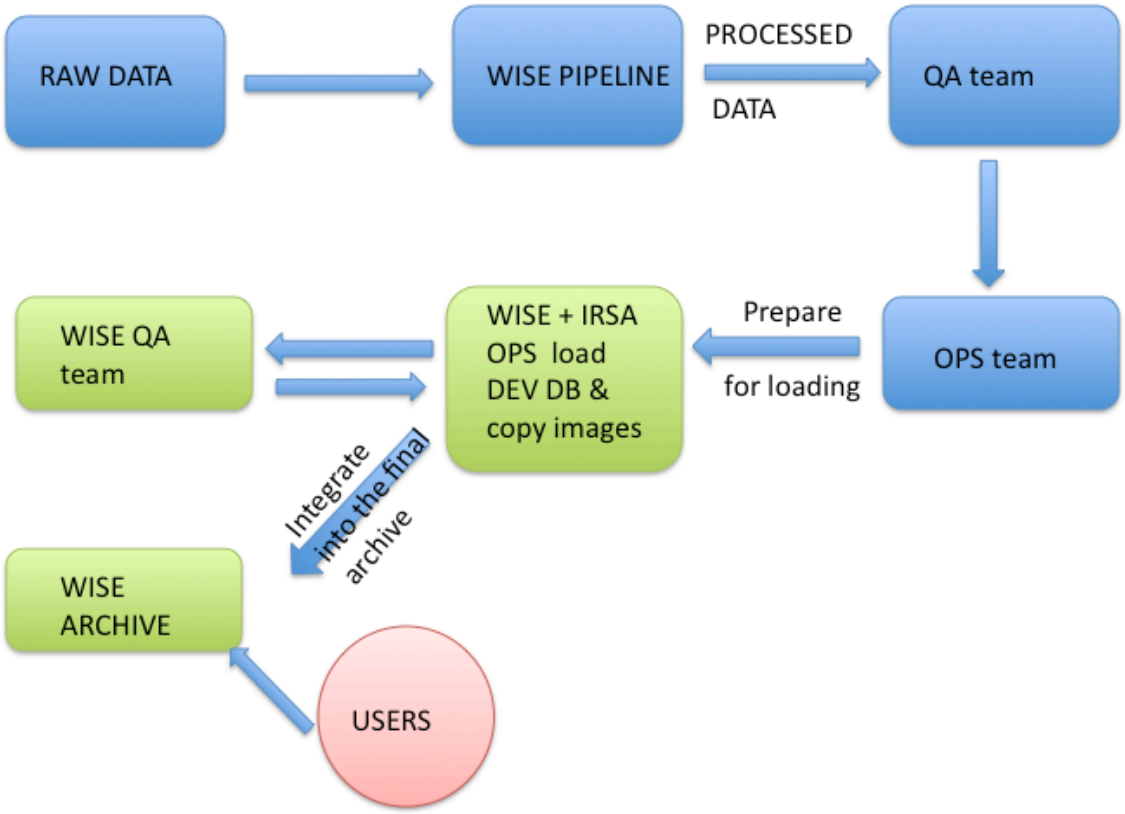

Figure 1. The data flow chart from the raw data to the final processed and archived data during the WISE mission operation.

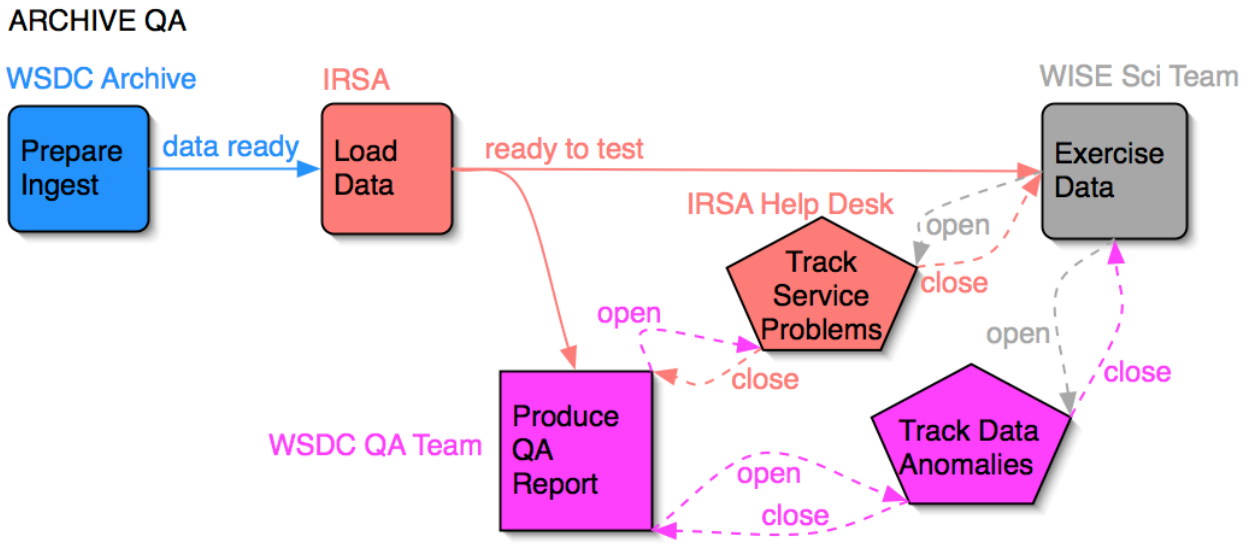

Figure 2. The detailed steps on how the WISE archive quality control and assurance is performed. We note that unless any problem is identified, the newly loaded data will be accessible by the science team as well as the WISE QA team.

3 TIMING REQUIREMENT FOR ARCHIVE LOAD

Because the archive needs to load twice a week, we need to make sure that for each loading, we can complete all tasks within a turn-around time of 3 days. Here is the summary of the time estimates for each task shown in Figure 1.

The data loading cycle starts at the time when the raw data has been processed by the pipeline.

1. The WISE QA team validates each batch of the WISE pipeline processing. Therefore, each WISE data load will be initiated by the WISE QA team. Specifically, this entails an email from the QA team to the WISE OPS and archive team that the processed data is ready to be loaded. Because the data are loaded into the IRSA archive twice a week, the QA team will initiate the loading every week on Monday and Thursday. Every Monday morning, there should be a message from the QA team to the OPS team to specify if the data is ready to go.
2. The second step is the actions taken by the WISE OPS team. Once the OPS gets a to-go signal from the QA team, they will run DBPREP and IMGPREP to prepare the data and stage at the specific places for the IRSA team to take over the next loading tasks. This step by the OPS team should take less than 5 hours. From the time when the OPS gets the message from the QA team, to

the time the OPS team finishing the preparation of the data for loading and sending the message to the IRSA team, it should be within 1 working day.

3. The WISE OPS will make a detailed weekly operation schedule, and the archive team will make sure that all of the required preparation work for each loading is planned in a timely fashion and can be executed on time.
4. IRSA database (GATOR) loading time: We have tested the speed of database loading. To load 66 million rows of source tables, it took only 4 hours to complete. This size table corresponds to 30 orbits of single frame data, about 2 days worth of data.
5. Fits image transfer: Setting a data transfer rate of 18MB/second, it takes 4hours to transfer 266GB of fits images from the WISE computer servers to the IRSA computer disks. This data volume is roughly two days data from the normal operation.
6. Finally, the IRSA team will release the newly loaded dataset on Monday and Wednesday each week.

5 TIME LINE FOR ARCHIVE LOADING AFTER THE WISE LAUNCH

Currently, the WISE is scheduled to be launched on December 7, 2009. The telescope cover will be lifted off on Dec. 23rd, so the telescope will start taking the real data after that. The mission operation is scheduled to start on Jan. 7, 2010. We expect to have 1 week to adjust the pipeline parameters, thus starting on Jan. 15th, there should be processed data for the archive loading. Between Jan. 15th and Feb. 28th, all of the processed data is regarded as preliminary processed data. During these 6 weeks, we plan to do the usual twice loading per week. Starting on March 1st, re-processing starts and this version of processed data is expected to ultimately replace the previously archived data. Therefore, after March 1st , we will need to have one extra loading per week in order to archive the reprocessed data taken during the first 6 weeks.