



# Science Data Processing and Data Products

Chris Pankratz

LASP/CU

(303)492-0696

[chris.pankratz@lasp.colorado.edu](mailto:chris.pankratz@lasp.colorado.edu)

# Abstract

---

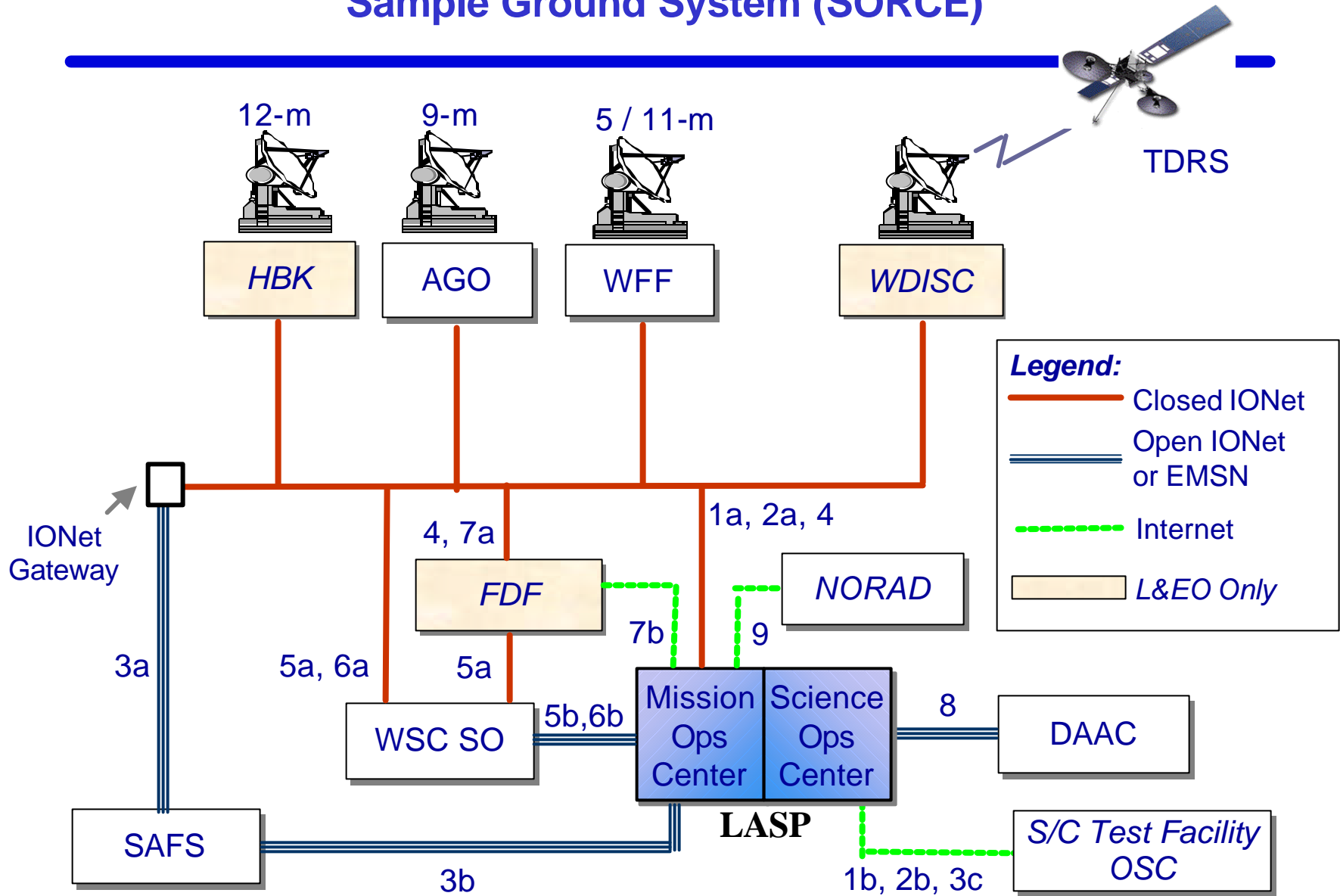
- The SORCE Science Data System will produce assigned EOS Standard Products using data acquired from the SORCE satellite. The SORCE Science Operations Center (S-SOC), located at the Laboratory for Atmospheric and Space Physics (LASP) in Boulder, Colorado, USA, will produce these science products along with associated metadata and documentation, delivering them to the Goddard DAAC for archival and distribution. The SORCE project will produce two principal science data products for delivery to the GSFC DAAC: Total Solar Irradiance (TSI) data and Solar Spectral Irradiance data. The TIM instrument measurements will be combined to produce representative daily and 6-hourly values of the Total Solar Irradiance. The SOLSTICE, SIM, and XPS measurements will be combined into merged daily and 6-hourly spectra, each containing representative irradiances reported from 1-2000 nm (excluding 31-120 nm, which is not covered by the SORCE instruments) on a uniform wavelength scale, which will vary from 1-34 nm over the entire spectral interval.
- Science data processing will be performed automatically within 24 hours of data reception and a preview release of science data is made available to the public within 48 hours, following preliminary data inspection. Appropriate corrections for some time-dependent processes (such as instrument degradation) will require in-flight calibration data several months into the future, so that the daily production of science products beyond provisional quality will lag by a period of several months. These finished data will be used to generate products that are distributed to the GSFC DAAC, therefore the DAAC will not receive SORCE data products until several months after initial data reception.

# Outline

---

- Ground System Overview
- Data Processing System Overview
- Data Products
  - Identification of Data Products
  - Data Availability
  - Data Configuration Management

# Sample Ground System (SORCE)



## Summary of Ground Data Flows

- NASA IP Operations Network (IONet) used for real-time operations
  - MOC, WFF, and WDISC are on Closed IONet
- NASA EOS Mission Services Network (EMSn) will be used for some non-real time data transfers; regular Internet will be used for others

	Data Stream	Source	Destination	Network	Daily Traffic
1a	Real-time telemetry	Station	MOC	IONet	64 kbps * 2 passes * 12 min
1b	Real-time telemetry	OSC Test	MOC	ISDN	64 kbps
2a	Real-time command	MOC	Station	IONet	2 kbps * 2 passes * 5 min
2b	Real-time command	MOC	OSC Test	ISDN	2 kbps
3a	Post-pass telemetry data files	Stations	C-SAFS	IONet	70 MB * 2 passes
3b	Post-pass telemetry data files	C-SAFS	MOC	EMSn	70 MB
3c	Telemetry data files	OSC	MOC	ISDN	25 MB
4	UTDF Tracking data	Station	FDF, MOC	IONet	< 1 MB * 2 passes
5a	Ground station schedules	WSC SO	Station, MOC	IONet	< 1 MB per week
5b	Ground station schedules	WSC SO	MOC	Internet	< 1 MB per week
6a	Pass summary messages	Station	WSC SO	IONet	< 1 MB per week
6b	Pass summary messages	WSC SO	MOC	Internet	< 1 MB per week
7a	IIRV spacecraft ephemeris	FDF	Stations, SO	IONet	< 1 MB for one week L&EO
7b	IIRV spacecraft ephemeris	FDF	MOC	Internet	< 1 MB for one week L&EO
8	Data products	SOC	DAAC	EMSn	130 MB
9	Two-line element sets	NORAD	MOC	Internet	< 1 MB per week

