The SOFIA Data Center (SDC)

Bernhard Schulz, Aaron Bryant, Jonas Früh, Benjamin Greiner, Michael Hütwohl, Christof Iserlohe, Thomas Keilig, Alfred Krabbe, Bastian Knieling, Karsten Schindler, Manuel Wiedemann, Oliver Zeile

SOFIA Data Center (SDC) / Institut für Raumfahrtsysteme (IRS), University of Stuttgart, Germany

Abstract

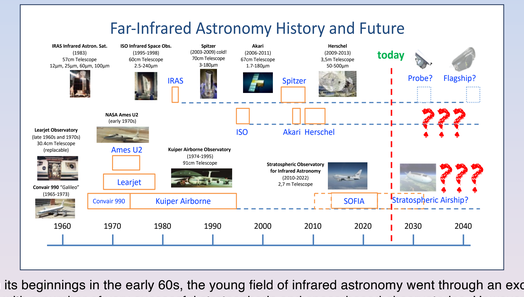
During 783 scientific flights, SOFIA, the Stratospheric Observatory for Infrared Astronomy of the space agencies of Germany and the United States, DLR and NASA, collected a considerable volume of scientific data, that is available through the Infrared Science Archive at IPAC. After the end of flight operations in September 2022, with only one year of ramp-down time for Science Mission Operations remaining, only a limited data reprocessing of SOFIA Observing Cycles 5 to 9 (2017 to 2022) could be achieved. In order to complete the job and ensure a full exploitation of the considerable investment in SOFIA, the SOFIA Data Center (SDC) was funded through DLR for a five-year post operational phase starting in July 2024.

The SDC aims to eventually provide reprocessed and improved Far-Infrared SOFIA science data in a VO-compatible astronomical archive and the technical and operational engineering data of the SOFIA telescope in the official data repository of the University of Stuttgart, DaRUS. We expect to work with the entire astronomical community to promote the exploitation and publication of valuable Mid- and Far-Infrared data, establish a powerful scientific resource for the post-mission phase, and bridge the gap until the next space- or stratospheric FIR observatory becomes available.

To this end the current focus is on several initiatives that include updating the Python based instrument pipelines (Redux), building of a new pipeline for converting and plate solving of the telescope's guide camera images for pointing reconstruction and scientific exploitation, and preparing a new kind of calibration for atmospheric water vapor for every flight. The FIFI-LS instrument pipeline, best known to the team, will serve as a trailblazer for data of the other three instruments EXES, FORCAST, and HAWC+, that the SDC will focus on, while the GREAT data is being reprocessed by the University of Cologne. An assessment of all the available data is under way as part of the preparation for ingesting both, raw- and processed data into the DaCHS VO-publishing infrastructure. To implement the SDC web portal for the astronomical community, we are planning to use Firefly, that also powers IRSA, to present the data optimally tailored and re-ordered by astronomical needs.

This poster provides the overview of the SDC's activities with some aspects being detailed in additional presentations at this conference. Once the 5-year operational phase of the SDC is complete, the archive will be transferred to the new German Astrophysics Center (DZA) as a long-term home.

Motivation



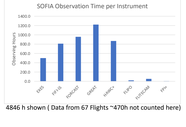
With its beginnings in the early 60s, the young field of infrared astronomy went through an exciting time with a number of very successful stratospheric and space-based observatories. However, the premature end of SOFIA, the last regularly scheduled observatory sensitive to the FIR, after less than half of its design lifetime of 20 years, leaves the field with an uncertain future. The earliest realistic chance for an infrared space mission is the NASA Probe mission, that is currently being competed for by an infrared and an X-ray mission.

The SOFIA Data Vault

Daten

Data Volume in DCS

	LEVEL_0	LEVEL_1	LEVEL_2	LEVEL_3	LEVEL_4	Total
EXES	46.6	2013.9	819.6	43.7	0.0	**2923.7**
FIFI-LS	0.0	972.4	21.1	125.7	35.9	**1155.1**
FLITECAM	0.3	31.1	46.8	38.2	0.0	**116.3**
FORCAST	5.6	94.3	549.7	148.0	1.2	**797.8**
FPI_PLUS	0.0	22.4	43.4	0.3	1.1	**67.3**
GREAT	0.0	2494.6	0.0	3056.7	75.7	**5627.0**
HAWC_PLUS	13334.1	0.7	18.5	10.1	4.9	**13368.3**
HIPO-BLUE	0.0	46.8	0.0	0.0	0.0	**46.8**
HIPO-RED	0.0	74.8	0.0	0.0	0.0	**74.8**
Total	**13386.5**	**5755.1**	**1498.0**	**3422.7**	**118.8**	**24177.0**



- SOFIA Raw Data ~141 TB
- Science Data ~24 TB
- Total about **5300 h** (SLT)
- Telescope guider cameras FPI+, WFI und FFI were always active (**>5300 h**).

