A Science Data System Approach For The SMAP Mission

David Woollard, Oh-ig Kwoun, Tom Bicknell, Richard West, Kon Leung
NASA Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive, Pasadena, CA 91109-8099

Abstract— **Though Science Data System (SDS) development has not traditionally been part of the mission concept phase, lessons learned and study of past Earth science missions indicate that SDS functionality can greatly benefit algorithm developers in all mission phases. We have proposed a SDS approach for the SMAP Mission that incorporates early support for an algorithm testbed, allowing scientists to develop codes and seamlessly integrate them into the operational SDS. This approach will greatly reduce both the costs and risks involved in algorithm transitioning and SDS development.**

I. Introduction

Soil Moisture Active Passive (SMAP) Mission is one of four missions recommended by the National Research Council (NRC) Earth Science Decadal Survey for launch in the 2010-2013 time frame [2]. The mission will use an L-band radar (active) and radiometer (passive) instrument to globally measure soil moisture and its freeze/thaw state. The mission inherits many concepts and characteristics from the former Hydrosphere State (HYDROS) mission with some differences in operations goals [4].

The science data system (SDS) is one of the key ground system elements of any science mission, including SMAP. The SDS is responsible for converting satellite downlink data into science data products for science and application users. While consideration of SDS design has traditionally begun only after science algorithms have matured, we have found that this practice introduces significant cost and risk when transitioning science code into the mission's operational SDS. Indeed, in surveying lessons learned from a number of previous JPL-led Earth science missions, the importance of early engagement in the SDS design effort has become increasingly apparent.

The SMAP mission represents the first example of a paradigm shift in which the SDS is well represented and actively engaged early in the mission conceptualization phase. In this paper, we present some of the major challenges for the SMAP SDS and some prospective solutions that would help meet those challenges in an effective and efficient manner.

II. SMAP MIssion Characteristics

The basic science requirements and instruments employed for SMAP are essentially unchanged from HYDROS [9]. Like HYDROS, SMAP carries both radar and radiometer instruments. The SMAP mission is required to produce estimates of surface soil moisture (top 5 cm depth) at 10 km resolution at a 3-day average interval. It is also required to produce estimates of land freeze/thaw transitions in the region north of 45° N, at 3-km resolution at a 2-day average interval.

The radar and the radiometer share the aperture of a 6-m deployable mesh reflector conically scanning at 14.6 rpm at a constant look-angle of 40°. This enables a 1000-km swath with a global revisit of 2~3 days. The L-band radiometer has V/H/U polarizations capable of 40 km resolution. The L-band radar has HH/HV/VV polarizations with 1~3 km resolution in SAR mode and 30 x 6 km resolution in real-aperture mode [9].

TABLE I. SMAP DATA PRODUCT SUITE

Level 0 Products	
L0b	Radiometer and LoRes/HiRes Radar
Level 1 Products	
Level 1 Radar Data Products	
L1B_S0_LoRes	LoRes Radar σ_O in Time Order
L1C_S0_HiRes	HiRes Radar σ_O, Gridded
Level 1 Radiometer Data Products	
L1B_TB	Radiometer TB in Time Order
L1C_TB	Radiometer TB, Gridded
Geophysical Science Data Products on Earth Grid	
L3_SM_HiRes	Radar Soil Moisture
L3_SM_40Km	Radiometer Soil Moisture
L3_SM_A/P	Radar/Radiometer Soil Moisture
L3_F/T_HiRes	Freeze/Thaw State
L4_F/T	Freeze/Thaw Model Assimilation on Earth Grid
L4_SM	Soil Moisture Model Assimilation on Earth Grid

To meet science requirements, there are a total of 13 different SMAP data products planned (see Table 1) [3]. The SMAP data product availability requirements vary with product use cases; with product turn-around time ranging from 12 hours (provisional Level 1 products mostly for science team use) to 12 months (production Level 4 products for public release) [3].

