National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology

SMAP SDS Resources for Data Product Validation

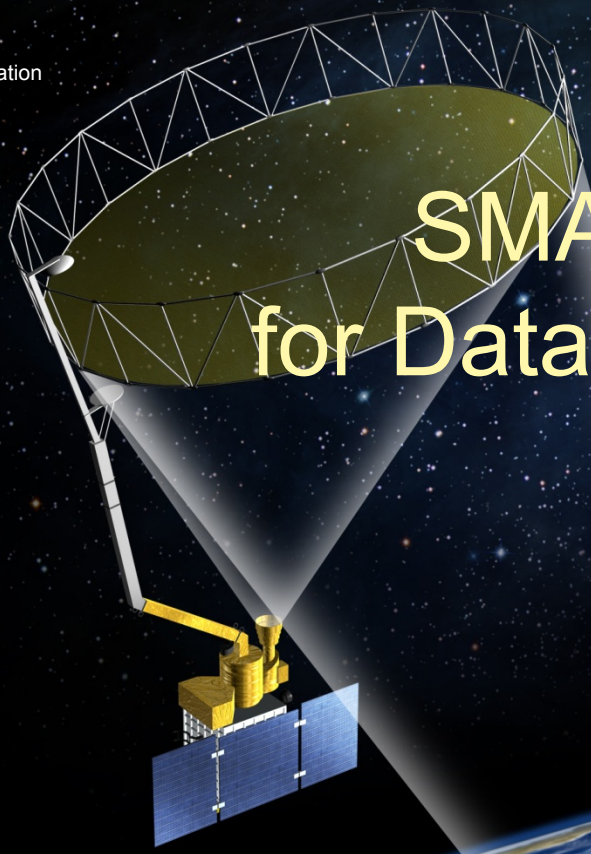
Soil Moisture Active Passive Mission

SMAP

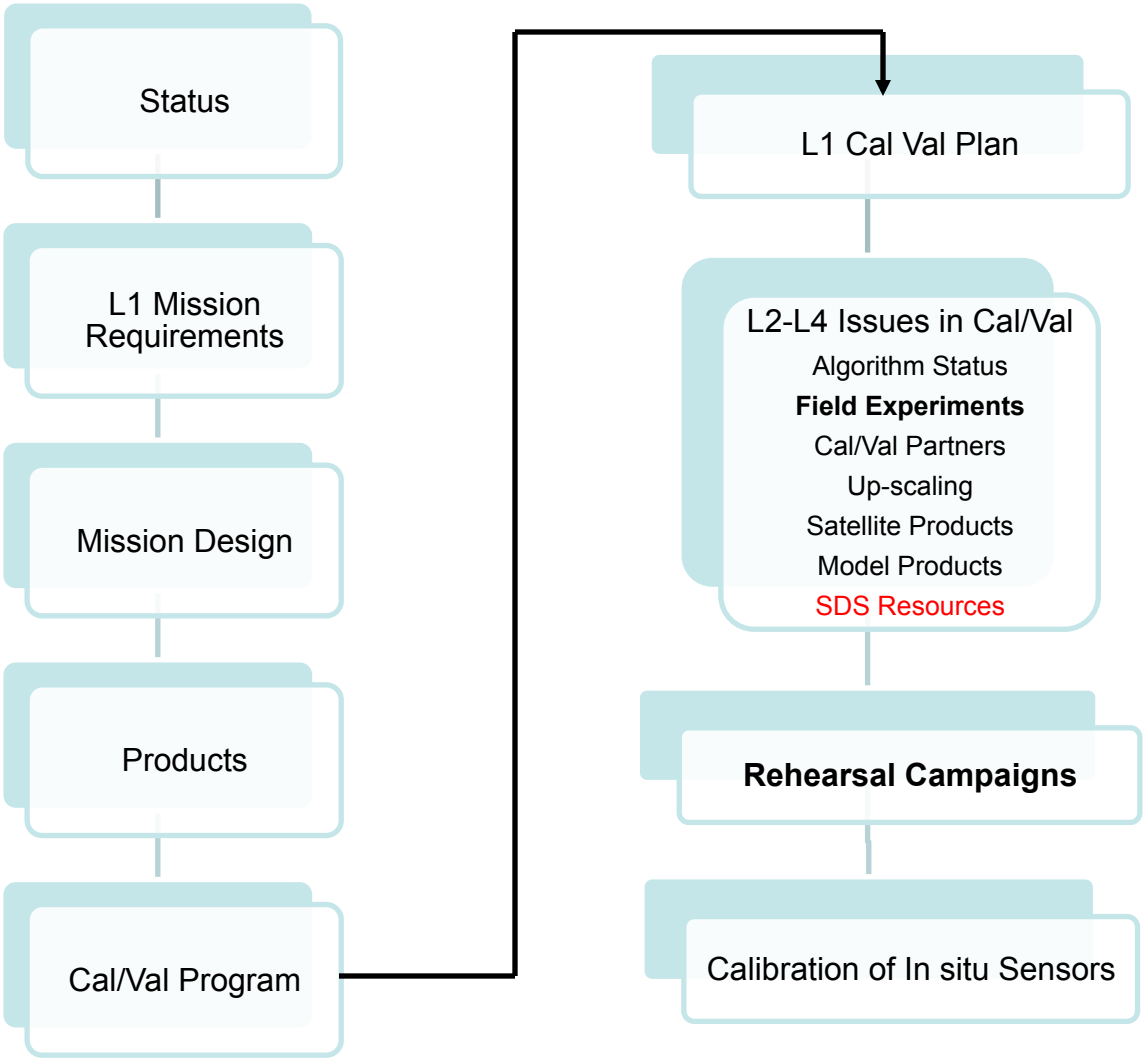
Barry Weiss
Jennifer Cruz
David Cuddy

3rd Cal/Val Workshop

