

A. APPENDIX 7 – DRAFT NGSS PROJECT DATA MANAGEMENT PLAN

A.1 Purpose and Scope

This Project Data Management Plan (DMP) describes project-level policies, schedules and plans for handling, processing, distribution and archiving of the science data for the Next Generation Sky Survey. The NGSS mission is described in section A7.2, and science data processing, archiving and distribution is described in section A.7.3.

The scope of this document covers the ground data system as it applies to the mission science data, from downlink to final archiving. This document does not apply to operating the mission or spacecraft, which is managed separately by the NGSS project. The PDMP identifies the primary and supplementary science data products of NGSS, systems associated with data handling, and the roles and responsibilities, and operational interfaces affecting those systems.

A7.2 NGSS Science Overview

NGSS will map the full sky with unprecedented sensitivity in four mid-infrared wavelength bands from 3.5 to 23 microns. The primary scientific objects of the survey are to study the nature and evolutionary history of the luminous infrared galaxies, and to measure the space density, mass function and formation history of brown dwarfs in the solar neighborhood. NGSS will produce the data set that will allow the identification of the most luminous galaxy in the universe, and the closest star to the Solar System. Additional scientific objectives include measuring radiometric distances and albedos for nearly all known asteroids, the density of old stars in galaxies at redshift $z=1$, the luminosity function of faint protostars in nearby star formation regions, and the evolution of circumstellar disks. All of these objectives address directly the objectives of the NASA Space Science Enterprise Strategic Plan.

A7.3 Roles and Responsibilities

IPAC - Science Data Processing and Archive Center

IPAC is responsible for NGSS science data processing, distribution and archiving. IPAC works with the PI and Science Team to develop algorithms for the data processing pipeline, metrics for science data quality verification, and to develop an effective archive that is accessible to the NGSS team and to the scientific user community. IPAC develops a long-term archive plan for the Level 0 science data and the main science data products from the mission. The plan describes the contents of the archives, the size of the archives, and the schedule for delivery. The archive plan is to be based on the services and policies of the Infrared Science Archive (IPAC), NASA's designated archive center for infrared and sub-millimeter astronomy missions.

White Sands Complex (Ground tracking station)

The White Sands Complex (WCS) is responsible for receiving spacecraft engineering and science data telemetry, validating downlink content and integrity, and transfer of the engineering and Level 0 science data to other Ground Data System facilities (SOC, MOC, SDL, IPAC). WCS also maintains an interim science data archive for retransmission to IPAC in the event of tape failure.

General Dynamics Spaceplex (Ground station data management)

General Dynamics Spaceplex (GDS) is responsible for managing the data holdings at WCS. GDS notifies WCS to either release science data from the interim archive or to resend data tapes after receiving from IPAC the results of science data tape validation.

Principal Investigator and Science Team

The NGSS PI and Science Team work with IPAC to develop efficient algorithms for data processing, and metrics and tests for science data quality. The PI and Science Team work with IPAC to validate NGSS Data Products during the data processing stage, and in

preparation for data product releases. The PI and Team also contribute to the preparation for the mission explanatory documentation.

NGSS Project Office (JPL)

[Valerie – need to add appropriate words about Project responsibilities here]

Responsible for ensuring policies and implementation plans for data releases. Coordinate with NASA headquarters, making sure HQ personnel are informed of pending NGSS information and data releases.

[Questions - Do any other partners have responsibility? BATC – maintain engineering data archive? SDL – Provide instrument data from ground test to IPAC for development of processing software?]

A7.4 Data Flow

The NGSS end-to-end data flow is described in section F.7.3 and Figure FO4-1. In this section, we summarize the flow of science data from downlink to final archiving.

Compressed, encoded science and engineering data are transmitted to TDRSS for relay to the White Sands Complex (WCS). The ground system at WCS unpacks and decodes the science data to Level 0 status, and writes them to digital linear tape (DLT). WCS validates the integrity of the telemetry and reports the status to SOC, MOC and IPAC. Spacecraft and instrument engineering data are routed to BATC for health monitoring and mission assurance. A subset of the engineering data is made available via ftp to SDL, UCLA and IPAC for spacecraft and instrument trending analysis and integration into science data processing.

WCS stores the science data in an interim archive for up to 30 days in the event a retransmission to IPAC is necessary.

Science data tapes are shipped from White Sands to IPAC twice per week. IPAC reads the data tapes, validates contents and reports the status of the validation to GDS so that that the data can be released from the WCS interim archive.

IPAC conducts *preliminary processing* of the Level 0 science data shortly following their arrival via tape from WCS. Processing is

performed using preliminary instrumental, photometric and astrometric calibration parameters, derived during ground test and the in-orbit validation period. The preliminary Image Atlas and Source Catalog are generated from this preliminary processing.