## Other Ground System Components

---

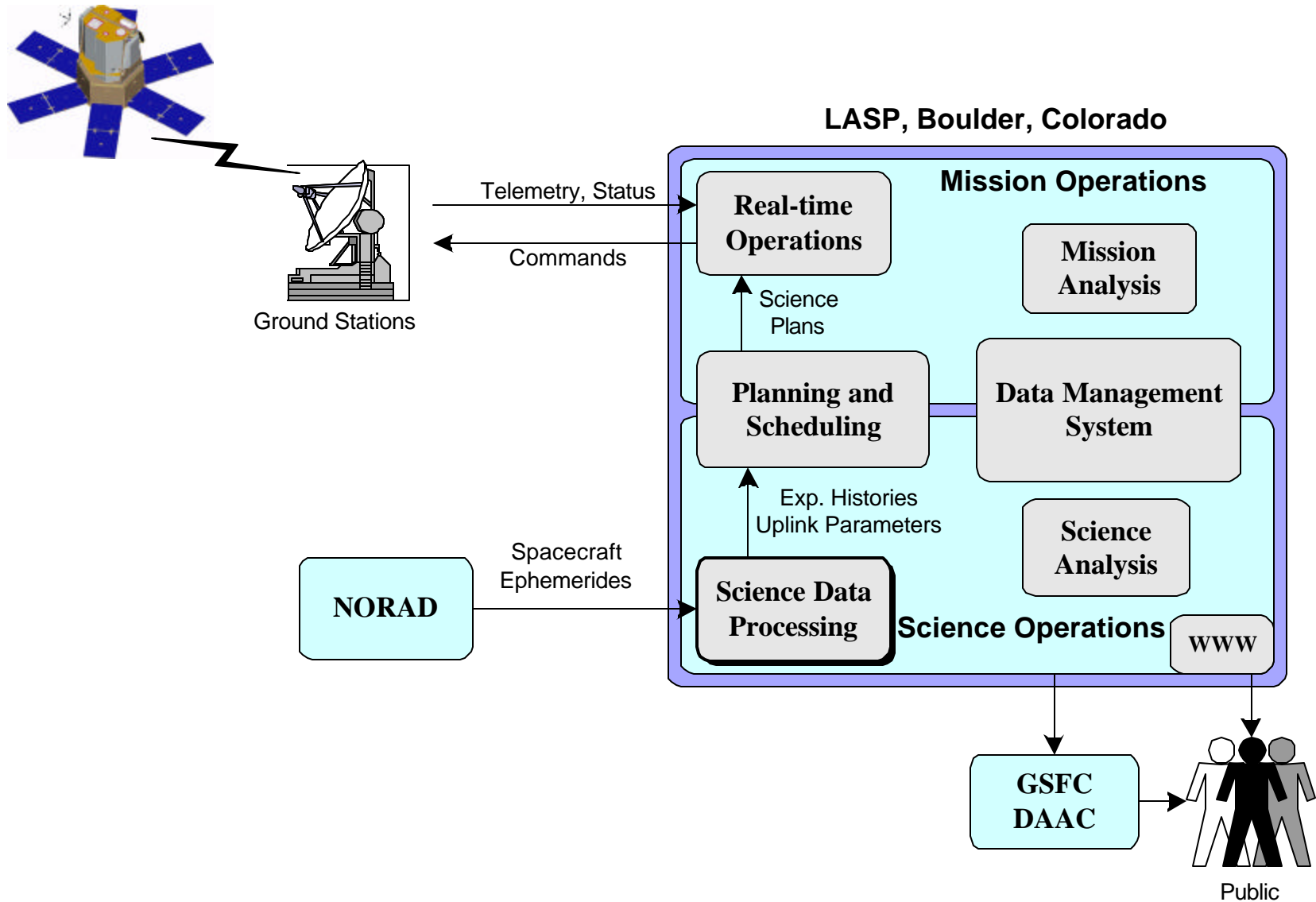
- Mission Operations Center (MOC) located at CU/LASP
  - Handles all real-time spacecraft activities, bus-related non-real time activities, and processing of telemetry data up through Level 1
- Science Operations Center (SOC) located at CU/LASP
  - Performs science data processing above Level 1
  - Archives selected data products with the DAAC
- Flight Dynamics Facility (FDF) at GSFC
  - Performs orbit determination and prediction during first week on orbit
- White Sands Complex Scheduling Office (WSC SO)
  - Schedules GN and SN contacts for SORCE
- Standard Autonomous File Server (SAFS) at GSFC
  - Collects playback telemetry data files from ground stations and forwards to MOC
- Distributed Active Archive Center (DAAC) at GSFC
  - Archives SORCE data and distributes to the general science community
- OSC Spacecraft Test Facility, Dulles Virginia
  - Used for the MOC's pre-launch tests with the SORCE spacecraft and testing (both prelaunch and during the mission) with the spacecraft Flatsat simulator

## Data Processing System Overview

---

- All data processing will occur locally at LASP in Boulder, Colorado at the SORCE Science Operations Center (SOC)
- LASP will process, manage, analyze, and distribute science data products via the SORCE Science Data System
- Algorithms are implemented as described in the SORCE Algorithm Theoretical Basis Document (ATBD) to produce science data products
- Data will be processed within 24 hours of reception, after all possible telemetry data retransmissions
- *Preview* products (unvalidated with preliminary time-dependent corrections) are released within 48 hours of data reception
- Fully-calibrated (*finished*) science products are produced and publicly released ~3 months after data reception, following in-flight calibration analysis and application.
- *Finished standard* (Level 3) data products are generated and distributed to the GSFC DAAC, along with associated metadata
- Reprocessing will take place following algorithm and/or calibration data updates, recalculating only those products that are affected by a change

# Data System Principal Interfaces



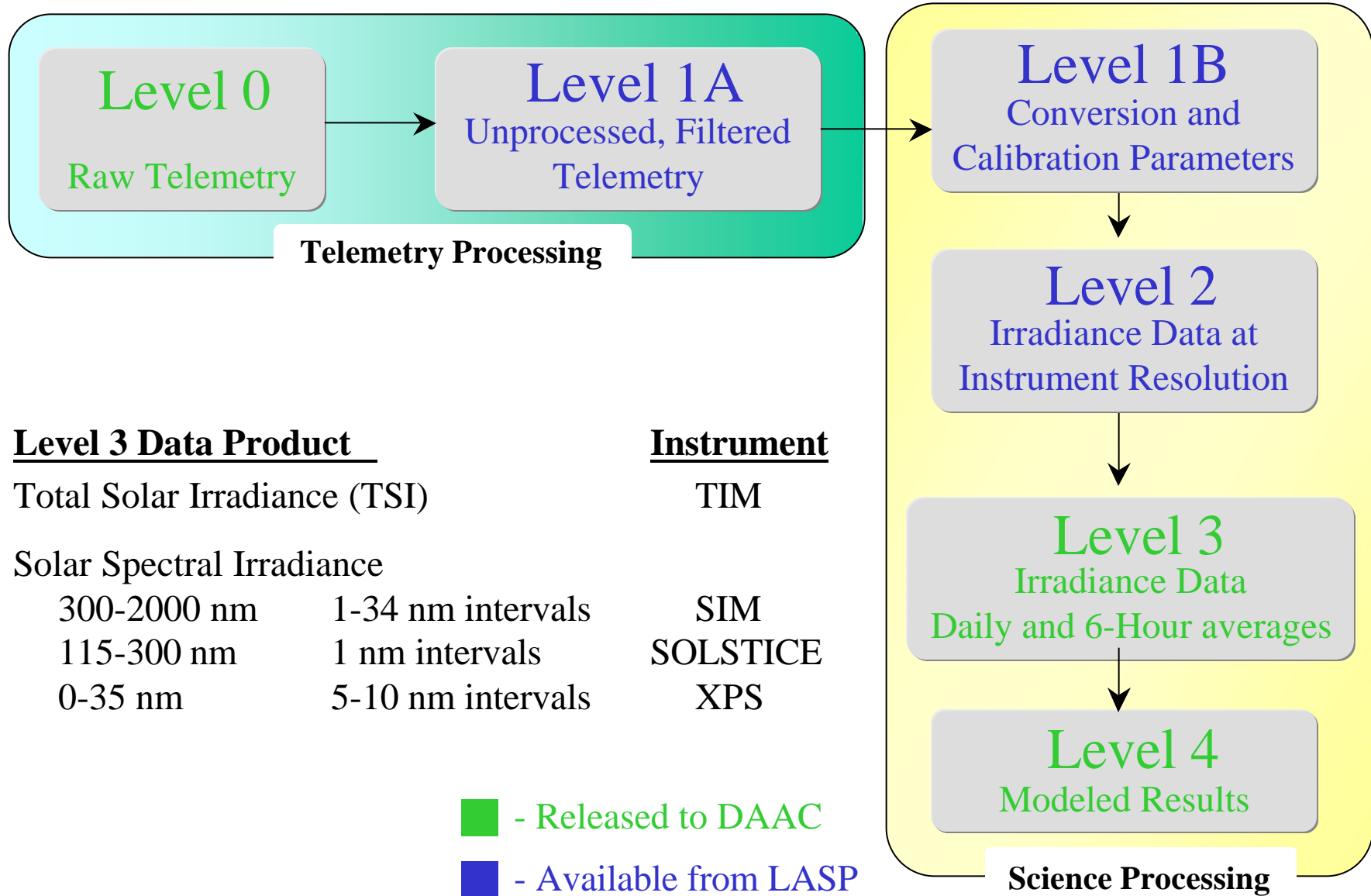


# SORCE Data Level Definitions

---

- **Level 0**
  - Reconstructed, unprocessed instrument and spacecraft data at full space-time resolution; any and all communications artifacts, e.g., synchronization frames, communications headers, duplicate data removed
- **Level 1A**
  - Unprocessed instrument data at full resolution, time-referenced, sorted by experiment and/or data type
- **Level 1B**
  - All processed information required for conversion into physical units, including correction and calibration factors and attitude/ephemeris parameters
- **Level 2**
  - Sorted instrument data (level 1A) processed into physical units with all applicable conversions and calibrations (level 1B) applied
- **Level 3 (Standard Data Products)**
  - Processed instrument data, time-averaged and/or spectrally resampled onto a uniform wavelength scale (spectral measurements only)
- **Level 4**
  - Model-based Calculations

# Science Processing Flow



# Science Data Products

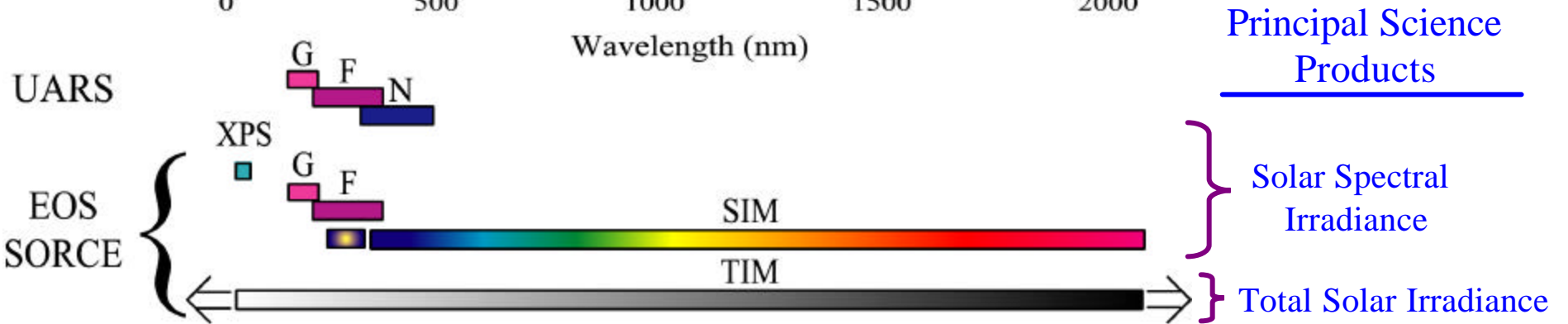
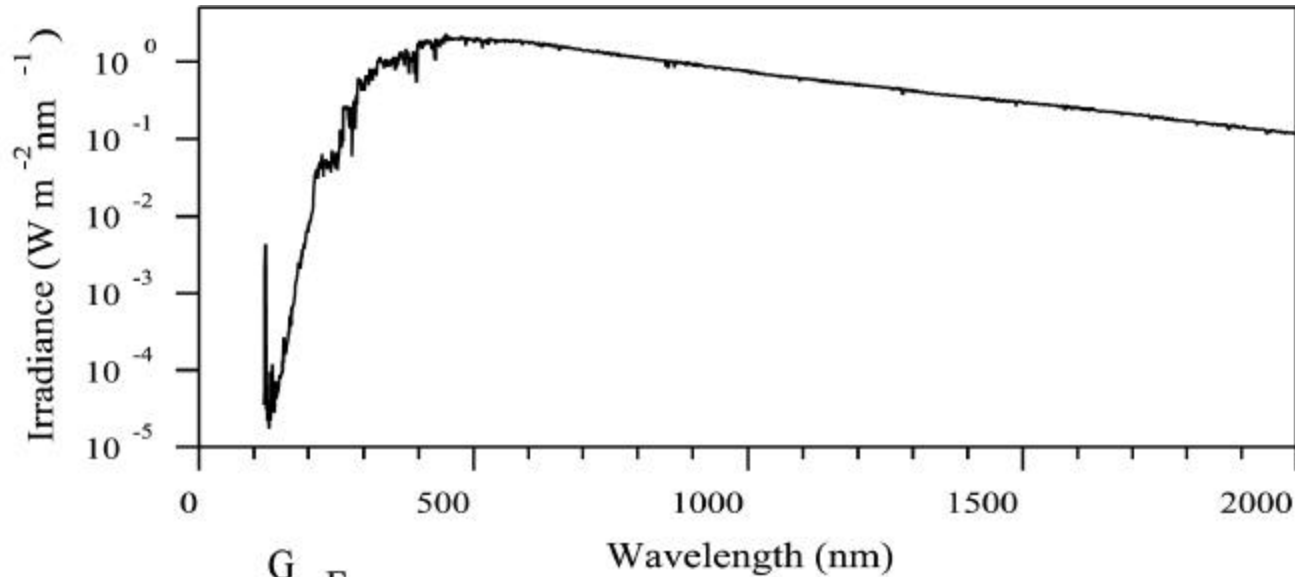
	Level 2	Level 3
TIM	<ul style="list-style-type: none"> <li>Total Solar Irradiance at ~60 second time resolution</li> </ul>	<ul style="list-style-type: none"> <li>Total Irradiance averaged over 6 and 24 hours</li> </ul>
SIM	<ul style="list-style-type: none"> <li>Absolute ESR solar irradiances at 10 second time resolution at ~40 wavelengths</li> <li>Calibrated diode irradiances at full instrument resolution (0.25-34 nm, ~400 <math>\lambda</math>'s) every 1 second.</li> </ul>	<ul style="list-style-type: none"> <li>Irradiances on fixed wavelength scale (1-34 nm resolution, ~400 <math>\lambda</math>'s) and averaged over 6 and 24 hours</li> <li>Integrated 1.5 mm band averaged over 24 hours</li> </ul>
SOLSTICE	<ul style="list-style-type: none"> <li>Solar irradiances at instrument resolution (0.1 nm) and 1 second time resolution</li> <li>Stellar irradiances at instrument resolution (1-5 nm) and 1 second time resolution</li> </ul>	<ul style="list-style-type: none"> <li>Irradiances integrated to 1 nm resolution and averaged over 6 and 24 hours</li> <li>200-300 nm integrated irradiance averaged over 24 hours</li> <li>Mg II core-to-wing ratio (6-hour and daily averages)</li> <li>Atomic Line Integrations (daily and 6-hour averages, plus individual extractions)</li> <li>Irradiances at instrument resolution (0.1 nm) averaged over 6 and 24 hours</li> <li>Occasional stellar spectra at instrument resolution (1-5 nm)</li> <li>Stellar irradiance integrations (~few wavelengths per orbit)</li> </ul>
XPS	<ul style="list-style-type: none"> <li>Irradiances at instrument resolution (5-10 nm) and ~6 min time cadence</li> </ul>	<ul style="list-style-type: none"> <li>Irradiances at instrument resolution (5-10 nm) averaged over 6 and 24 hours</li> </ul>

\* Available from LASP

\* Available from the GSFC DAAC

# Data Product Spectral Coverage

## Representative Solar Spectrum



## Data Product Availability

---

- Level 0 and Level 3 data products will be delivered to the GSFC DAAC in HDF5 format. Links to data access sites will appear on the SORCE web page after launch.
- Level 0 (telemetry)
  - Delivered daily to the GSFC DAAC (<http://daac.gsfc.nasa.gov/>)
    - Initial delivery shortly after launch
- Lower-level Science Products (Levels 1B and 2)
  - Data will be available on the same delivery schedule as the standard science products. Designation of a distribution site is in progress.
- Standard Science Products (Levels 3+)
  - Initial public data release 120 days after launch – 3 months following instrument commissioning and initial on-orbit characterization
  - *Preview* (unfinished) data will be released 48 hours after collection
  - **Finished Total Solar Irradiance (TSI)**
    - Monthly releases of TSI data to DAAC with three month latency (i.e. each release will contain data three months old)
    - Each release incorporates data for the entire mission into two files (6- and 24-hour averages)
  - **Finished Solar Spectral Irradiance**
    - Daily releases of spectral data with three month latency
    - Each release includes spectral irradiances, fits to key atomic lines, the Mg II c/w ratio, and other relevant solar/geophysical parameters
- Data processing algorithm packages will be delivered to the DAAC after each major software release

# Data Configuration Management

---

- The configuration management of data is literally built-in the the SORCE data processing system
- Each data product is versioned and will have an associated pedigree of dependencies defining that version, including
  - The **Software version** of modules used in the calculation
  - Any **Job control parameters** that specify, for instance, a tolerance criterion that can be changed without requiring a software release
  - **Versions of other managed data products** that were used by the algorithm to generate the given data product (e.g. Level 3 product generation requires Level 2 irradiance data and a degradation correction – two dependencies)
- This **pedigree** defines the state of the data processing system used to generate a particular data product and **is generated at runtime by the processing software itself**
- Since the pedigree is generated at run-time, a change to one algorithm or calibration dataset will cause the automatic propagation of version changes necessitated by the applicable dependency tree
  - including the execution of those algorithms required to facilitate the generation of new data
  - Only those algorithms that are required to execute as part of the version change are executed. Algorithms for data products not requiring a version change do not execute
- Benefits:
  - Enhanced processing performance – Only selected algorithms must execute
  - Minimized data volumes – Data are stored only for those algorithms that execute, and only if new data were required

# Data Processing Operations

---

- Production processing will take place on a Sun Solaris multiprocessor server autonomously following data downlink for the preceding calendar day
- Processing will occur for each complete GMT calendar day.
- The system is designed to accommodate frequent reprocessing in a highly-optimized manner
  - Software for each instrument and data product is configuration managed and executed independently
  - As algorithm or calibration data changes occur, only those data products with traceable dependencies from the changed item require reprocessing. All other data products remain static.
  - This technique optimizes both storage and CPU requirements by limiting reprocessing to only those data that need reprocessing, thereby speeding reprocessing activities and reducing data volumes
- Data processing algorithms are triggered, as required, i.e.
  - Data production process flow will follow the tree of dependencies between distinct data products, established within the algorithms themselves
  - Processing will start by initiating calculation for each high-level (level 3) data product, thereby triggering processing of each required lower-level product (dependency), as necessary.
  - Calculations for each of these dependencies will then trigger processing of its dependencies, etc.

# Summary

---

- Preview (unfinished) SORCE data products will be available approximately 48 hours following data reception
- Finished SORCE data products will be available four months after launch, followed by daily releases with a three-month lag
- SORCE will benefit from a highly optimized data system that will
  - Store all data used in the generation of scientific data, including all Level 1B (calibration) parameters
  - Minimize the time and storage requirements taken by reprocessing activities by obviating the need to recalculate that that already exist in the database and have not changed
  - Facilitate improved accounting of data products, and provide access to individual calibration factors, thereby easing data inspection and diagnosis
  - Ease future extension and modification of the system by utilizing modern object-oriented principles with minimized coupling between software modules.



## Applicable URLs

---

- **SORCE web site** – <http://lasp.colorado.edu/sorce>
- **Hierarchical Data Format (HDF)** – <http://hdf.ncsa.uiuc.edu/HDF5/>
- **NASA Goddard DAAC** – <http://daac.gsfc.nasa.gov/>