Nov. 14-16, 2012



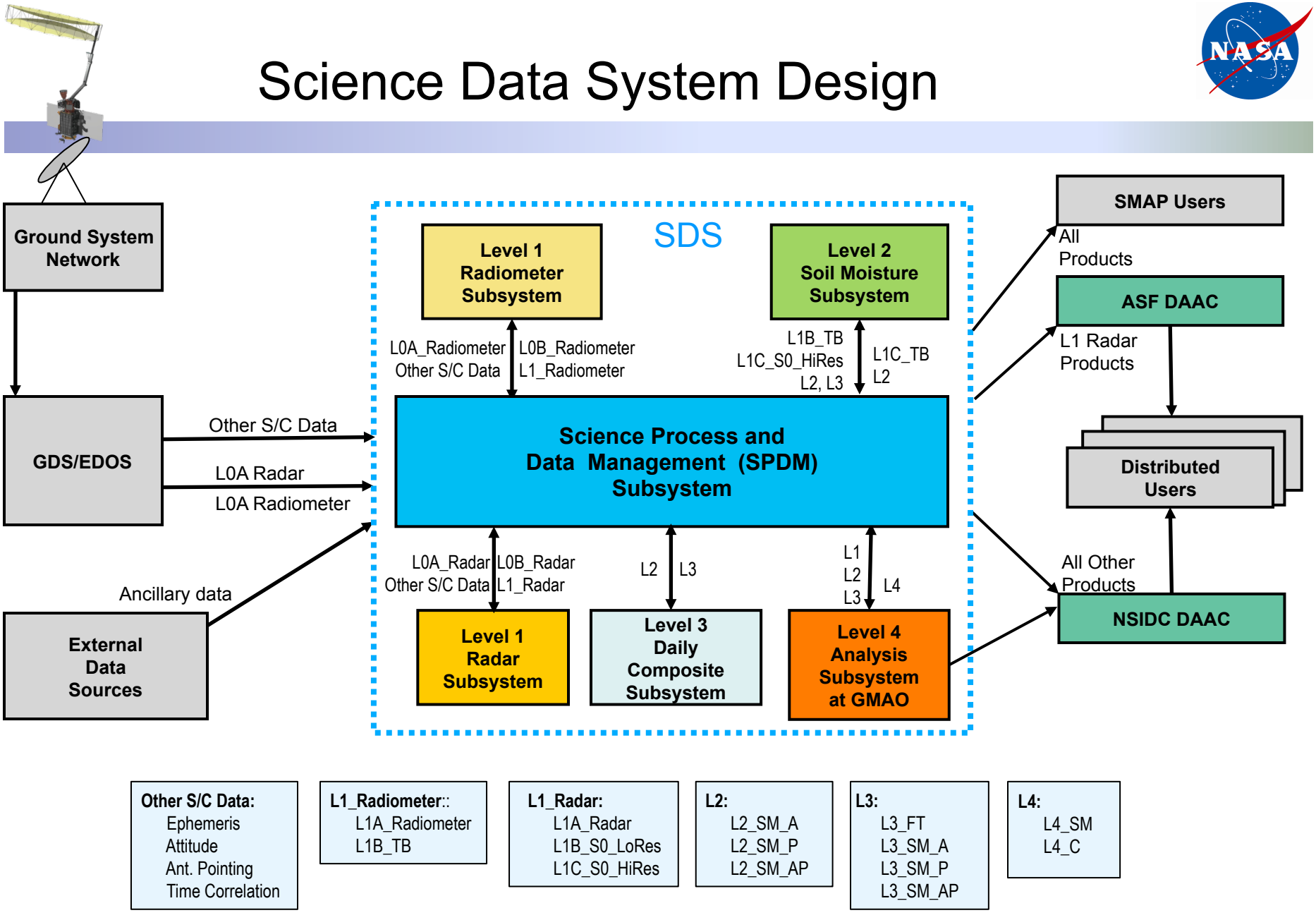
Workshop Overview



SDS Resources

- Science Data System Design
- Science Data System Environments
  - *Research and Analysis*
  - *Algorithm Development*
  - *Science Data Systems Operations*
- NASA ESDIS Validation Standards
- Data Provision