III. SMAP LEVEL 1 RADAR PRODUCTS GENERATION

SMAP Level 1 radar processing will produce two basic products. A low-resolution time-ordered product (L1B_S0_LoRes) will be produced continuously, providing complete coverage of Earth's surface at 30 km resolution. A high-resolution gridded product (L1C_S0_HiRes) will be produced at 1 km resolution for land areas. The low-resolution product is partially processed onboard the spacecraft to reduce the data volume transmitted by the spacecraft down to Earth. The high resolution SAR product is processed on the ground from high-rate raw data collected by the radar. Block floating point quantization (BFPQ) reduces the data volume by a factor of 2. The high-resolution data volume is further reduced by selecting only data taken during half of each conical scan, excluding a 300 km nadir gap out of a 1000 km swath width, and only over land areas. Both products are produced for two linear co-polarized channels (HH and VV), and one for cross-polarized echo data (HV).

Once on the ground, the high-rate data are processed through a standard SAR processing system. After BFPQ decoding, the radar samples are put through range compression followed by azimuth compression. Due to the short synthetic aperture time (32 ms), azimuth compression can be implemented with a relatively simple time domain process. Once range-doppler pixels are obtained, the image data are multi-looked and located on the Earth's surface. Radiometric calibration is applied using a combination of geometry data and measured system parameters. Noise-only data are also collected to calibrate out changes in system performance over time. The processing flows for both high- and low-resolution products are shown in Fig. 1. The low resolution processing flow does not include range or azimuth compression, but does include the geo-location and radiometric calibration steps.

The gridded product will likely be stored in a swath-oriented coordinate system with separate backplanes for key ancillary parameters such as the local incidence angle. The low-resolution time-ordered product will be stored as vectors of data that will include the location of each cell area.

IV. SMAP SCIENCE DATA SYSTEM (SDS) CHALLENGES

The SMAP SDS is responsible for the reduction of radar and radiometer data collected by SMAP to a variety of science data products for use by the science and application user communities. SDS functions commonly include data ingestion, data processing, data analysis, product verification, data/information management, and data/information storage. SDS development includes prototyping and developing scientific algorithms on testbeds by scientists and deploying those algorithms on operational systems by engineers. A functional diagram of the SMAP SDS is depicted in Fig 2.

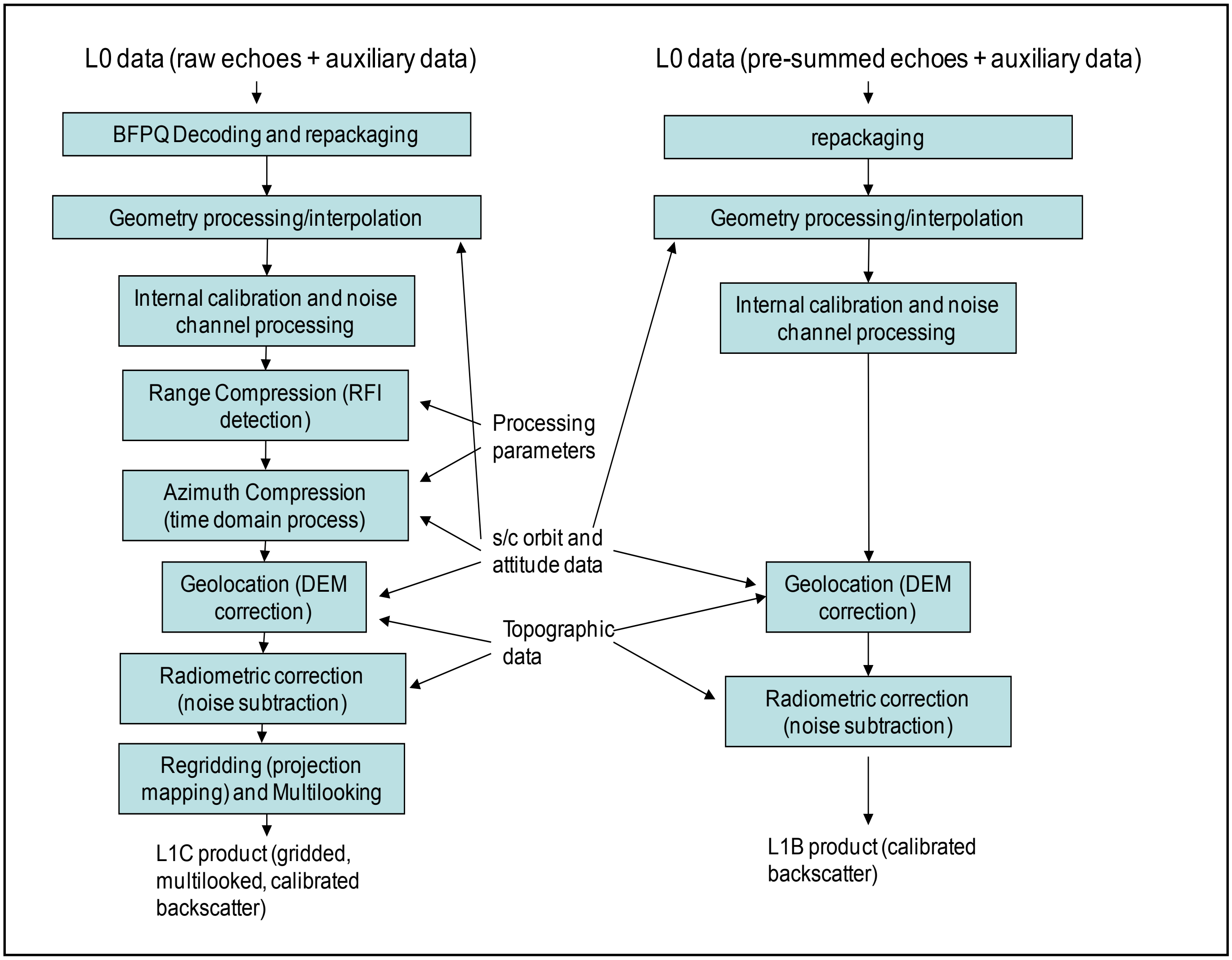

Figure 1. Processing flows for level 1 radar algorithms.

A. *Science Algorithm Testbed*

The primary challenge for SMAP SDS development is to reduce risks and costs in the implementation and deployment of the SDS.

Scientists developing new algorithms do so using methods and techniques to which they are accustomed. It is commonplace to see scientists prototype their algorithms in environments very different from the eventual operational environment of the SDS. SDS software developers are then required to either significantly modify or completely recode these prototype algorithms to fit them into the operational SDS infrastructure.

Considering the often rapid and iterative nature of the prototyping and development phases, the effort and process for integrating scientific algorithms into the SDS can become very costly and laborious. Development of the science data algorithms and their conversion to science executables (often called Product Generation Executives, or PGEs) have become the dominant factors in SDS development cost for most, if not all, recent Earth science missions [10].

A recent survey of existing Earth science mission's science data system development methodologies confirms that science algorithm development represents the lion share of science data system development costs; often taking up more than 50% of the overall SDS budget in the development phase [10].

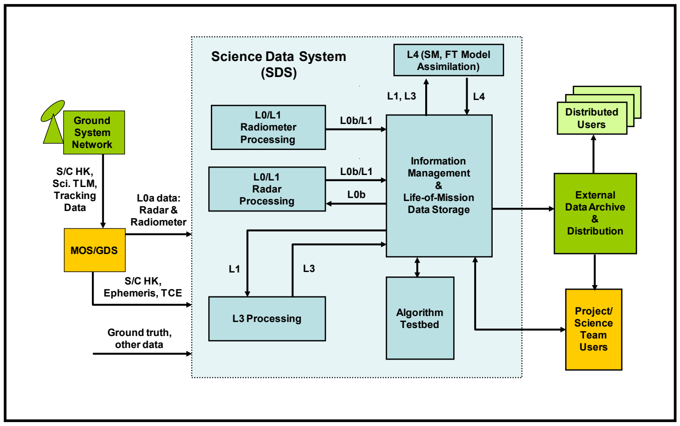

Figure 2. SDS Functional Block Diagram [6].

In order to reduce these development costs, as well as the inherent risks in porting, refactoring, and often rewriting scientific algorithms to make them "operational," we propose making available a science development environment that is as close to the full- scale target operational system as possible. This approach enables a rapid and seamless transition of prototype science code to operation in a manner that helps meet the cost and risk reduction challenges of SDS development. With the often repetitive, iterative, and *ad hoc* nature of science algorithm development, a proper testbed environment can also promote efficiency in algorithm development efforts with an effective information management system that tracks algorithm versions, inputs and the associated test results.

B. *Science Data System Framework*

In the past, mission SDS's were invariably custom built from scratch with little attention to re-use. This practice is responsible for the high cost and risk experienced by most missions.

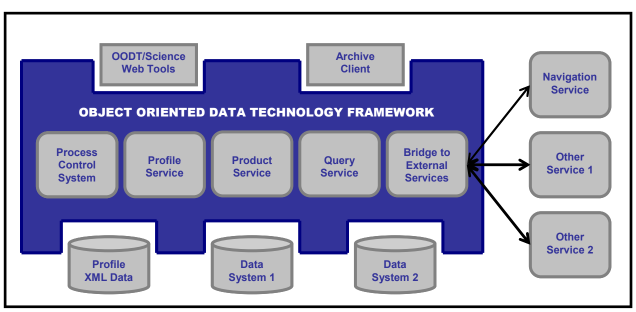

Figure 3. OODT Framework.

By adopting a common, componentized framework that can scale to mission needs and allow for easy technology insertion, missions can start to reap the benefit of lower costs and higher reliability through re-use of proven structures and components.

C. *OODT Framework for the SMAP SDS*

The OODT framework, initially developed with funding from NASA's Office of Space Science in 1998, is a well recognized software framework for sharing data across heterogeneous, distributed data repositories [1]. OODT ties together loosely coupled systems in a virtual data grid enabling data production, discovery, access, and distribution. Within the OODT framework (see Fig. 3), the Process Control System (PCS or formerly Catalogue & Archive Service - CAS), shown in greater detail in Fig. 4, is well suited for satisfying the SMAP SDS's complex data processing, storage, and information management functions.

PCS has been in continuous development at JPL and has been successfully deployed in a number of different scientific applications [11]. PCS incorporates three core components: (1) Workflow Management for modeling data processing tasks, (2) Resource Management for deployment of computation in cluster and Grid environments, and (3) File Management for metadata cataloging and data product archiving (See Fig. 4).

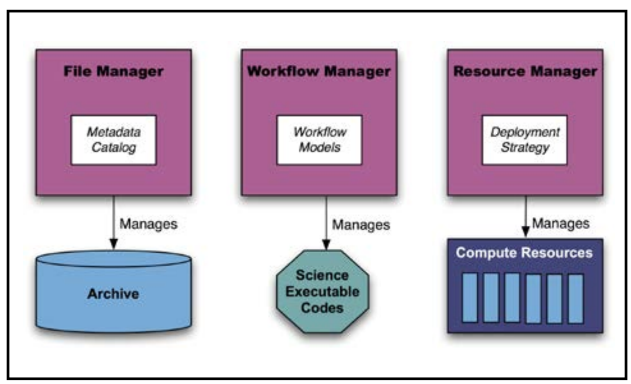

Figure 4. Process Control System components.

The mapping between the SMAP SDS functional blocks depicted in Fig. 2 and the PCS components in Fig. 4 can be described as follows: Each 'data processing' block (inclusive of L0 to L4 and the algorithm testbed in Fig. 2) represents a Product Generation Executable (PGE) that executes a particular science algorithm. Each PGE is connected to the PCS via a standard/common interface that is capable of accommodating virtually any type of PGE programming language without the need for modification to the PCS. These PGEs are executed and managed by the PCS Workflow Manager according to pre-defined processing workflow for each product. The PCS Resource Manager is responsible for the actual deployment of workflows, optimally matching tasks with available resources.

The 'Information Management' block in Fig. 2 is embodied by all three PCS components in orchestrating the various data products generation functions as well as managing and monitoring the associated data and information.

D. *Implementation Approach*

In the early stages of the mission, a scaled-down version of the PCS framework will be assembled for the science team to support algorithm development and evaluation activities. This replaces the conventional *ad hoc* algorithm testbed approach by putting in place early in the project a framework that benefits science algorithm development needs. This framework will be provided with design guidance to scientists developing SMAP science algorithms. PCS-supplied standards-based interfaces wrap scientists' code to provide a standard interface to PCS framework, which allows the scientist to leverage data management to annotate data with algorithm descriptors and other provenance information. In other words, scientists can conduct large-scale test runs, track anomalous behavior, and rerun experimental workflows, without inventing a framework for these activities.

This approach has multiple benefits. The PCS framework used for the algorithm testbed can be expanded to an operational SDS without additional development efforts. A science code can evolve and be integrated into the operational system without modifying the scientist's original interface. The PGE concept allows scientists to develop fine-grained software components that maximize the reuse of their code in any commonly used programming language. In addition, the PGE interface allows science algorithms to independently evolve without impacting software engineers' effort to improve the SDS infrastructure to meet the engineering challenges imposed by SMAP, including throughput, robustness, and scalability [11].

Furthermore, the PCS is platform independent. When multiple testbeds need to be deployed at different scientists' organizations, the framework of PCS and its design guidance for scientists provides a unified algorithm development environment across the science team, such that the risk and costs of developing and integrating science code into the SMAP SDS can be significantly reduced.

V. CONCLUSION

We have described a Science Data System approach for the SMAP mission that can greatly reduce risks and costs in SDS development. The proposed approach is based on a well-established software framework, OODT (Object Oriented Data Technology), which has been applied in support of missions such as OCO (Orbiting Carbon Observatory) and NPP Sounder PEATE. Leveraging OODT's rich heritage is expected to significantly reduce the risks and costs of SDS implementation for SMAP through re-use and lessons learned. Another keystone of this approach is the establishment of an algorithm testbed using the operational framework early in the project cycle. Allowing science algorithm development to be accomplished in the same environment as the operational system, the risks and costs for migrating the development science codes to the operational PGE's can be greatly reduced.

ACKNOWLEDGEMENT

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

[1] OODT Website, http://oodt.jpl.nasa.gov/.

[2] "Earth Science and Applications from Space - National imperatives for the next decade and beyond," NRC Committee on Earth Science and Applications from Space: A Community Assessment and Strategy for the Future, The National Academic Press, 2007.

[3] T. Bicknell, K. Leung, and B. Vollmer, "Soil Moisture Active and Passive (SMAP) Mission: Mission Concept Review - Science Data System," NASA JPL, June 24, 2008.

[4] K. Kellogg, "Soil Moisture Active and Passive (SMAP) Mission: Mission Concept Review - Project Overview," NASA JPL, 2008.

[5] M. Lapointe and R. Girard, "HYDROS L1C_S0_Hi_Res data product Algorithm Roadmap," Canadian Space Agency, March 8, 2005.

[6] K. Leung, "SMAP Science Data System (SDS) Phase A Budget Review," NASA JPL, 2008.

[7] P. O'Neil, E. Njoku, and J. C. Shi, "HYDROS 40-km Earth-Gridded Soil Moisture (L3-40km_SM) Algorithm Roadmap (draft)," NASA GSFC, JPL, & UCSB, April 7, 2005.

[8] M. Spencer, J. Piepmeier, S. Chan, and F. Rogez, "Hydros Instrument Design," in *HYDROS Science Team Meeting*. San Francisco, 2004.

[9] M. Spencer and C. Richardson, "Soil Moisture Active and Passive (SMAP) Mission: Mission Concept Review - Instrument," NASA JPL, 2008.

[10] D. Woollard, D. Freeborn, E. Kay-Im, and S. LaVoie, "Case Studies in Science Data Systems: Meeting Software Challenges in Competitive Environments.," presented at The 10th International Conference on Space Operations, Heidelberg, Germany, 2008.

[11] D. Woollard, N. Medvidovic, Y. Gil, and C. Mattmann., "Scientific Software as Workflows: From Discovery to Distribution.," *IEEE Software - Special Issue on Developing Scientific Software*, 2008.