4846 h shown (Data from 67 Flights ~470h not counted here)

SOFIA's eight science instruments collected a considerable amount of data during her operational lifetime. The data comprises the instrument science data, instrument and observatory housekeeping data and all images taken in parallel by the three telescope guider cameras. After observing flights ceased, there was only limited time and funding to perform data reprocessing. These activities focused on Cycles 5 - 9 and ended after one year.

Aims of the SOFIA Data Center

- Creating a powerful scientific resource for the time after the mission
- Bridging the expected 10-20 year gap in regular new FIR data availability
- Ensuring the quality of the SOFIA data archive and its documentation to secure its scientific productivity for the long term
- Enabling future scientific data exploitation without the need for expert knowledge of SOFIA subtleties
- Completing the work within 5 years with subsequent transfer to DZA
- Staying in sync with the current SOFIA data repository hosted within IRSA

SOFIA during flight testing with chase plane (Credit: NASA/DLR)

Redux Status: SOFIA Instrument Pipelines

- Internal development at GitHub Enterprise Instance of Uni Stuttgart
  - Guest accounts for outside collaborators available on request.
- Public at: https://github.com/SOFIA-Data-Center/sofia_redux
- Forked from: https://github.com/SOFIA-USRA/sofia_redux and pushes from the internal repository
- GitHub.com contributions through issues or pull requests welcome and will be synced to internal
- Version releases will update existing PyPI package sofia-redux
  - currently outdated 1.3.3 from 2023
- Unit tests continuously checked with GitHub Actions CI/CD workflow (internal runner)

The data pipelines for the different science instruments are organized under an umbrella called "Redux" that provides a homogeneous user interface, requiring only two inputs, a file list and a configuration file.

This package is essential for both, the scientist doing research and requiring an optimally reduced dataset, and the SDC in its quest to re-reduce the entire SOFIA dataset with various improvements.

References

Young, E. T., Becklin, E. E., Marcum, P.˜M., et al. 2012, ApJL, 749, L17

Roby, W., Wu, X., Ly, L., et al. 2013, Astronomical Data Analysis Software and Systems XXII, 475, 315

Demleitner, M. 2014, Astronomical Data Analysis Software and Systems XXIII, 485, 309

Clarke, M., Vacca, W. D. & Shuping, R.˜Y. 2015, Astronomical Data Analysis Software and Systems XXIV, 495, 355

Pfüller, E., Wolf, J., Wiedemann, M. 2018, Journal of Astronomical Instrumentation, Volume 7, Issue 4, id. 1840006

Fischer, C., Iserlohe, C., Vacca, W., et al. 2021, PASP, 133, 055001

Iserlohe, C., Fischer, C., Vacca, W. D., et al. 2021, PASP, 133, 055002

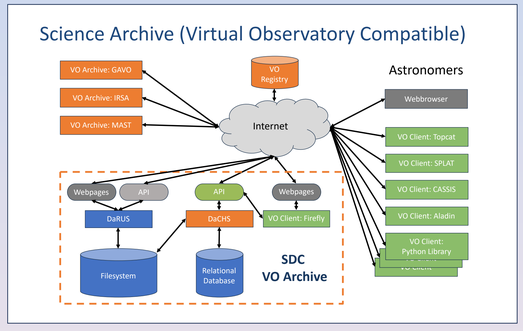
Iserlohe, C., Vacca, W. D., Fischer, C., et al. 2022, PASP, 134, 085001