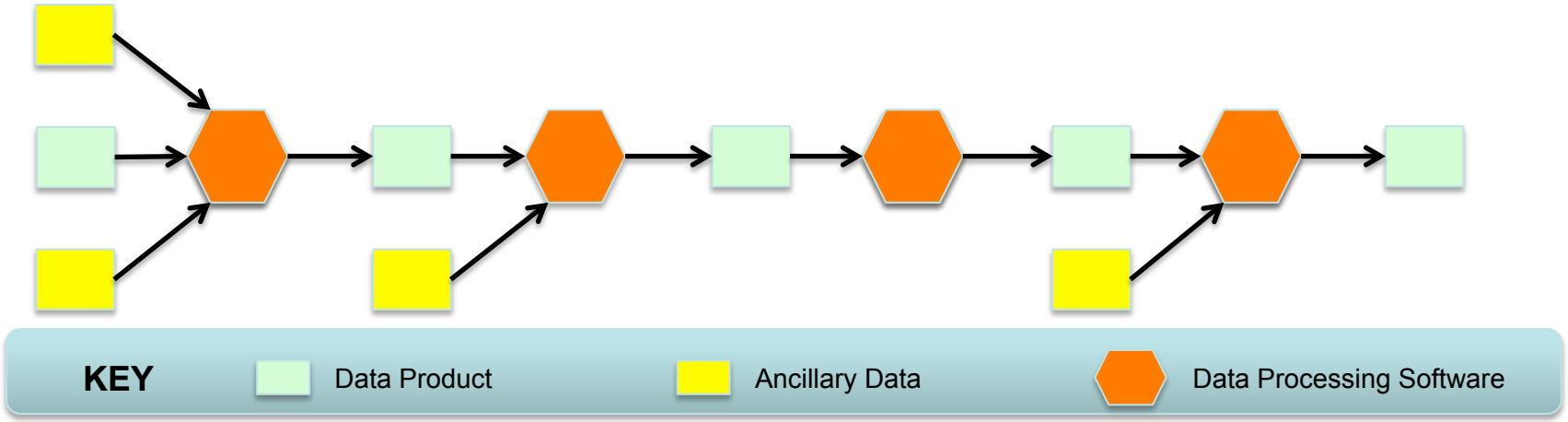
Science Data System Design



Other S/C Data:	L1_Radiometer::	L1_Radar:	L2:	L3:	L4:
Ephemeris	L1A_Radiometer	L1A_Radar	L2_SM_A	L3_FT	L4_SM
Attitude	L1B_TB	L1B_S0_LoRes	L2_SM_P	L3_SM_A	L4_C
Ant. Pointing		L1C_S0_HiRes	L2_SM_AP	L3_SM_P	
Time Correlation				L3_SM_AP	

Science Data Systems Operations

- Automated facility
  - Automatically ingests all telemetry and ancillary data required for data processing
  - Automatically runs data through complete pipeline from Level 0 to Level 3
  - Analogous system at GSFC processes the Level 4 data
- Content and function are under rigorous configuration management
  - Any change to operations requires approval from the Change Control Board
  - The system manages and tracks all versions of operating software
  - The system manages and tracks all versions of data products



SDS Environments

SDS supports five distinct computational environments. Three are of interest to the cal/val team:

Research and Analysis

Tools that assess current output, provide insight for correction and improvement

Software Development

Workspace to implement new software or existing software with modified parameters

Science Data Systems Operations

Automated Data Production for Product Users
Employs Rigorous Configuration Control

Science Data Systems Integration and Test

Ensures that new or modified software will function in operations environment

Offline Algorithm and Staging and Input System (OASIS)

Flexible Shadow of Science Data System Operations
Runs specific test algorithms and parameters in a selected stream with selected data

Research and Analysis Environment

- The cal/val team needs an environment where:
  - *They can assess and prepare validation data sets*
  - *They can execute code that provides effective data validation*
  - *They can research methods that have the potential to improve SMAP algorithms*
- The Research and Analysis Environment must:
  - *Access requisite validation data*
  - *Access requisite SMAP data products*
  - *Host software favored by individual cal/val team members*
    - Matlab, IDL…
- The SDS testbed at JPL that will host a Research and Analysis environment
  - *Will require an adequate number of Matlab, IDL, or other required analysis software licenses*
- Remote mission partners can configure hardware to host a local Research and Analysis environment

Software Development Environment

- Output from cal/val team activities will generate recommendations to change or enhance SMAP algorithmic software
- The ADT and SDS will employ the Software Development Environment to:
  - *Modify algorithms based on validation results*
  - *Test the behavior of the modified algorithms*
- The Software Development environment must host the following:
  - The same operating system used in SMAP Operations
  - The same suite of compilers used in SMAP Operations
  - The same version of HDF5 and other baseline product development software used in SMAP Operations
- The SMAP testbed will host a software development environment
  - *Access will be restricted to the Algorithm Development Team and the Science Data System Team*

NASA Earth Science Data Maturity Levels

Beta

- Products intended to enable users to gain familiarity with the parameters and the data formats.

Provisional

- Product was defined to facilitate data exploration and process studies that do not require rigorous validation. These data are partially validated and improvements are continuing; quality may not be optimal since validation and quality assurance are ongoing.

Validated

- Products are high quality data that have been fully validated and quality checked, and that are deemed suitable for systematic studies such as climate change, as well as for shorter term, process studies. These are publication quality data with well-defined uncertainties, but they are also subject to continuing validation, quality assurance, and further improvements in subsequent versions. Users are expected to be familiar with quality summaries of all data before publication of results; when in doubt, contact the appropriate instrument team.

