Open Science in Action:

SWxSOC's Role in Accelerating Data Release and Cloud Processing for Heliophysics Missions

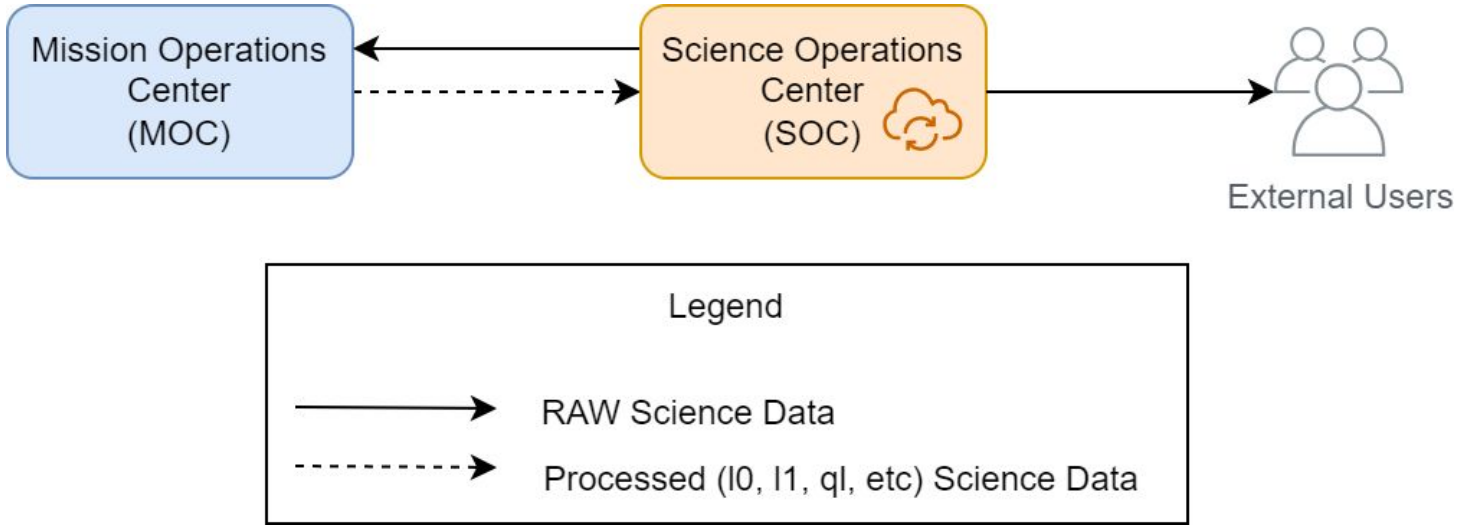
Damian Barrous-Dume[1] (dbarrous@navteca.com),
Amy Catherine Rager[2] (amy.c.rager@nasa.gov),
Andrew Robbertz[2] (andrew.robbertz@gd-ms.com),
Daniel Skeberdis[3] (daniel.skeberdis@nasa.gov),
Steven Christe[4] (steven.d.christe@nasa.gov),
Steve Kreisler[5] (stephen.kreisler-1@nasa.gov),
Tony Mercer[6] (tonyfig@berkley.edu),
William R Paterson[7] (william.r.patterson@nasa.gov)

[1] Community Coordinated Modeling Center, NASA GSFC, Navteca, Greenbelt, United States
[2] General Dynamics Mission Systems, Fairfax, United States
[3] a.i. solutions, Inc., Lanham, MD, United States
[4] NASA GSFC, Solar Physics Lab, Greenbelt, United States
[5] Columbus Technologies and Services, Greenbelt, MD, United States
[6] Space science laboratory, Berkeley, CA, United States
[7] NASA Goddard Space Flight Center, Geospace Physics Laboratory, Greenbelt, MD, United States

What is a Ground System?

Mission Operations Center (MOC) - a facility responsible for the daily management and control of a spacecraft, overseeing its health, status, and commanding functions to achieve the objectives of the mission.

Science Operations Center (SOC) - a facility that manages and processes data from scientific missions, provides operational support, and aids in coordinating various ground-based instruments or observation systems. They get forwarded the data from the Mission Operations Center (MOC).