Vacca, W. D., Iserlohe, C., Shenoy, S., et al. 2023, PASP, 135, 085001

Main Tasks

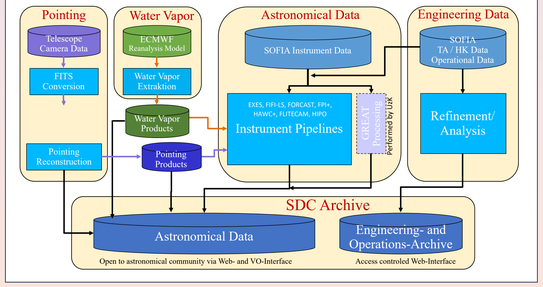
- Update of metadata to enable reprocessing of Cycles 0 - 4
- Create an automatic reprocessing environment allowing for an iterative approach
- Reprocess instrument data (GREAT instrument will be separately done in Cologne)
- Iteratively implement additional pipeline improvements
- Develop supporting software tools
- Create a SOFIA Science Archive hosting raw and processed data
- Reorganize science data rather by astrophysical criteria than operational ones
- Create an Operations- and Engineering-Archive
- Update documentation where necessary
- Support scientists through webinars, conferences and direct advice
- Support engineering research by interested parties

VO-Archive and Data Reprocessing



The SDC Archive is planned to be VO compatible to enable a broad exposure of the data to science searches and minimize the need for new software development through use of available VO-client and server software. DaCHS (Demleitner 2014) will be used to create the VO-archive and provide an API to the internet. The Firefly software suite (Roby et al. 2013) will be used to access this API as a VO-Client and create the archive webpages.

Reprocessing Workflow

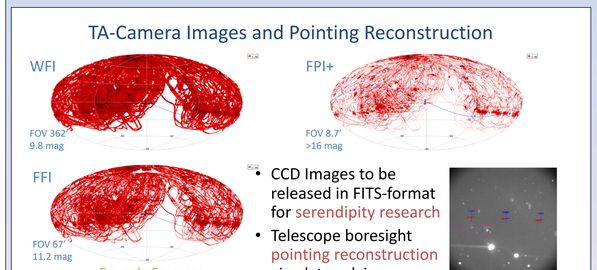


The SDC will build an automatic reprocessing system that can fully process a set of related observations into a final higher-level product without user interaction. It will be using the instrument pipelines originally constructed by the SMO as well as additional new modules that integrate improved pointing information, corrections for water vapor and other improvements into the processing. The necessary pipeline parameters will be stored in the database for each group and can be refined as our understanding of the data improves. The grouping of related observations will be driven by scientific considerations. Groupings that arose only from operational constraints will be eliminated.

Summary

- The SOFIA Data Center (SDC) at the University of Stuttgart will provide an adequate post-processing and archival phase for SOFIA.
- The active phase for the SDC is 5 years.
- Within this period the data will be substantially improved, be placed into a well searchable and documented online-archive, and scientists will receive expert support in understanding residual instrumental effects and finding valuable, yet unpublished data.
- At the end of this period the archive contents are planned to be transferred into a permanent archive at the DZA.

Improved Pointing and Atmospheric Water Vapor Correction

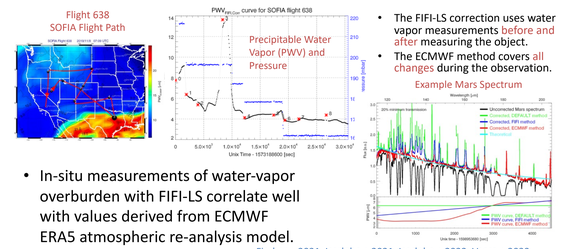


- CCD Images to be released in FITS-format for serendipity research
- Telescope boresight pointing reconstruction via plate solving

Example FPI+ image

The three CCD cameras WFI, FFI, and FPI+ (Pfüller et al. 2018) of the telescope assembly (TA) recorded images through all science flights. The SDC will provide the images with FITS headers, publicly release them and generate detailed telescope pointing reconstruction products for every flight that can be used by the data processing pipelines.

Water Vapor Calibration



- The FIFI-LS correction uses water vapor measurements before and after measuring the object.
- The ECMWF method covers all changes during the observation.
- In-situ measurements of water-vapor overburden with FIFI-LS correlate well with values derived from ECMWF ERA5 atmospheric re-analysis model.

Fischer + 2021, Iserlohe + 2021, Iserlohe + 2022, Vacca + 2023

A new method to calibrate the atmospheric transmission mainly due to its water vapor content is based on a global atmospheric reanalysis provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). The method was validated using dedicated, in-flight FIFI-LS observations. It can provide corrections for all data from all SOFIA flights even after the fact (Fischer et al. 2012, Iserlohe et al. 2021, Iserlohe et al. 2022, Vacca et al. 2023). In particular, data from instruments with low to intermediate spectral resolution will benefit from an implementation of this method into their corresponding processing pipelines.