NASA Earth Science Validation Levels

- Stage 1:
  - Product accuracy is estimated using a small number of independent measurements from selected locations/time/ground-truth/field-program efforts
- Stage 2:
  - Product accuracy is estimated over a significant set of locations/time-periods by comparison with reference in situ and other suitable reference data.
  - Spatial and temporal consistency of the products and with similar products has been evaluated over globally representative locations/time-periods
  - Published in peer-reviewed literature
- Stage 3:
  - Product accuracy has been assessed
  - Uncertainties and associated structure are well quantified against reference in situ and other suitable reference data
  - Uncertainties are characterized in a statistically robust way over multiple locations/time-periods representing global conditions
  - Spatial/temporal consistency of the products and with similar products has been evaluated over globally representative locations/time-periods
  - Published in peer-reviewed literature
- Stage 4:
  - Results for stage 3 are systematically updated when new versions are released and as time-series expands

* Reference: http://science.nasa.gov/earth-science/earth-science-data/, dated 17 Aug 2010

Science Data Validation and Delivery Timeline

In Orbit Checkout

Level 1 Cal/Val

Level 2/3/4 Cal/Val

Launch

Months since Launch

Science Operations

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

- Unreleased SMAP Products Available to the Cal/Val Team

- Unreleased SMAP Level 2 to 4 Products Available to Cal/Val Team
- Beta Level 1 Products Available to the Entire User Community

- Validated Level 1 Products Available to the Entire User Community
- Beta Level 2/3/4 Products Available to the Entire User Community

- Validated Data Products at all Levels Available to the the Entire User Community

Data Access

- Two major methods will be available for data access:
  - *The Data Centers will provide data products*
    - Cal/val team members will have access to data products during the cal/val period
      - Products generated during this period will not be distributed to the general user community
    - Data Centers will have a list of cal/val team members
      - Data Centers will provide access to a distribution site that requires a password
  - *SDS will set up an sftp site*
    - This site will require a password for access
    - The site will provide access to the validation data the cal/val team requires for research
    - The site will provide space for cal/val team members to share data and results
    - SDS and cal/val team leads will collaborate to organize the site and specify procedures

Validation Data Sets

- SDS will provide access to several categories of validation data sets at the sftp site. These include:
  - Output from various cal/val tools
  - Core Validation Site data – includes tower sites
  - Sparse Network data
  - Model Products – includes matchups
  - Field Campaign data
  - Alternative satellite data
  - Analysis products generated by the Operations System
- Cal/val team leads will manage the paths to validation data sets
  - SDS will provide guidelines for management of these data sets
- Cal/val team members will provide validation data sets
  - Arrange for automated ingestion for those that are frequently delivered
  - Specify delivery location for those that are infrequently delivered.

Data Provision for Phase I Rehearsals

- For Phase I Rehearsals
  - *SDS operational facilities will not yet be implemented*
  - *SDS will leverage existing hardware current in use for software development*
- To enable an exchange of information and data, the SDS will need to know
  - *Which cal/val partners have ftp sites and plan to use them*
  - *Which cal/val partners will need assistance to enable data transfer*
  - *Approximately how much data will be transferred to and from the site*
  - *What feedback information might the cal/val teams like to review?*
    - Results from upscales or matchups?
  - *Other foreseeable needs for Phase I rehearsals?*

Backup

Data Used in SMAP Operations

- SPDM manages the storage of data products and ancillary data that are required or generated by SDS Operations
  - Data generated by SDS operations
    - SMAP Data Products
    - QA Products and Analysis Products
    - Ancillary data used for standard processing
  - Ancillary data required for data processing and calibration
    - Instrument Data
    - Ancillary Data required for instrument processing and geophysical retrievals
    - Mission generated Parameter Files and Look-Up Tables
  - SDS and Science Team will configure means to make these data available based on access criteria
    - Level of cal/val involvement
    - Security and ITAR related issues

SMAP Data Products

Data Product Short Name	Short Description	Gridding (Resolution)	Latency*
L1A_Radar	Radar raw data in time order	-	12 hours
L1A_Radiometer	Radiometer raw data in time order	-	12 hours
L1B_S0_LoRes	Low resolution radar σ_o in time order	(5x30 km)	12 hours
L1B_TB	Radiometer T_B in time order	(36x47 km)	12 hours
L1C_S0_HiRes	High resolution radar σ_o (half orbit, gridded)	1 km (1-3 km)**	12 hours
L1C_TB	Radiometer T_B (half orbit, gridded)	36 km	12 hours
L2_SM_A	Soil moisture (radar, half orbit)	3 km	24 hours
L2_SM_P	Soil moisture (radiometer, half orbit)	36 km	24 hours
L2_SM_A/P	Soil moisture (radar/radiometer, half orbit)	9 km	24 hours
L3_F/T_A	Freeze/thaw state (radar, daily composite)	3 km	50hours
L3_SM_A	Soil moisture (radar, daily composite)	3 km	50 hours
L3_SM_P	Soil moisture (radiometer, daily composite)	36 km	50 hours
L3_SM_A/P	Soil moisture (radar/radiometer, daily composite)	9 km	50 hours
L4_SM	Soil moisture (surface & root zone)	9 km	7 days
L4_C	Carbon net ecosystem exchange (NEE)	9 km	14 days

* Mean latency under normal operating conditions. Latency is defined as time from data acquisition by the observatory to availability to the public data archive. The SMAP project will make a best effort to reduce these latencies.