Who we are at the SWxSOC?

Development of an **Open Source** multi-mission Science Operation Center to support space weather missions that provides file processing/calibration pipeline in alignment with NASA's **Open Science initiative (OSSI)**.

Services we provide:

Analysis

Provide a Cloud-based data analysis environment

Alerting

Provide alerts for instrument anomalies and track instrument performance metrics over time

Data Backup

Be able to serve as a primary and backup low-cost data center

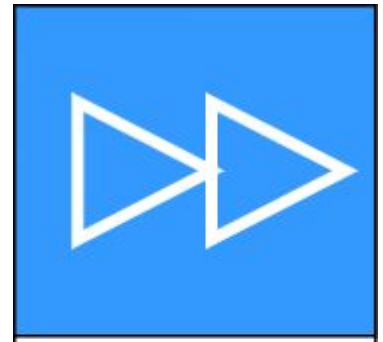
Data Processing

Pipeline which processes science files to higher data level CDF files

Monitoring

Graphical interfaces for monitoring data processing status and Telemetry and science data visualization

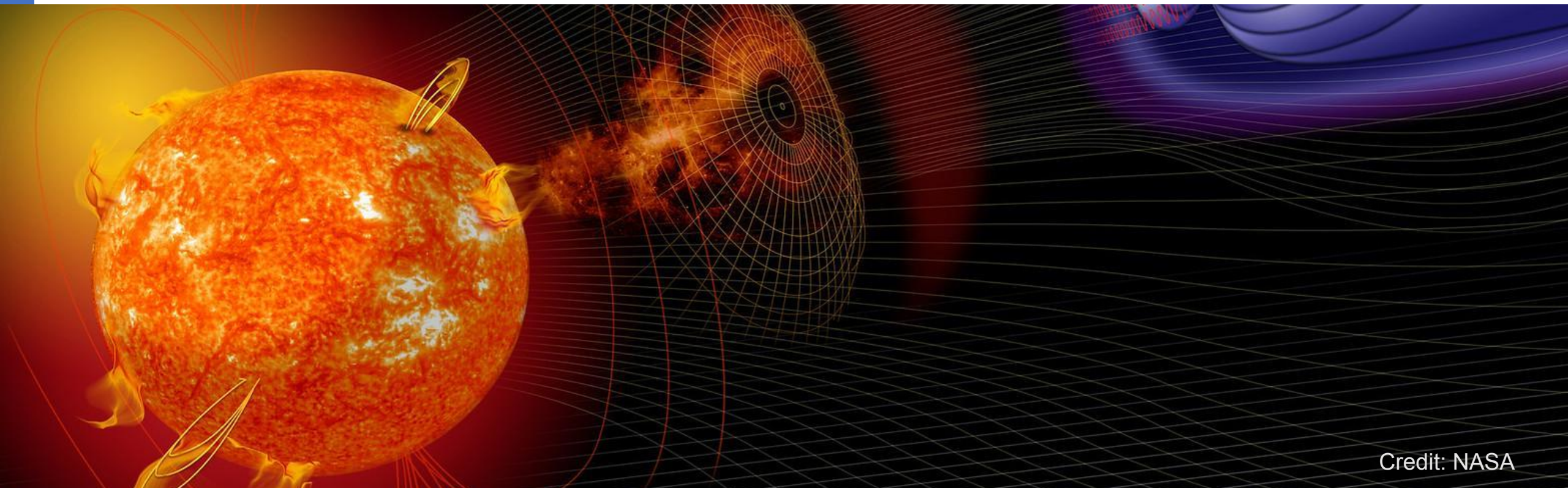
Telemetry Forwarding



We work with a Mission Operations Center to capture and forward telemetry and science data to whichever location

Supported Missions:

- Heliophysics Environmental and Radiation Measurement Experiment (**HERMES**), consists of 4 in-situ instruments.
- solar PolArization and Directivity X-ray Experiment (**PADRE**), consists of two solar-observing x-ray instruments.



