Data Processing Plan for SOFIA Science Instrument

THIS CONTRACT DOCUMENT IS UNDER NASA CONFIGURATION CONTROL

Date: June 07, 2011
Revision: A

DFRC
Dryden Flight Research Center
Edwards, CA 93523

ARC
Ames Research Center
Moffett Field, CA 94035

German Space Agency, DLR
Deutsches Zentrum für Luft
und Raumfahrt

Data Processing Plan for SOFIA Science Instrument

This Contract Document is Under NASA Configuration Control

NASA APPROVALS:

Pamela M Marcum	6/13/11
Pamela Marcum / ARC NASA / SOFIA Project Scientist	Date
Gelsomina Cappuccio	6/10/11
Gelsomina Cappuccio / ARC NASA / SOFIA Science Project Manager	Date

USRA

UNIVERSITIES SPACE RESEARCH ASSOCIATION

STRATOSPHERIC OBSERVATORY FOR INFRARED ASTRONOMY
NASA AMES RESEARCH CENTER • MS N211-3 • MOFFETT FIELD, CA 94035
• 650-604-1984 (fax) • WWW.SOFIA.USRA.EDU •

7 June 2011
11-HJH-2496

Mr. Donald J. Nickison
NASA Ames Research Center
Mail Stop N211-3
Moffett Field, CA 94035-1000

Dear Mr. Nickison:

In accordance with the Document Requirements of Contract NAS2-97001, Stratospheric Observatory for Infrared Astronomy, Universities Space Research Association is pleased to provide SCI-US-PLA-PM17-2010, Revision A, the Data Processing Plan for SOFIA Science Instruments, a type 1 document.

If you would like to discuss any of this material, please do not hesitate to call.

Respectfully,

Helen J. Hall

Helen J. Hall
USRA SOFIA Associate Director for Program Management

HJH/rsm

Data Processing Plan for SOFIA Science Instruments

SCI-US-PLA-PM17-2010

Date: May 16, 2011
Revision: A

DFRC
Dryden Flight Research Center
Edwards, CA 93523

ARC
Ames Research Center
Moffett Field, CA 94035

German Space Agency, DLR
Deutsches Zentrum für Luft und Raumfahrt

Data Processing Plan for SOFIA Science Instruments
SCI-US-PLA-PM17-2010, Rev. A

AUTHOR:

William Reach, USRA SOFIA Associate Director for Science — Date 5/26/11

APPROVAL:

Helen J. Hall, USRA SOFIA Associate Director for Program Management — Date 6-7-11

REVISION HISTORY

REV	DATE	DESCRIPTION
-	05-16-2011	Initial Release
A	05-24-2011	Changes made in response to request made at the Science PMB on 5/24/11

Data Processing Plan for SOFIA Science Instruments SCI-US-PLA-PM17-2010, Rev. A

1. OVERVIEW

Science data from all SOFIA observations must be calibrated and stored in the Data Cycle System archive, in order to provide a permanent record of the observation, enable guest investigators to access their data during their (nominal 1 year) period of exclusive access, and to allow future use by other investigators. This Plan applies to Facility-Class instruments, both from initial complement of instrumentation and the instruments (or upgrades) that will be selected through NASA solicitations for Science Instruments. All first-generation Facility-Class as well as all next-generation SOFIA science instrument (SI) teams are required to develop and deliver a data analysis pipeline to SOFIA Science Mission Operations (SSMO). After the instrument is commissioned, during nominal operations, SSMO will run the data analysis pipeline after each flight, perform calibrations, archive both raw and reduced data, and maintain appropriate access rights.

The data analysis pipelines are not expected to run without manual intervention, though all routine steps should be automated to the extent possible, with the ability to script tasks that are eventually well-enough understood that they can be automated. An Instrument Scientist (primary and deputy) will be assigned to each selected instrument during development; these Instrument Scientists will be responsible for interfacing between the SI team and SSMO, including arranging for pipeline installation and learning the pipeline operation. During commissioning flights, the pipeline must be demonstrated. Acceptance of a facility instrument will depend upon successful demonstration of pipeline validity and operability. After acceptance the pipeline Data Processing Engineers and Scientists at SSMO will operate the pipeline, calibrate the observations, assess data quality, and release data through the archive.