** Over outer 70% of the swath.

SMAP Data Products

Data Product Short Name	Description	Pixel Resolution	Granule Extent
L1B_S0_LoRes	Low Resolution Radar σ_o in Time Order	5x30 km (10 slices)	Half Orbit
L1C_S0_HiRes	High Resolution Radar σ_o on Swath Grid	1 km	Half Orbit
L1B_TB	Radiometer T_B in Time Order	39x47 km	Half Orbit
L1C_TB	Radiometer T_B	36 km	Half Orbit
L2_SM_A	Radar Soil Moisture	3 km	Half Orbit
L2_SM_P	Radiometer Soil Moisture	36 km	Half Orbit
L2_SM_AP	Active-Passive Soil Moisture	9 km	Half Orbit
L3_FT_A	Daily Global Composite Freeze/Thaw State	3 km	North of 45° N
L3_SM_A	Daily Global Composite Radar Soil Moisture	3 km	Global
L3_SM_P	Daily Global Composite Radiometer Soil Moisture	36 km	Global
L3_SM_AP	Daily Global Composite Active-Passive Soil Moisture	9 km	Global
L4_SM	Surface & Root Zone Soil Moisture	9 km	Global
L4_C	Carbon Net Ecosystem Exchange	9 km	North of 45° N

SMAP Product Delivery

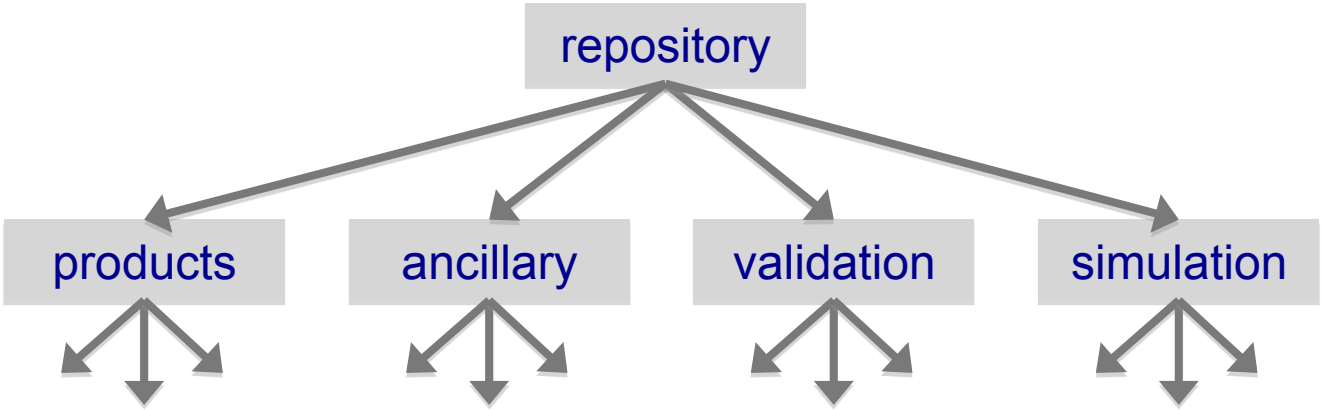
Data Product Short Name	Description	Initial Availability After Commissioning	First Validated Products after Commissioning	Latency to User Community after Acquisition
L1B_S0_LoRes	Low Resolution Radar σ_o in Time Order	3 months	6 months	12 hours
L1C_S0_HiRes	High Resolution Radar σ_o on Swath Grid	3 months	6 months	12 hours
L1B_TB	Radiometer T_B in Time Order	3 months	6 months	12 hours
L1C_TB	Radiometer T_B	3 months	6 months	12 hours
L2_SM_A	Radar Soil Moisture	6 months	12 months	24 hours
L2_SM_P	Radiometer Soil Moisture	6 months	12 months	24 hours
L2_SM_AP	Active-Passive Soil Moisture	6 months	12 months	24 hours
L3_FT_A	Daily Global Composite Freeze/Thaw State	6 months	12 months	50 hours
L3_SM_A	Daily Global Composite Radar Soil Moisture	6 months	12 months	50 hours
L3_SM_P	Daily Global Composite Radiometer Soil Moisture	6 months	12 months	50 hours
L3_SM_AP	Daily Global Composite Active-Passive Soil Moisture	6 months	12 months	50 hours
L4_SM	Surface & Root Zone Soil Moisture	6 months	12 months	7 days
L4_C	Carbon Net Ecosystem Exchange	6 months	12 months	14 days

SMAP Data Product Volumes

Data Product Short Name	Description	Daily Volume (GBytes)	Yearly Volume (TBytes)
L1B_S0_LoRes	Low Resolution Radar σ_o in Time Order	9.002	3.288
L1C_S0_HiRes	High Resolution Radar σ_o on Swath Grid	42.988	15.701
L1B_TB	Radiometer T_B in Time Order	1.958	0.715
L1C_TB	Radiometer T_B	0.472	0.172
L2_SM_A	Radar Soil Moisture	2.144	0.783
L2_SM_P	Radiometer Soil Moisture	0.015	0.006
L2_SM_AP	Active-Passive Soil Moisture	0.149	0.054
L3_FT_A	Daily Global Composite Freeze/Thaw State	2.349	0.343
L3_SM_A	Daily Global Composite Radar Soil Moisture	10.014	1.218
L3_SM_P	Daily Global Composite Radiometer Soil Moisture	0.046	0.006
L3_SM_AP	Daily Global Composite Active-Passive Soil Moisture	0.842	0.102
L4_SM	Surface & Root Zone Soil Moisture	16.284	1.784
L4_C	Carbon Net Ecosystem Exchange	0.604	0.026