ONCE UPON A TIME IN...

the Ground System SOCs of before

- Modernizing and adapting legacy codebases to meet the demands of new missions and evolving requirements
- Many codebases remain isolated within specific mission silos, accessible only to a limited number of individuals
- This is the definition of **Closed Science**

What is Open Science?

Open Science is a transparent, collaborative research and development approach which promotes the sharing of data, methods, and findings globally.

The Four Pillars of the Open Science Initiative (OSSI)

OPEN (**TRANSPARENT**) SCIENCE
scientific process and results should be visible, accessible, and understandable

OPEN **(ACCESSIBLE)** SCIENCE
data, tools, software, documentation, and publications should be accessible to all (FAIR)

OPEN **(INCLUSIVE)** SCIENCE
process and participants should welcome participation by and collaboration with diverse people and organizations

OPEN (**REPRODUCIBLE**) SCIENCE
scientific process and results should be open such that they are reproducible by members of the community

How the SWxSOC aligns with Open Science

Ground System SOCs should fundamentally already align with the values of **open science**, ensuring their operations are transparent, accessible, inclusive, and reproducible.

The Four Pillars of the Space Weather Science Operations Center

OPEN (**TRANSPARENT**) SCIENCE

All of the infrastructure as code and algorithms are made available publicly and open-source on our GitHub page.

OPEN **(ACCESSIBLE)** SCIENCE

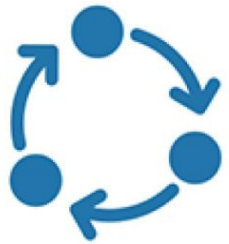
Since our code is publicly available, there are minimal barriers to access. You're welcome to contribute to our codebases, provided you adhere to the contribution guidelines.

OPEN **(INCLUSIVE)** SCIENCE

Anyone is free and welcome to create a pull request and contribute once again if they follow our contribution guidelines.