All science data gathered during on-orbit mission operations undergoes *final data processing* following the end of on-orbit mission operations and the release of the preliminary data products. Final processing incorporates optimal parameters derived from analysis of spacecraft, instrument and preliminary processing results over the life of the mission.

Science data processing, quality assurance and archiving for NGSS is conducted using a highly automated software system modeled closely on the successful 2MASS Automated Processing Pipeline (2MAPPS). The basic processing steps are:

- Unpack and decompress the science data from tape, validate contents, and prepare for input to the data pipeline
- Combine time stamped engineering data with individual science data frames
- Subtract dark response, linearize and remove the flat field response for each image
- Identify and tag cosmic rays, artifacts and bad pixels
- Combine images from adjacent orbits
- Characterize the PSFs on the combined images, detect sources, and characterize using PSF-fitting algorithms
- Perform position reconstruction using 2MASS sources identified in each frame
- “Bandmerge” extractions from individual bands into single sources and measure upper limits for non-detected bands
- Identify and tag suspected artifacts among the extracted sources
- Perform flux calibration
- Perform quality analysis for Image and extracted Source data
- Report results of science data quality analysis to Science Operations at UCLA for the purpose of tracking mission sky coverage

The calibrated images and source extractions produced from pipeline data processing are loaded into an image archive and working source database. These archives are built within the framework of the Infrared Science Archive (IRSA) at IPAC so that the NGSS Science and Team and partners can access the data using the extensive image visualization and database query tools already developed by IRSA.

The NGSS Image Atlas and Source Catalog are derived from the pipeline-processed data contained in the working archives. After validation and review, these products are served to the astronomical community via the IRSA interfaces.

A7.5 Data Products

IPAC works with the NGSS PI and Science Team to produce two primary data products that enable the key scientific investigations that embody the mission objectives. The products are an *Image Atlas* and an extracted *Source Catalog*. An Explanatory Supplement describing the NGSS mission, instrument and the data products will accompany the release of the Catalog and Atlas.

Image Atlas

The NGSS Image Atlas (section E.4.3.3) consists of photometrically calibrated, position-tagged FITS images in the four bands. The images in the four bands are resampled to 1.0"/pixel and registered using the same coordinate grid for each band. If the mission lifetime is extended in Phase F, a separate Atlas will be made from each six month sky coverage. The pixels from each of the >5 frames will be combined to produce the best estimate of the flux at each position on the sky.

Source Catalog

The NGSS Source Catalog contains accurate equatorial positions and flux densities (or upper limits) measured in the four survey bands. Sources are detected and measured on the combined frames from adjacent orbits for maximum sensitivity, using a detection threshold that scales with confusion noise. A point-spread-function (PSF) is used to estimate positions and intensities on the combined images. The requirement for photometric

accuracy for the NGSS source Catalog is 7%. Absolute positions of high SNR NGSS sources are reconstructed relative to the 2MASS astrometric reference frame, and will be accurate to ~0.5" RMS with respect to that frame. Lower SNR sources follow the full-width at half-maximum/SNR growth curve.

If the NGSS mission is extended in Phase F, the Source Catalog will contain positions and brightness estimates for both sky coverage epochs, with estimates of flux variation and proper motion.

Ancillary Data Products

NGSS will also produce such ancillary products as are necessary to capture spacecraft, instrument or science data quality as are necessary for science data processing and scientific interpretation. These ancillary products are most likely tabular in form, and may capture large segments of engineering data.

A7.6 Schedule and Latency

Data Product Release

The objective of the NGSS mission is to make the science data products available to the community as rapidly as possible. To facilitate this, data release takes place in two stages. The first release occurs six months after the end of nominal on-orbit lifetime. The first NGSS data release consists of a preliminary Image Atlas and Source Catalog containing source with SNR>20 in unconfused regions of the sky. The first release is drawn from the preliminary processing of the science data made while the spacecraft is operational. The final NGSS data release, made 17 months after the end of on-orbit operations, contains the final Image Atlas and Source Catalog containing sources as faint as SNR~5, and is accompanied by more extensive quality analysis, validation and documentation.

Science Data Processing

IPAC conducts *preliminary processing* of the Level 0 Science Data within one week of their arrival via tape from the ground station. Processing is performed during on-orbit mission operations using an automated software pipeline that uses preliminary instrumental, photometric and astrometric calibration parameters, derived during ground test and the in-orbit validation

period. The preliminary Image Atlas and Source Catalog are generated from this preliminary processing.