Data Repository

- Four major data categories in repository
  - Each data data category has its own storage organization

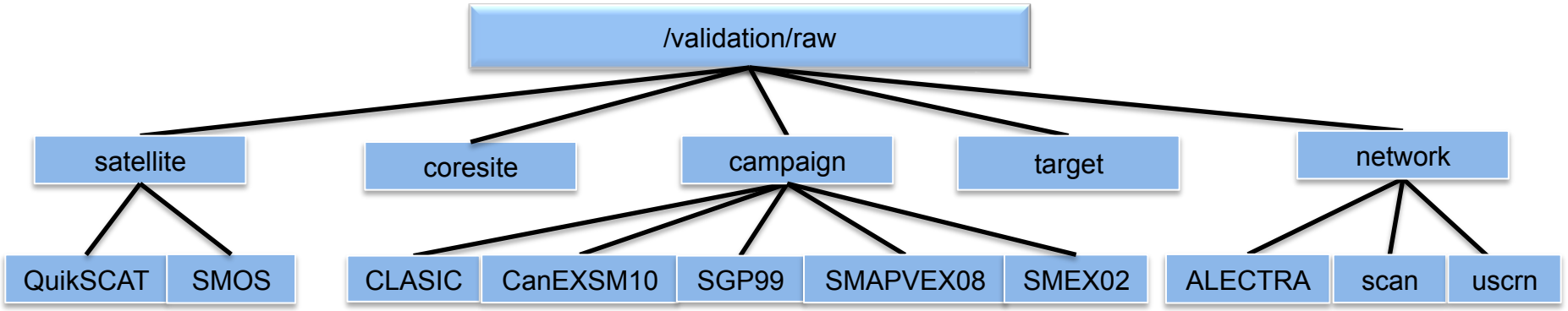


Life of Mission (LOM) Storage

- Life of Mission (LOM) Storage includes:
  - All SMAP input and output data products
  - All QA products
  - All run log files
  - All ancillary data required for processing in SDS operations
  - SPDM tracks and records all data in LOM Storage:
    - File creation information
    - Information about the job that created the file
- Direct access to LOM Storage will be limited
  - Most cal/val users will gain access to the data products in LOM Storage via the Data Centers
  - During Level 1 cal/val, the Data Centers will provide data products to cal/val participants before they are released for delivery to the user community

Tracking Validation Data Sets

- A model of the top level for storage of raw validation data
- A matching tree is available for processed and/or matchup validation data
- The tree was designed for science use
  - Will expand the model for validation data sets required by the instrument team
- Analogous directory trees will be constructed for processed validation data and matchup validation data
- Team members may place data in locations
  - README files are essential



Tracking Ancillary Data Sets

- Categorization of ancillary data:
  - By type – Vegetation index, Landcover, Soil Texture, Roughness
  - By source and usability – Raw data, Intermediate data and Operations data
- Raw and Intermediate Data appear under the /ancillary/masters directories
- Operations Data sets appear under the /ancillary/versions directories

Type	Raw Data and Intermediate Data	Operations Data
Vegetation Index	/ancillary/masters/vi/[category]/raw|proc	/ancillary/versions/vi/[category]
Landcover	/ancillary/masters/landcover/[category]/raw|proc	/ancillary/versions/landcover/[category]
Soil Texture	/ancillary/masters/soiltex/[category]/raw|proc	/ancillary/versions/soiltex/[category]
Roughness	/ancillary/masters/roughness/[category]/raw|proc	/ancillary/versions/roughness/[category]
….	/ancillary/masters/…./[category]/raw|proc	/ancillary/versions/…/[category]
….		

Tracking Ancillary Data Sets

- Categorization of ancillary data:
  - By type – Vegetation index, Landcover, Soil Texture, Roughness
  - By source and usability – Raw data, Intermediate data and Operations data
- Raw and Intermediate Data appear under the /ancillary/masters directories
- Operations Data sets appear under the /ancillary/versions directories

Type	Raw Data and Intermediate Data	Operations Data
Vegetation Index	/ancillary/masters/vi/[category]/raw|proc	/ancillary/versions/vi/[category]
Landcover	/ancillary/masters/landcover/[category]/raw|proc	/ancillary/versions/landcover/[category]
Soil Texture	/ancillary/masters/soiltex/[category]/raw|proc	/ancillary/versions/soiltex/[category]
Roughness	/ancillary/masters/roughness/[category]/raw|proc	/ancillary/versions/roughness/[category]
….	/ancillary/masters/…./[category]/raw|proc	/ancillary/versions/…/[category]
….		