This document describes the overall requirements for data processing of science instrument data. It is accompanied by the following documents:

- Interface Control Document: DCS-SI-01 (Includes DCS Archive Keyword Dictionary)
- Project Data Management Plan: USRA-DAL-1162-00
- SOFIA Science Project Software Management Plan: SCI-AR-PLA-PM20-2004

2. DESCRIPTION OF DATA PRODUCTS

Four levels of data products are recognized by SSMO:

- Level 1 data are essentially raw from the instrument but in standardized format (FITS or CLASS),
- Level 2 data have been corrected for instrument artifacts (e.g., flats, darks, bad pixels),
- Level 3 data have been flux calibrated (e.g. BUNIT keyword takes image to MJy/sr),
- Level 4 data are high-order products possibly combining multiple exposures (e.g. mosaics, spectral cubes)

The software and detailed procedures for generating Levels 2-3 must be delivered to SSMO and validated as part of the instrument commissioning review. All data products of any level must adhere to the Keyword Dictionary and file naming conventions specified in the DCS_SI_01.

2.1. Level 1

Level 1 products must be produced directly on-board as part of instrument operations (SOF-AR-SPE-SE01-2008, SOFIA Science Instrument Specification, ¶3.1.2). The Level 1 products will be ingested into the Data Cycle System at SSMO and will become part of the permanent archive. Thus the files must be fully self-documented in terms of the method of acquisition, instrument, and telescope configuration. The requirements on header keywords are specified in the DCS-SI-01 ICD.

2.2. Level 2

The pipeline shall produce Level 2 products and be integrated into the Data Cycle System at SSMO. The Level 2 pipeline will either remove artifacts or place sufficient standard information into the data header or extensions such that standard astronomical software can remove the artifacts. Ancillary products used in the data processing are considered part of the archive and must be included as part of a Level 2 pipeline processing via header keywords, image extensions, or pointers (header keywords) to ancillary files. A method for validating Level 2 products must be delivered both as part of validating the pipeline (this validation will be performed by SSMO before instrument commissioning as part of the instrument acceptance criteria) and as a procedure for operations.

Imaging pipelines: The Level 2 pipeline will demodulate the chops and nods (as needed) such that a proper image difference is obtained. Astrometric refinement and simple image distortions shall be dealt with by including appropriate FITS-standard header keywords or data structures; bad or otherwise flagged pixels shall be identified via a mask. Image artifacts such as latent images or other effects that are not cleanly removed by the pipeline shall be identified via a mask with values defined in the header. The actual image shall be corrected for all well-understood artifacts, whether derived from calibration files or from other observations that are part of the same night's observing. For imaging pipelines, a general guideline is that the result must be capable of being displayed using a standard visualization tool such as ds9, and manipulated and mosaiced using a standard image analysis tool such as iraf or idlastro.

Spectroscopic pipelines: An accurate wavelength or frequency calibration must be derived and included for each observation. For one-dimensional (e.g. heterodyne) spectra, the frequency calibration shall be included in the header via standard keywords or supplied in an extension. The data format shall be either standard FITS format or in a form that can be directly imported into GILDAS/CLASS such that the frequency, calibration, and pointing information can be used as per standard practice in coaddition and generation of spectral grids. For two-dimensional spectral data files (e.g. echelle), the level 2 data product must be accompanied by images or an algorithm that specifies the wavelength and spatial location of each pixel; if those calibrations are the same for all images then they can be supplied as calibration products into the archive, with a header keyword clearly specifying which calibration product applies. The basic Level 2 product will have the same pixelization (for imaging spectrographs) as the original data, with the

pixel locations fully specified. In order to produce one-dimensional spectra for scientific analysis, slit spectrographs must produce either or both of the following: a tool for making spectral extractions from the Level 2 products, or a Level 2.5 product, specifically, a rectified, two-dimensional, spectral image, with wavelength and spatial distortions removed.

2.3. Level 3

For Level 3 flux calibration, the SI teams must provide both the pipeline and a calibration parameter file. The calibration parameters will be determined from processing and analysis of the calibration sources and then applied to the appropriate science data and subsequently stored in the Archive.
Level 3 products will require association with absolute calibration observations from the same flight, and will require validation. The prescription for selection of the calibrator shall be specified in a Calibration Plan, and the data analysis procedure shall include rules for associating calibration files with their science observations, via header keywords. The calibration database and association procedure will be worked out with the Data Cycle System team at SSMO before the Acceptance Review for the instrument.

2.4. Level 4

Level 4 products are highly desired, but in cases where they are readily generated by existing tools (e.g. iraf), such products are not required and instead a recommended procedure should be provided. Emphasis is to be placed on instrument or SOFIA-specific processing tools such as corrections for instrument or atmospheric instabilities. It is likely that such tools will be interactive. Any SOFIA-specific tools required to analyze the data must be provided as part of the data processing software delivery. The goal is that the Level 3 data are of sufficiently high quality and usability that they can serve as the basic archival data files.

2.5. Calibration Data

In addition to the pipeline that analyzes science data, the procedure or pipeline required to analyze calibration data must also be provided. This includes for example generation of on-board per-flight or per-flight-series calibrations like darks, flats, bad pixel masks, wavelength calibrations. Calibration procedures and pipelines must be consistent with the data analysis pipeline and interoperable to the extent possible.

3. SOFTWARE CODING STANDARDS

The data analysis pipelines shall be developed in standard and common coding languages. Approved languages include c, c++, idl, perl, and python. Only commonly-available and maintained libraries (such as fitsio, wcslib, idlastro, scipy, numpy) are to be used.

4. SOFTWARE DEVELOPMENT PLAN

There will be two development phases for the data processing pipeline. During the first phase, the Science Instrument team will be in active development and responsible for configuration control and planning at the “home” institute. Intermediate releases will be made to SOFIA

Science Mission Operations, for incorporation into the Data Cycle System both in terms of data products (for the archive and database), and for executive functionality. Interface Control Documents will be developed for each SI to ensure clarity and ensure compliance. The SI pipeline generates Level 2 products from the raw data, and Level 3 (calibrated) data from the Level 2 products.

Each Science Instrument team shall schedule a preliminary delivery, of scope adequate to test interfaces, at least 6 months before instrument delivery. The delivery must include sample data products at each level (including calibration products), and a version of the processing code that is at least skeletal, such that a version of the processing code (potentially with limited functionality) can be interfaced with the DCS pipeline executive and data handling.

- The data products shall comply with FITS standards and the SOFIA keywords as defined in the Data Cycle System Keyword Dictionary and the DCS-SI Interface Control Document.
- The Level 2 and Level 3 products are to be included in this delivery.
- Ancillary files such as bad pixel masks, darks, etc. (and the method for generating those files) shall be included.
- The calibration procedure shall be outlined, including processing steps and method for incorporating the results into the Level 3 products. (The astronomical aspects of calibration are to be covered in a separate Calibration Plan.) The Level 3 products are to be in physical units (MJy/sr for specific surface brightness, Jy or Kelvin for in-beam spectra) with the BUNIT keyword set appropriately and enough information in the header to trace back to the calibration products used.
- For images, the Level 2 (and 3) products shall have a World Coordinate System (WCS) with the center of the image specified in CRVAL keywords, the rotation in CROT, and scale in CDELT (or the CD matrix) as per FITS standards. The WCS shall be derived from the telescope control system using keywords provided on-board by the MCCS. Distortion shall be dealt with as per FITS standard (such as the TAN-SIP keywords that allow second-order distortion); more complicated distortion shall be documented fully, with a method for determining the location of the center of each pixel. Level 2 data products should have the same pixels as the original data; however an additional distortion corrected product can be included as an extra result from Level 2.
- A clear specification of the pipeline method for coadding, and combining chop-nods shall be provided.
- For imaging (including slit) spectrographs, a complete specification of the wavelength of each pixel shall be provided either within the FITS header, in an extension, or in an ancillary product that is specified in the header of the Level 2 product.
- For single-element spectrographs, the data products shall be FITS or CLASS format together with a complete specification for determining the frequency of each spectral element.

The second phase for pipeline development marks a transition in configuration control from that of the SI development institute to that of the SOFIA project. At least 3 months before the acceptance review for the instrument, a full delivery shall be made, including:

- Requirements or Specification Doc
- source code
- Software Version Description Doc (includes build and deploy instructions and list of known issues)
- SI-DCS ICD updates (including pipeline and calibration specification files per DCS_SI_01)
- Users Guide,
- Test plan,
- and other supporting documentation.

The Users Guide shall be targeted at the Science Mission Operations Center staff, who will be operating the pipelines after acceptance of the instrument. The Instrument Scientist assigned to the each SI will assist in producing the Users Guide. Either in this document or in a separate Manual, the software requirements and compilation procedure shall be specified.

The specific SI pipeline Test Plan shall include test cases that address both pipeline performance and scientific validity, as outlined in the general SOFIA Pipeline V&V Plan (in work). Incremental deliveries will be made during the commissioning process. After acceptance of the instrument, the software will be configuration controlled as described in the SOFIA Science Project Software Management Plan.