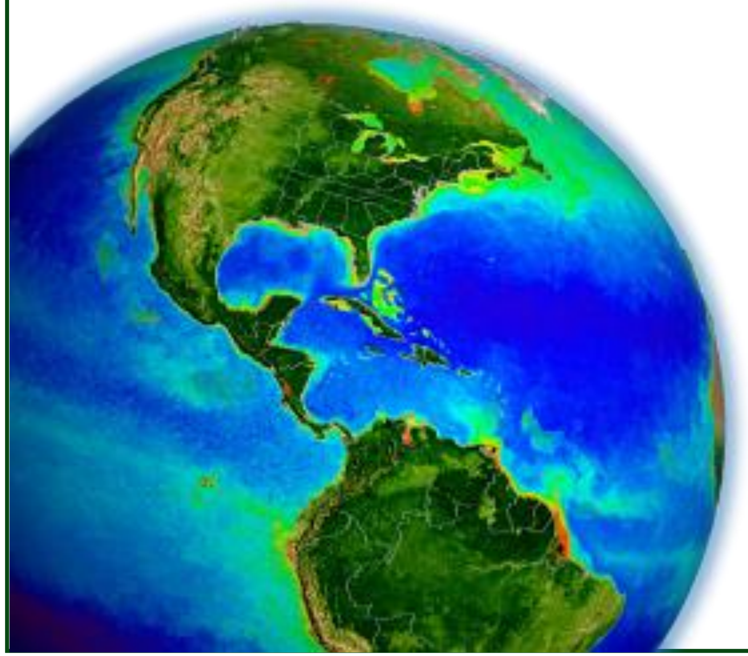
PACE Science Data Segment

Bryan Franz
PACE SDS Manager

Ocean Ecology Laboratory
NASA Goddard Space Flight Center



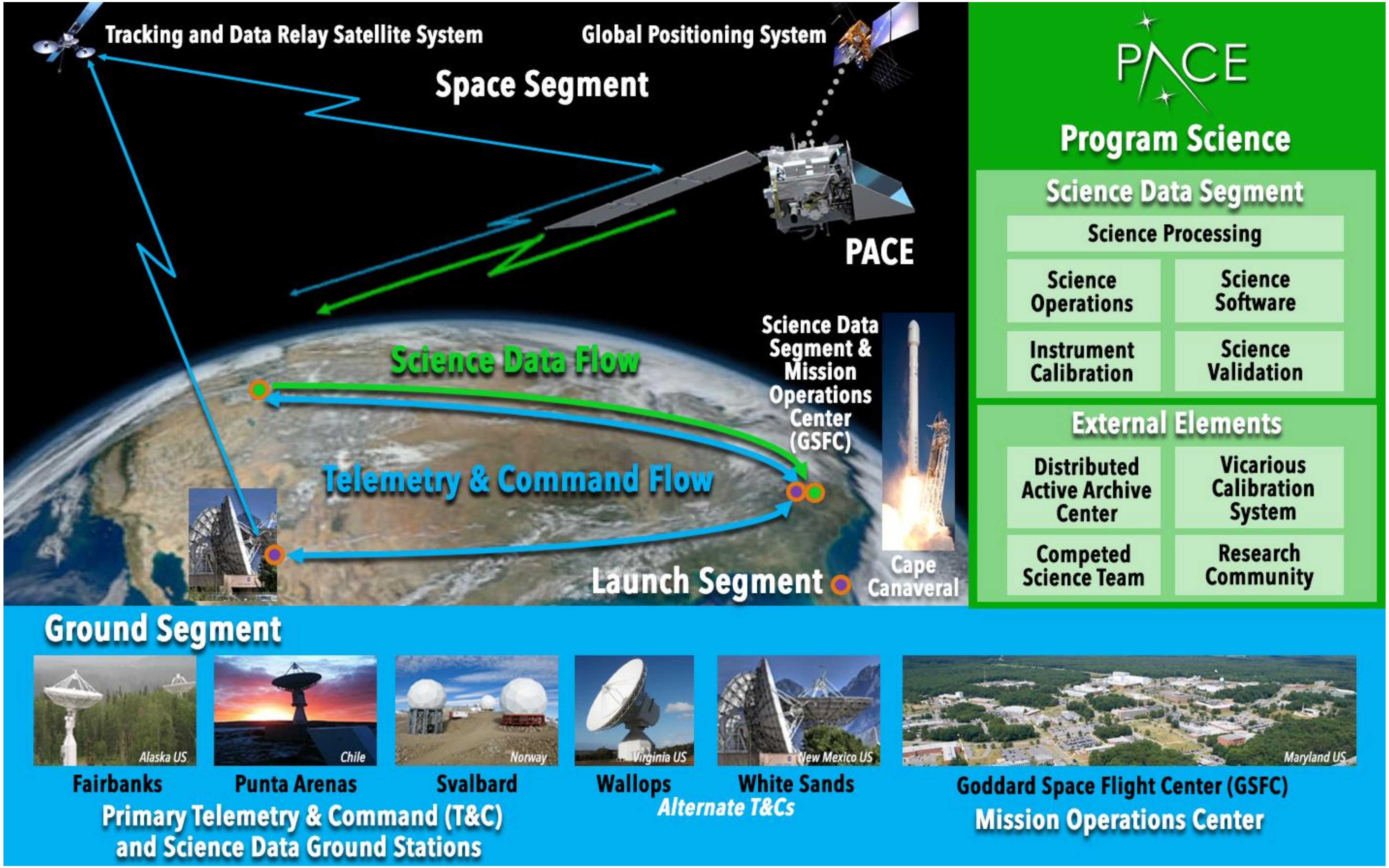
PACE Science and Applications Team

1 May 2020

Contents

1. Roles and Responsibilities of the PACE Science Data Segment
2. Science Processing Software Development Approach
3. Product and Algorithm Selection Process
4. Current Status of Implementation
5. Simulated Data
6. Final Thoughts

PACE Mission Architecture



Science Data Segment Organization

Ocean Ecology Laboratory (616)
C. Del Castillo, Lab Chief
B. Franz, Deputy Lab Chief

PACE SDS Manager
B. Franz

SDS SE
F. Patt

IT Systems
P. Smith

PACE Project Science
J. Werdell

OB.DAAC*
S. Bailey

Science Operations
F. Patt
- Scheduling (L. Hong)

Data Processing
J. Wilding

Calibration
G. Meister
- Vicarious (S. Craig)

Science Software
B. Franz, S. Bailey

- Software Eng (D. Shea)
- L1A (L. Hong)
- L1B/GEO (J. Gales, G. Fireman)
- L1C (M. Montes)
- OCI OC (M. Zhang)
- OCI AER (S. Anderson)
- OCI CLD (W. Robinson)
- Polarimetry (M. Gao)

HARP2 (X. Xu, UMBC)
SPEXone (O. Hassekamp, SRON)

Simulated Data (P. Castellanos, GMAO)

Validation
J. Werdell

- Protocols (A. Mannino)
- Operations (C. Proctor)
- Datasets (V. Calzado)
- In situ Radiometry (S. Craig)

* ESDIS Project

PACE Science Data Segment

Instrument Scheduling:

- delivery of integrated activity schedules to MOC

Science Data Processing:

- acquisition of Level-0 data
- processing to ocean color, aerosol, cloud, polarimetry
- delivery to DAAC

Calibration Support:

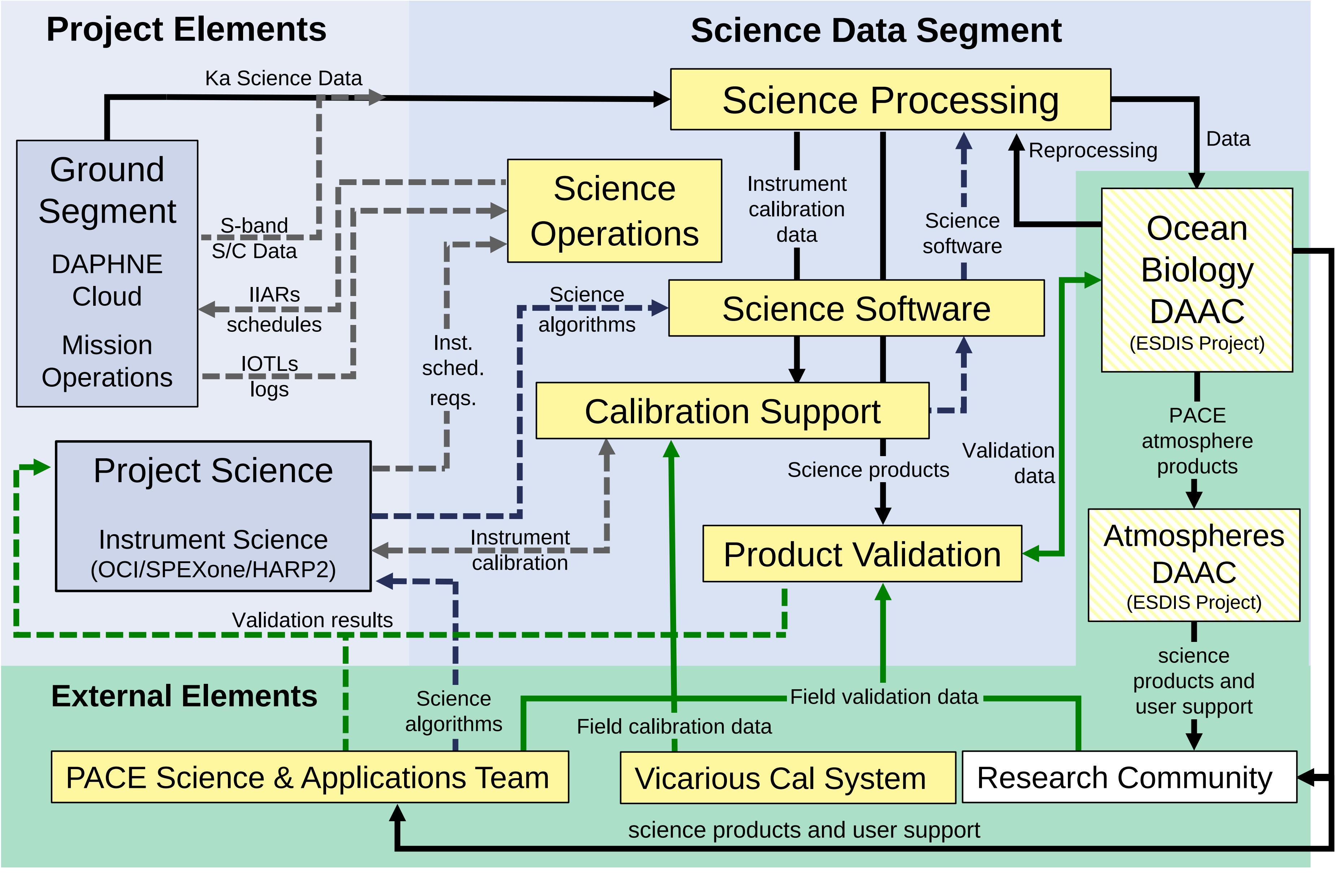
- vicarious cal analyses
- specialized processing

Product Validation Support:

- in situ processing, protocols, quality control
- match-up analyses

Science Software Development:

- proc chain development
- algorithm integration and testing (life of mission)