Tracking Ancillary Data Sets

- Categorization of ancillary data:
  - By type – Vegetation index, Landcover, Soil Texture, Roughness
  - By source and usability – Raw data, Intermediate data and Operations data
- Raw and Intermediate Data appear under the /ancillary/masters directories
- Operations Data sets appear under the /ancillary/versions directories

Type	Raw Data and Intermediate Data	Operations Data
Vegetation Index	/ancillary/masters/vi/[category]/raw|proc	/ancillary/versions/vi/[category]
Landcover	/ancillary/masters/landcover/[category]/raw|proc	/ancillary/versions/landcover/[category]
Soil Texture	/ancillary/masters/soiltex/[category]/raw|proc	/ancillary/versions/soiltex/[category]
Roughness	/ancillary/masters/roughness/[category]/raw|proc	/ancillary/versions/roughness/[category]
….	/ancillary/masters/…./[category]/raw|proc	/ancillary/versions/…/[category]
….		

Simulated Data Sets

- Simulated data sets – mostly for pre-launch use
  - *Model SMAP products in HDF5 format*
  - *Level 1C Radar, Level 1B Radiometer and higher are based on GloSim2*
- These data sets will be critical for Phase II Rehearsals
- These data sets will be made available to the cal/val team before launch.
  - *National Snow and Ice Data Center*
  - *SMAP sftp site*

Overview of GloSim2

- Overview of full simulation
  - *Orbit sequence based on 8-day exact repeat cycle (117 orbits). Initial orbit of cycle has ascending node longitude 0° at crossing time 18:00 UTC.*
  - *LSM inputs provided by GMAO 1-year Global Nature Run. (Mv, T10) at 9km spatial/1-hour temporal resolution*
- New features in GloSim2:
  - *Radiometer Tb integration over full main beam (formerly to 3 dB FOV)*
  - *Multi-datacube implementation of the radar forward model, where datacubes are mapped to landcover/crop types.*
  - *Choice of 5 dielectric models (Dobson, Mironov [default], Wang/Schmugge, Hallikainen, Zhang/Shi); model defaults to Zhang/Shi for frozen soil (Tsurf < 0C).*
  - *10-day NDVI climatology is interpolated in time to compute VWC; choice of VWC algorithms (original Jackson quadratic formula, and Ray Hunt's VWC algorithm with stem factor [default]).*
  - *Surface roughness (RMS height in cm, s) map includes applied variances about the mean values for each landcover type.*
  - *Updated ancillary inputs (MODIS IGBP dominant landcover type, water body fraction, soil texture)*

Ancillary Data in GloSim2 SMAP Simulations

The table below lists ancillary datasets that are being used for SMAP GloSim2 simulations. They are reflective of the datasets that SMAP may use in Level 2 processing. The table does not reflect the ultimate choice of ancillary data sets for the operational Level 2 algorithms.

Data Type	Resolution	Size	Notes
Soil Moisture (Mv)	9-km EASE grid/hourly	25 MB	GMAO Nature Run
Surface Temperature (Ts)	9-km EASE grid/hourly	25 MB	GMAO Nature Run
NDVI climatology	1 to 36-km EASE grid/10-day (plus max/min grids)	1.02 GB max	Bindlish/Jackson; interpolated in time at the grid cell to use in VWC algorithm
Water Body Fraction	1 to 36-km EASE grid	0.51GB max	MODIS land/water database (generated by S. Chan)
Landcover type	1 to 36-km EASE grid	0.51GB max	Dominant IGBP class at each grid cell (generated by S. Chan)
Sand/Clay fraction, bulk density	1 to 36-km EASE grid	0.51GB max	See N. Das soil texture memo; data are scaled to byte type to save space
Crop type	1-km EASE grid	0.51GB	Based on USDA, Canada, and Europe crop databases; other regions assigned
Surface roughness (s)	1-km EASE grid	2.04GB	Based on Hydros/J. Johnson means, with added variance by IGBP class
Radar Datacubes	3-channel sigma0, function of roughness, dielectric constant, and VWC (280,280,140)	0.125GB each, 2.0 GB total	Currently 16 cubes including 4 crop types, assigned by IGBP/crop indices

GloSim2 Radiometer Simulation Flow

MATLAB LINE INPUT

Specify orbit ID range (1-5,400)

Run Sim Cycle

Run orbit/instrument sampling

- Read SMAP time in UTC @ 12-ms interval
- Read 3-dB FOV arrays of lat/lon's, antenna gain, and incidence

Read global 1km semistatic data

- Read annual max NDVI
- Read NDVI at two adjacent 10-day periods
- Interpolate NDVI for a given DOY; interpolate again within a single half orbit if necessary

Read global 1-km static data

- Sand Fraction
- Clay Fraction
- Bulk Density
- IGBP Dominant Land Cover
- Surface RMS Roughness

Read global 9-km dynamic data

- Read GMAO hourly 10-cm soil moisture; read two files within a single half orbit if necessary
- Read GMAO hourly 10-cm surface temperature; read two files within a single half orbit if necessary

Static Data

Dynamic Data

STEP 1: Sample at FOV lat/lon's

- 246,575-by-338 elements per half orbit

STEP 2: Water pixels within FOV

- Compute dielectric constant using Klein-Swift model
- Compute water TB using Fresnel equations, along with FOV incidence

