**1. INTRODUCTION**

SOFIA is an airborne telescope for infrared and sub-millimeter astronomy. Its design combines the challenges of ground and space observatories. Contrary to satellite observatories, SOFIA hosts a large variety of instruments observing in wavelength ranges from 0.3 to 600 microns which will be upgraded over time. This will produce a large diversity of data types which will likely increase as new generations of instruments are operated. The SOFIA Data Cycle System (DCS) will support, for both facility and visitor instruments, all aspects of data processing including archiving and pipelining of data of level 1, (raw), 2 (processed-instrumental and background removal), 3 (flux calibrated), and 4 (higher processing- mosaicing and source catalogs).

The DCS will provide uniform, extendable and supportable framework for all aspects of the data cycle.

Data processing includes all steps required to obtain good quality flux calibrated data of spectroscopy, imaging, fast-acquisition, polarimetry, etc. Processing each data type requires a sequence of unique or common algorithms. Processing each data type requires a sequence of unique or common algorithms that are developed in various instruments and distributed in a variety of environments. In addition, these algorithms may require user-interaction or fine tuning of input parameters in order to return good quality data.

**2. CONCEPTS AND ASSOCIATIONS**

The Astronomical Observation Request (AOR) is the link between scientific and calibration data of the same observation type. It defines the parameters necessary for the observation and post-processing. Therefore, it will identify the reduction pipeline and its parameters. For each level 2 product, the Pipeline Pedigree (PP) records the pipeline generating the data, the parameters, the processing date and the data involved in the process. AOR and PP concepts have been implemented and are operative in DCS. A similar concept will be necessary to track calibration activities. DCS will include a Flux Calibration Parameter (FCP) which will support the calibration process in order to document and reproduce the same results as needed. AOR, PP and FCP are characterized by unique key numbers that identify them as well as information about the data involved in the process.

**3. ARCHITECTURE**

The Challenge of Data Reduction for Multiple Instruments on SOFIA

**3.1 PIPELINING APPROACHES**

A pipeline is a collection of algorithms which are run in a particular order. The DCS will store pipelines coded in IDL, Python and C++ which are delivered by the instrument teams in association with a XML file describing the pipeline recipes. With the appropriate pipeline specification, DCS can currently run pipelines in any language with no modification of the code as soon as the pipeline is delivered as an executable, likely the same that runs in SMO machines outside DCS. Because DCS does not have a knowledge of the details of the pipeline execution after it is called, we name this pipeline “blackbox”. Level 2 blackboxes are applied based on the specifications of the AOR - which is detailed before the flight as part of the observation planning process – and the details of the process are recorded on the PP - which is created after peeling – as explained in section 2. This approach is currently implemented in the DCS and embraces automatic pipelining at EoF and manually pipelining initiation within the same framework. This answers the need for re-pipelining with the goal of improving the quality of the Level 2 data by fine-tuning pipeline parameters after manual inspection or applying an improved version of the pipeline. Although, this approach represents an important effort saving on the implementation and maintenance of the pipelines it lacks of the advanced functionalities that the DCS could offer including parallel execution of processes of a single pipeline, status report, and intermediate user intervention.

We plan to implement the current functionality with another approach allowing human intervention. User interaction is required for step-by-step data processing and intermediate data analysis. These will be performed using a DCS graphical interface tool which runs user-data interaction process and analysis tools locally (outside DCS) after downloading updated algorithms from DCS. As a long term goal, DCS will integrate user-pipelining within the same pipeline as automatic pipelining. For that purpose, pipelines will be delivered as a collection of functions (modules) performing a portion of the pipeline and XML files describing them. The pipeline recipe (another XML file) will describe how modules are executed, the order of execution and how data is transferred between modules. Technically, the pipeline manager objects (pipe_man) are in charge of executing specific modules (module->process method) or the whole pipeline (pipe_man->run method). This new approach fits in the actual black box structure by calling run method as the pipeline executable. When implemented within DCS, pipe_man will be able to process modules in parallel, control their execution, and allow user data analysis. In addition, pipe_man will manage modules in different computer languages for the same pipeline thus, reducing the number of algorithms in the system. Instrument teams will be encouraged to use existing algorithms when developing their pipelines, resulting in a common library of algorithms which will decrease the efforts of the instrument teams for developing pipelines and of the DCS team for maintaining and upgrading them.