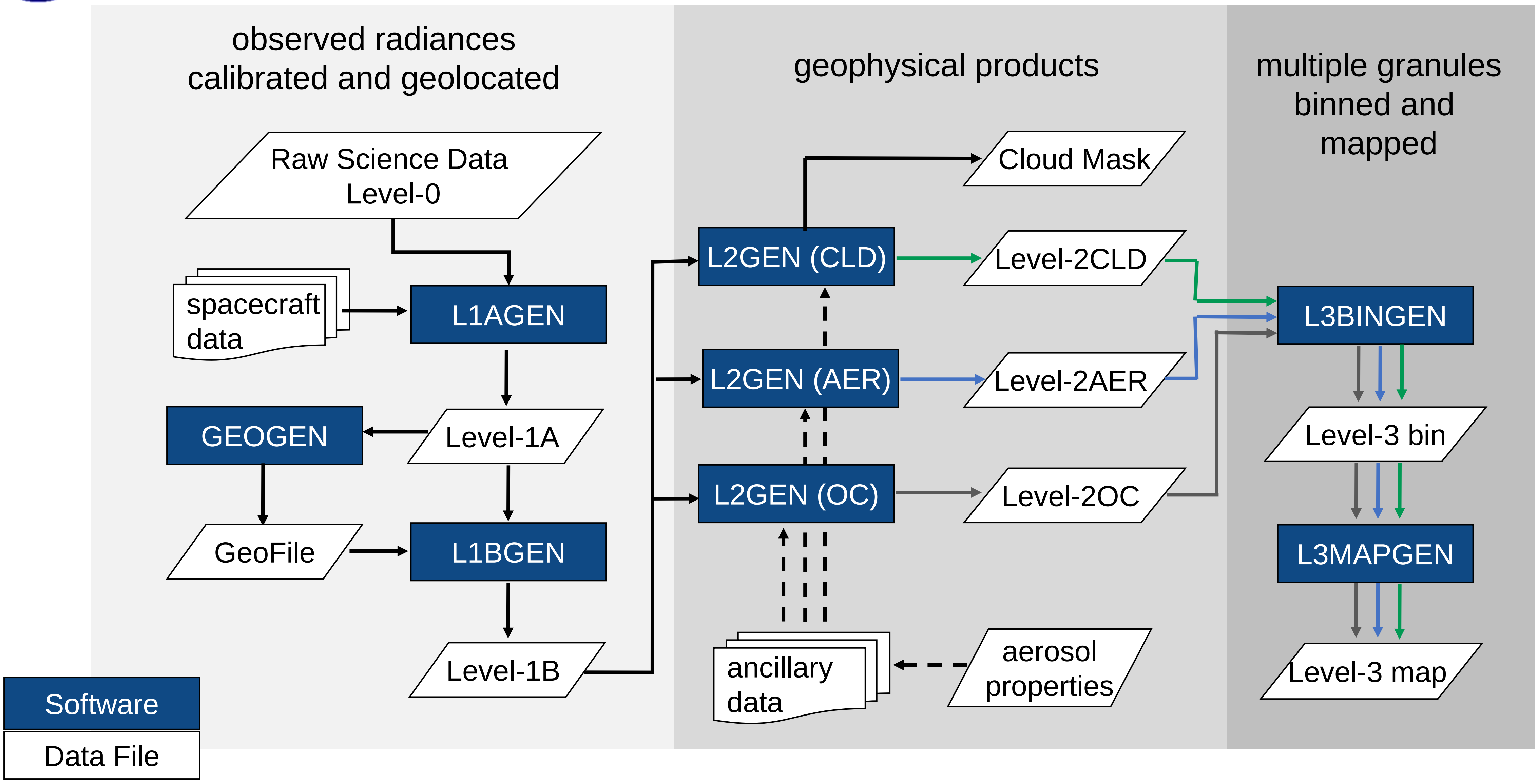


Science Data Product Level Definitions

Data Level	Description
Level 0	Lowest level science data (e.g., CCSDS packets)
Level 1A	Uncalibrated science data in self–describing archive format (netCDF4)
Level 1B	Calibrated, geo-located science data as observed
Level 1C	Calibrated, geo-located, co-registered (resampled) science data
Level 2	Science products derived from Level-1B/C
Level 3	Temporally and spatially composited science products

- All data above Level-0 will be netCDF4 and conform to NASA ESDS meta-data standards.
- Data format specifications are maintained by SDS in concurrence with OB.DAAC.
- Level-2 files will be organized into product suites specific to each Discipline (e.g., OC, AER, CLD), with potential for further subdivision (e.g., OC IOPs)
- Sample data formats (draft):
  https://oceancolor.gsfc.nasa.gov/fileshare/sean_bailey/pace/OCI_Level-2_Data_Structure.cdl
  https://oceancolor.gsfc.nasa.gov/fileshare/sean_bailey/pace/OCI_Level-1B_Data_Structure.cdl

Science Software Components (OCI)



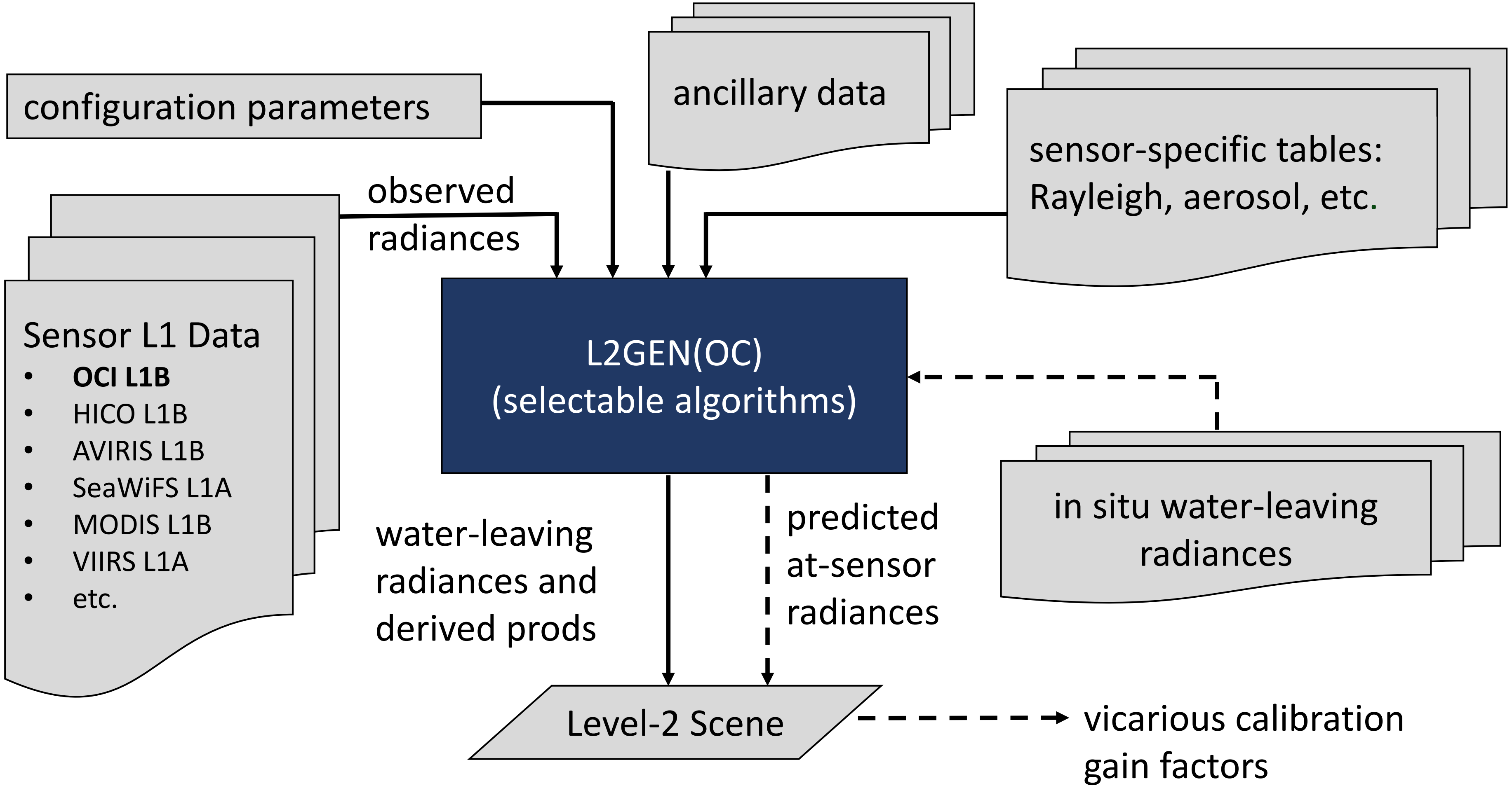
Multi-Mission, Multi-Algorithm Software Framework

L2GEN(OC) has the capability to run various AC algorithms and a multitude of derived product algorithms.

Algorithms and desired output products are selectable at runtime via parameter files.

Sensor-specific elements are abstracted in the L1 reader and sensor-specific LUTs.

L2GEN(AER) and L2GEN(CLD) are being developed with a similar framework.



Product and Algorithm Selection Process

SDS implements algorithms and produces the science data products selected by Project Science

Project Science will utilize the process described in the PACE Science Data Product Selection Plan

- https://pace.oceansciences.org/docs/PACE-SCI-PLAN-0143-DPSP_20190314.pdf

Key elements of the Plan include:

- establishment of a Science Operations Board (SOB) consisting of representation from NASA Program Management, Project Science, Science Data Segment, and DAAC, to oversee the selection process and consider resource requirements as well as science value
- establishment of a stepwise approach for implementation, test, evaluation, and maturation to standard product

Science Data Product Maturity

Standard

Standard science data products (*with associated standard algorithms*) will be produced by the SDS and delivered to the NASA-assigned DAAC for permanent archival and distribution. For PACE, standard products will include all **required** mission products, as well as any additional products selected as standard through the process described in Section 2.5 of the PACE Science Data Product Selection Plan.

Provisional

All new products (*or alternative algorithms*) will be defined as provisional at start of production, and they will remain provisional until some level of validation has been performed, and until performance, documentation, and science value is deemed sufficient for reclassification as standard products.

Test

Test products are those that have been implemented by the SDS into production-capable science code, to enable assessment of resource requirements and feasibility for global production.

Special

Applications, region-specific products, alternative algorithms to be distributed by the DAAC but not necessarily produced by SDS.

Required OCI Science Products

Discipline	Product	Current Algorithm Implementation
Ocean Color (OC)	Water-leaving reflectances centered on (±2.5 nm) 350, 360, 385, 412, 425, 443, 460, 475, 490, 510, 532, 555, and 583, 617, 640, 655, 710 (15 nm bandwidth) and 665, 678 nm (10 nm bandwidth)	• Ibrahim et al. 2018, 2019 (MBAC) • Planning hyperspectral Rrs (5-nm@2.5-nm sampling)
	Concentration of chlorophyll-a	• Hu et al. 2012 (OCx-CI), with updates
	Diffuse attenuation coefficients 400-600 nm	• Werdell et al. 2013 (GIOP), with updates • Spectral sampling TBD
	Phytoplankton absorption 400-600 nm	
	Non-algal particle plus dissolved organic matter absorption 400-600 nm	
	Particle backscattering 400-600 nm	
	Fluorescence line height	• MODIS approach, nFLH
Aerosol (AER)	Total aerosol optical depth at 380, 440, 500, 550 and 675 nm	• Deep Blue & Dark Target implemented, runtime selectable • No 380 AOD yet implemented
	Fraction visible aerosol optical depth of fine mode aerosols over oceans at 550 nm	
Clouds (CLD)	Cloud layer detection for optical depth < 0.3	• MODIS/VIIRS approach, minus thermal dependence (A. Sayer)
	Cloud top pressure of opaque (optical depth > 3) clouds	• O2 A-band approach (A. Sayer)
	Optical thickness of liquid clouds	• Nakajima-King, based on CHIMAERA implementation • LUTs not yet tuned for OCI band passes
	Optical thickness of ice clouds	
	Effective radius of liquid clouds	
	Effective radius of ice clouds	
	Water path of liquid clouds	
	Water path of ice clouds	
	Shortwave radiation effect	• Zhang et al. 1995, with updates (B. Cairns)

Polarimeter Processing Implementation Status

- Level-1A to Level-1B/C for SPEXone: initial code provided by SRON, testing with airSPEX data
- Level-1A to Level-1B/C for HARP-2: initial code provided by UMBC, testing with airHARP data
- Implementation of consensus Level-1C approach for OCI, SPEXone, HARP2 (presentation by K. Knobelspiesse) is underway.
- testing aerosol retrieval code from SRON with airSPEX data & simulated SPEXone L1C
- testing aerosol/ocean color simultaneous retrieval code (MAPOL, Gao et al. 2018, 2019) on RSP, airHARP, airSPEX data
- testing throughput on experimental parallel processing cluster (Poseidon)

Simulated OCI Data

MODIS-based Simulator

Process MODIS to Rrs/Chl/AOT/Angstrom. Replace water-leaving radiance and atmosphere LUTs with OCI hyperspectral models. Use retrieved aerosol info to select aerosol model from OCI LUT and propagate back to TOA. Fill-in the clouds and land with some realistic spectral shape scaled to the MODIS TOA radiances.

Computationally inexpensive. Useful for sensitivity testing of sensor artifacts, sanity check on algorithm implementation, SDS data flow and processing mechanics.

https://oceancolor.gsfc.nasa.gov/fileshare/sean_bailey/pace/PACE_OCI_SIM.20190321T002001.L1B.V4.nc

GMAO Simulator

Vector RT-based with atmosphere properties from GMAO nature run and water radiances from Gregg & Rousseaux model.

Computationally expensive. Independent data source for algorithm testing (cloud, aerosol, ocean color), as well as testing mechanics of SDS processing software.

Considerations for Algorithm & Product Advancement

Heritage algorithms for required products are flight-tested, and by design they can meet our mission requirements, but the goal is to go well beyond heritage.

Advanced algorithms that utilize multiple instruments are anticipated, but likely too risky for at-launch production (anticipate parallel development, provisional production).

Any algorithm implemented or operated by PACE SDS will be distributed as open source (NASA ESDS policy).

Computational capabilities within SDS are substantial, but still limited. Target for operational production is 22x throughput (22 data days per calendar day).

Thank You

SDS (OBPG) Distributed Processing System

OBPG Computing Facility, NASA Goddard Space Flight Center

Building 28 Room W220
- 900 square feet dedicated to OBPG

Linux Servers
- 111 processing nodes (**~1821 concurrent streams**)
- 79 RAID storage nodes (**~6.8 PB online storage**)
- Other support servers

Network
- Co-located with SEN network facility
- 10 gigabit SEN connection, internal backbone

Environment
- Primary and backup A/C units
- 225 KVA Eaton Powerware UPS with 4 PDUs
- 27 racks, hot aisle/cold aisle configuration

Minimizing downtime
- on-site spares, self-maintenance
- in-place upgrades, highly-available services,
- backup net connection

SDS (OBPG) Parallel Processing System

Management Node

- 192GB of RAM
- 2 (two) 4 core Xeon X5550 CPUs @ 2.67GHz for a total core count of 8 (16 threads)
- 53T of ZFS mounted to all compute nodes
- Running the CentOS 7.7 Linux Operating System
- Running SLURM & OpenMPI

Compute Nodes (currently 4, expandable)

- 4 x 1TB (with space for up to 2TB) for a total of 4 Terabytes of RAM (8TB Max)
- 4 x 2, 32 core CPUs for a total of 256 AMD Gen-1 EPYC cores (512 threads)
- Running the CentOS 7.7 Linux Operating System