STEP 3: Land pixels within FOV

- Compute dielectric constant using Mironov model (thawed state)
- Compute dielectric constant using modified Dobson model (frozen state)
- Construct VWC array using (a) interpolated NDVI, (b) annual max NDVI, and (c) IGBP-based stem factor
- Construct 'b' and 'omega' arrays from IGBP-based lookup table
- Construct 'h' array from 's' array using the h = 0.1 * s relationship
- Compute land TB using Fresnel equations, along with FOV incidence and tau-omega model

STEP 4: Antenna synthesis

- Integrate water TB and/or land TB according to FOV antenna gain. Result: one TB per boresight location

TB Calculations

Radiometer Level 1B File

Noise-free time-ordered radiometer H- and V-polarized TB observations output at every 12 ms

GloSim2 Radar Simulation Flow

CMD LINE INPUT
Start Cycle
Number of Cycles

Run Sim Cycle

FOR EACH ORBIT:
- Generate Truth File
- Generate L1C File
- Generate Header files

Make SMAP Truth

Make L1C Radar

Initialize Radar Forward Model
- Read Datacube cubeset configuration file
- Allocate memory for N cubes
- Read datacubes
- Set up landcover->datacube mapping

Read NDVI Climatology (10-day, 1km resolution)
- Initialize NDVI data structure
- Get first NDVI map pair based on starting date of simulation
- Read maxNDVI map needed for VWC algorithm

ORBIT SWATH SETUP
- Shift swath template for orbit
- Get data start/end times
- Compute 1-km and 9-km EASEgrid coordinates for swath cells

Make Swath Arrays From Truth Data
- Cell Lat/Lon
- 1-km EASEgrid col/row
- Flags
- True Sigma0/Kp

BLUR RADAR
- Variable-length window as function of cross-track location
- Average true sigma0 along row

Read Global Static Ancillary Data
(1 km resolution)
- Soil Texture Maps
  - Sand Fraction
  - Clay Fraction
  - Bulk Density
- Water Body Fraction
- MODIS Dominant Landcover
- Crop type Map
- Surface Roughness (RMS "s")

Static Data

Read GMAO Nature Run LSM Data (hourly, 9km resolution)
- Initialize LSM data structure at simulation start date/hour
- Soil Moisture (sfmc)
- Surface Temperature (Tsurf)

Dynamic Data

* Update NDVI & LSM data structures as needed

FOR EACH SWATH CELL:
- Collocate with ancillary data
- Collocate & interpolate NDVI*
- Collocate LSM data*
- Determine landcover type
- Compute VWC
- Run Radar Forward Model
- Compute Kp

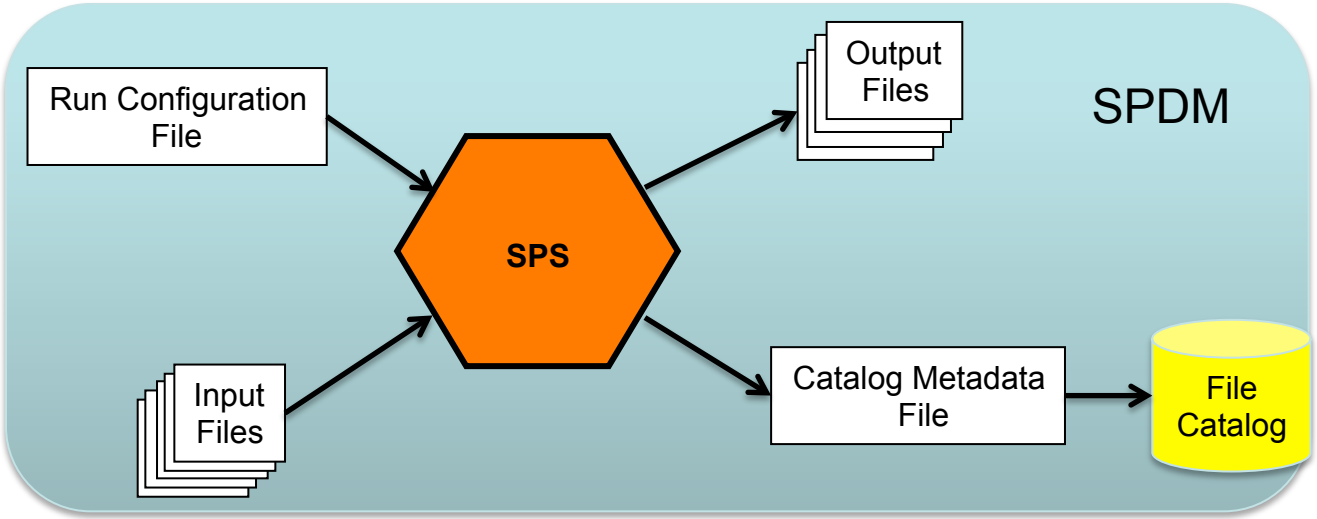
SMAP Truth File
1-km cells of collocated ancillary data and true σ_0 measurements

AT EACH CELL:
Generate fore & aft realizations of noisy sigma0 for HH,VV,HV channels

Radar Level 1C File
High resolution σ_0 on a 1 km SWATH Grid

SDS Integration and Test

- The Integration and Test environment provides a workspace to ensure that new or modified software will function in the SDS Operations Environment
  - Incorporates the full set of software and services used in SDS Operations on a smaller scale
  - Includes necessary adaptation to the Science Processing and Data Management (SPDM)



Offline Algorithm Staging and Input System (OASIS)