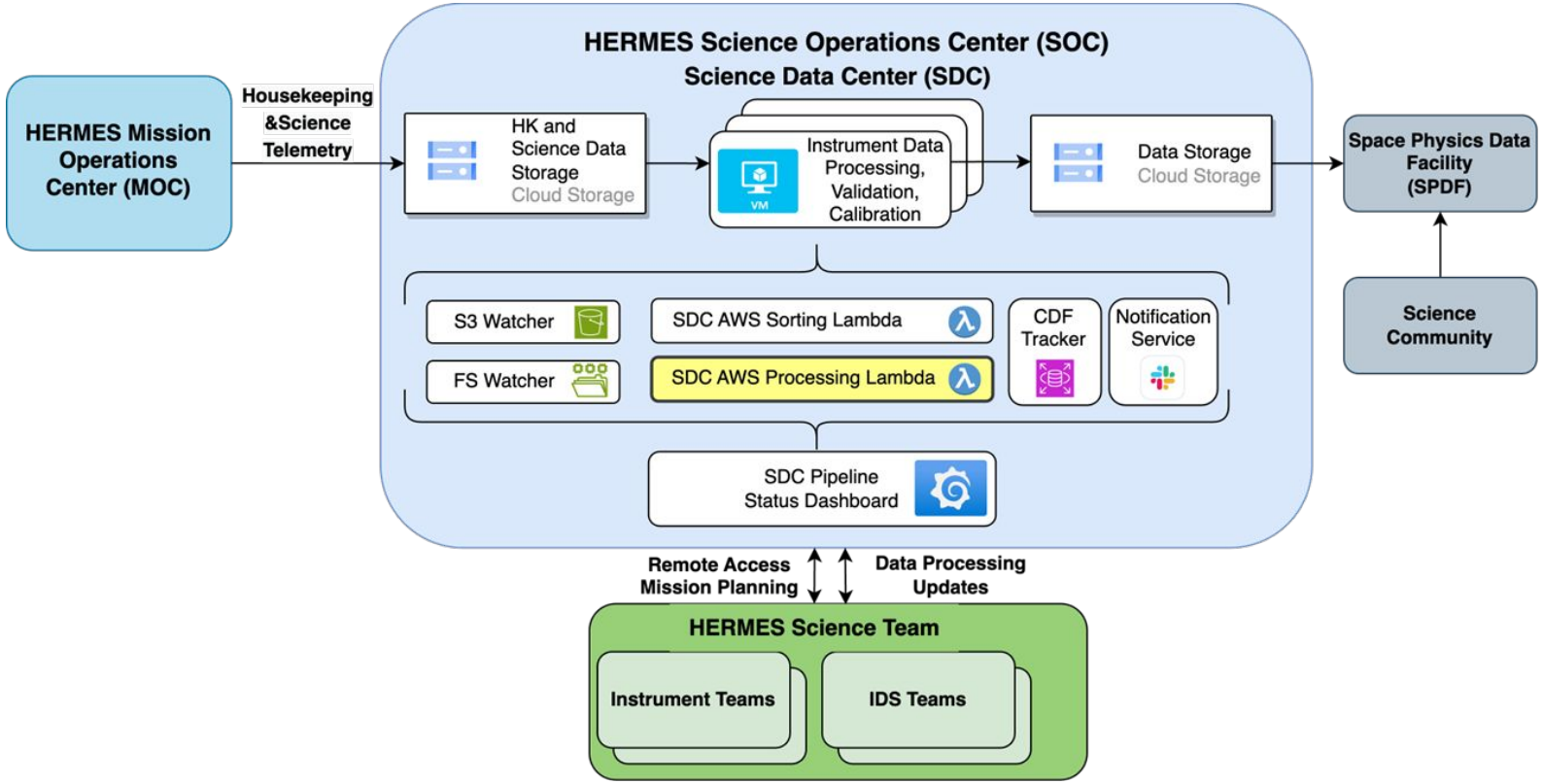
OPEN (**REPRODUCIBLE**) SCIENCE

Given that the code and algorithms are open-source, anyone can reproduce the calibrations for the Science files at their desired data levels, such as l1, l2, ql, and so on.



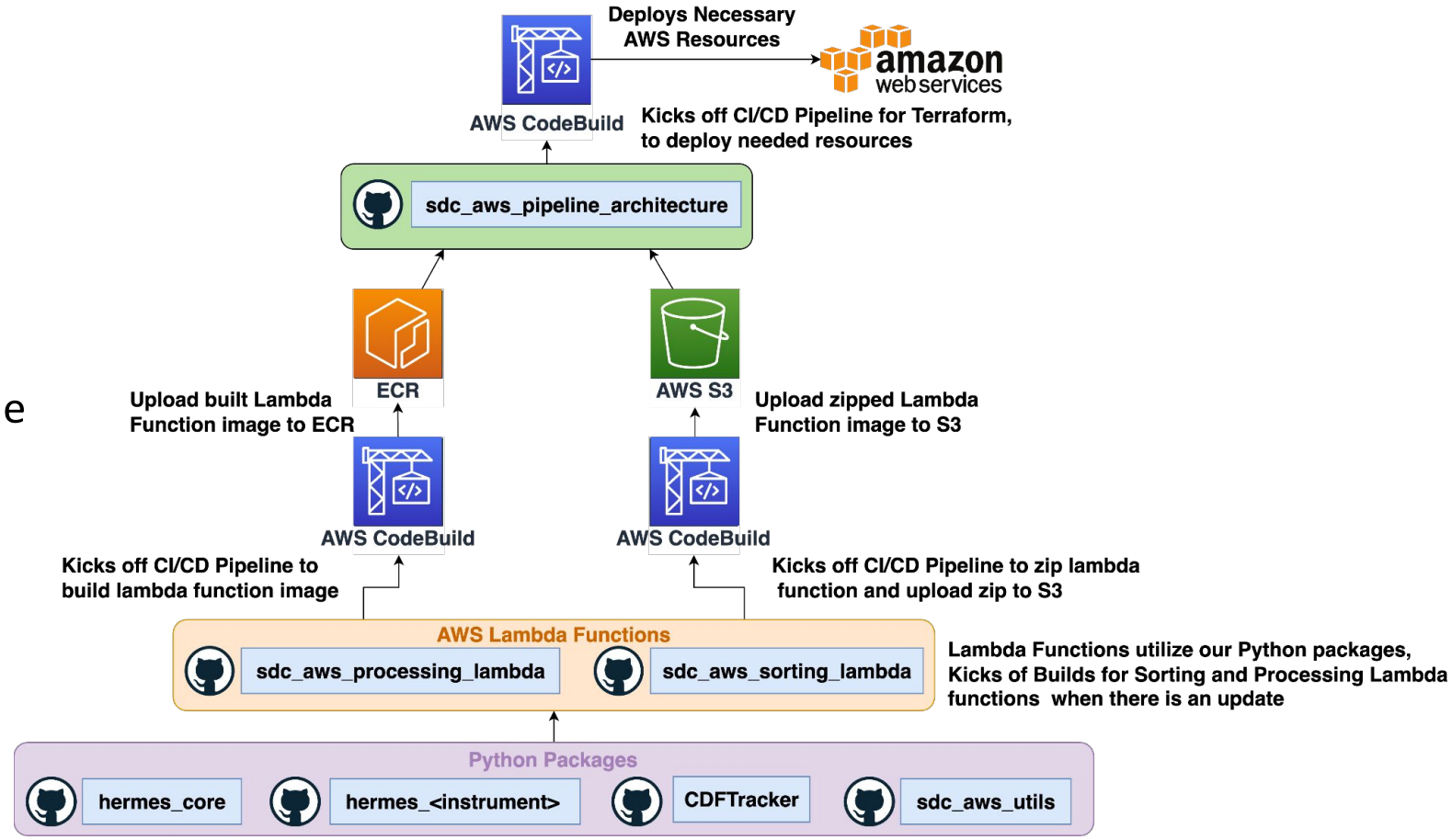
A look at our Evolving Open Source Data Processing Pipeline