All science data gathered during on-orbit mission operations undergoes *final data processing and validation* following the end of on-orbit mission operations and the release of the preliminary data products. Final processing incorporates optimal parameters derived from analysis of spacecraft, instrument and preliminary processing results over the life of the mission.

The NGSS science data processing and archive system at IPAC is sized to be able to keep up with the data acquisition rate. The science data received twice per week from White Sands is processing, validated and loaded into the working archives within one week.

Tape Validation

Within one week of their arrival at IPAC, IPAC validates the integrity of the Level 0 Science Data tapes sent by the White Sands Complex, and compares their contents with that reported by the ground station. If a data tape is unreadable, contains errors in the read validation, or does not contain the data specified in the downlink report, IPAC requests a resend of the data from White Sands.

Science Data Transfer

Level 0 science data are written to DLT tapes once per day at the White Sands Complex. The tapes are shipped to IPAC twice per week via commercial overnight service. The Level 0 science data will be archived for 30 days at the White Sands in an interim storage facility so that data can be resent in the event of tape failure. Science data are released from the interim archive upon confirmation of their receipt and validation of tape integrity and contents by IPAC.

Downlink and Verification

GDS validates the integrity and content of each downlink at White Sands **within 24 hours (TBD)** of the contact. The results of that validation, and the content of the downlink are reported to the NGSS SOC (UCLA), MOC (BATC) and Science Data Analysis Center (IPAC) within **36 hours (TBD)** of the downlink contact.

A7.7 Data Distribution

Data Access Policies

NGSS Data Products are released to the astronomical community and general public as rapidly as is technically feasible and scientifically sensible. There is no proprietary period on NGSS data, but the mission Data Products are not released to the archive for distribution to the community until they have undergone thorough Quality Assurance analysis by the PI and Science Team. During preliminary and final data processing and the data product validation period, the PI and Science Team have unrestricted access to the working archives for the purpose of assisting with the validation of spacecraft, instrument and data processing system performance.

[Questions – Who manages data access permission? Who approves data releases?]

Data Access Modes

Although daunting in size, the NGSS Data Products are comparable to the size of the 2MASS Image Atlas and Source Catalogs. Thus, plans for Science Team and end-user access to NGSS data follow the successful model used to enable access to the data products of 2MASS.

Primary access to the ~10 TB calibrated Image Atlas and ~1 TB Source Catalog is via the on-line services of IRSA at IPAC. These services currently include:

- GATOR – a web-based interface that enables complex, efficient queries of massive Catalogs;
- Interactive and Batch Image archive access tools that allow position, region or source-based image query requests;
- OASIS - a Java-based service that allows image/catalog interaction and for a wide variety of Image and Catalog data sets currently served by NASA and international data centers, not limited to IRSA.

IRSA has a number of additional services in development, under the auspices of the National Virtual Observatory, that will be deployed long before NGSS launches include. These services

Second Heading *First Heading*

emphasize intercomparison of massive data archives held in different locations.

NGSS science data pipeline processing yields calibrated Images and Source Extractions that are loaded into working databases built within the framework of IRSA at IPAC. The NGSS Science Team and project staff access these working archives using the same on-line IRSA services. Team access to the working archives and preliminary products is restricted using password protection on the IRSA sites.

The NGSS Source Catalog will also be distributed on DVD, or equivalent media, for bulk distribution. The bulk Catalog distribution will follow the on-line IRSA distribution by several months because of the long production time for commercial DVD vendors.

Interconnectivity

Because the NGSS working archives and release data products are developed with data distribution and archiving in mind, all necessary metadata and other descriptors will be in place to insure compatibility with the existing IRSA tools. This development philosophy that NGSS data products are fully interoperable and extensible through the IRSA infrastructure, with related data centers and services that use XML, SOAP, WSDL and UDDI protocols being developed under the NVO initiatives sponsored by NASA and NSF.

A7.8 Archive Strategy

Long Term Data Access

The NGSS data products will be “delivered” to IRSA, the NASA designated archive center for infrared and submillimeter astronomy missions. Upon completion of NGSS Phase E, IRSA will assume curator duties for the NGSS data products and Explanatory Documentation. Therefore, continued access to the NGSS data products for the astronomical community is assured with the continued operation of IRSA.

Deep Archive

Through the NGSS Phase E period, one copy of the repacked Level 0 raw data tapes is stored at IPAC, and one copy is stored in a commercial off-site secure data storage facility in the Los Angeles area. The commercial facility used will be the one selected by the SIRTf Science Center to archive SIRTf data tapes beginning in FY03.

At the completion of NGSS Phase E, one copy of the repacked Level 0 data tapes, and one copy of the processed data archive tapes containing pipeline output images and source files will be sent to the National Space Science Data Center for long-term deep archiving.