- The MOC sends binary files to the SOC Servers, which are then relayed to S3 using our containerized FSWatcher tool in Docker.
- Cloudwatch S3 Bucket events activate a lambda function for sorting and processing, placing files into the relevant instrument buckets.
- After processing, the containerized S3Watcher tool detects the new files and retrieves them for the server.
- The processed data is then distributed to the MOC, instrument teams, and SPDF.



A look at our Continuous Integration/Continuous Development set-up

- We have different GitHub repos for all of our codebase. Including repos for different python packages that are utilized by the different AWS Lambda functions using the Python runtime environment
- When a change is made to either the functions or packages that triggers them to be rebuilt and redeployed via our AWS Codebuild CI/CD Process
- All of the architecture is kept as Infrastructure as Code written in Terraform



Other tools and packages we offer

Mission Core & Instrument Packages: Sets of packages providing a structured framework for our primary processing pipeline, with adaptability for other pipelines.

CDFWriter: A component of the mission core package, this Python library facilitates reading, writing, and structuring CDF files for user-friendly access. Currently only supporting the HERMES mission.

CDFTracker: A Python package designed for monitoring Raw Binary and CDF Files within a Relational Database.

FSWatcher & S3Watcher: Containerized tools that, respectively, monitor directories for new files to upload to S3, and monitor AWS S3 buckets to download new files to a local machine in an event-driven fashion.

For more information on the tools and packages we offer, check out Andrew Robbertz Poster on them via this QR Code:

Challenges and looking forward…

Institutional Roadblocks

- Encountering barriers to open-sourcing due to concerns over intellectual property, although the process is improving with the Transform to Open Science (TOPS) mission being introduced.
- To streamline the open-sourcing process, US Government-generated IP should be made freely available rather than being vetted for proprietary potential, unless it is ITAR classified or poses a national security risk.

Modern Integration with CDF

- Navigating the compatibility of CDF with modern-day Python development can be challenging, particularly when it comes to installing the appropriate libraries for distinct Python versions and operating systems.
- Overcame these by creating custom Python packages (CDFWriter) to bridge the gap at least for HERMES as well as creating our own development docker container with pre-installed CDF binaries.

Seeking Community Feedback:

- As we continue on this open development journey, we are curious about best practices or strategies others have employed to navigate IP concerns and ensure broad access to government-developed resources.
- Also open to contributions to any of our open source offerings we have discussed.

Thank you for your time!

Any Questions?

For more information on SWxSOC, please visit our website: