

A Science Data System Approach for the DESDynI Mission

Oh-ig Kwoun, David Cuddy, Kon Leung, Dan Crichton, Chris A. Mattmann, Dana Freeborn
Jet Propulsion Laboratory, California Institute of Technology
Pasadena, USA
kwoun@jpl.nasa.gov

Abstract—Amongst the many key challenges to the Science Data System (SDS) for the DESDynI (Deformation, Eco-system Structure, and Dynamics of Ice) mission is the exceptionally large data volume (on the order of 5 tera-byte per day) acquired by the radar and the consequent huge volume of data products produced (on the order of 16 peta-bytes per year). This paper presents an SDS conceptual approach to effectively and efficiently support the mission. The features of this SDS approach include: 1) A modular functional architecture that is based on the proven Object Oriented Data Technology (OODT) based framework, 2) the utilization of a Testbed Concept that facilitates the morphing of scientific algorithms to operational codes, and 3) innovative data staging, storage and backup strategies. This SDS approach is expected to form a strong basis for helping DESDynI achieve its many science goals and objectives.

I. INTRODUCTION

DESDynI stands for Deformation, Eco-system Structure, and Dynamics of Ice and is one of four first-tier Earth Decadal Survey Missions (DS) identified in the National Research Council (NRC) Earth Science Decadal Survey [1]. It is the second DS mission assigned for execution at JPL, succeeding another DS mission SMAP (Soil Moisture Active Passive) that started in 2008. DESDynI is currently in the midst of its pre-mission study [2,3,7] and Mission Concept Review (MCR) preparation, leading to a potential project start in 2010.

DESDynI poses an unprecedented set of challenges to the data processing and handling system due to the sheer amount of data it collects with its radar and lidar instruments, and the complicated sets of data products it generates serving three major scientific disciplines. DESDynI's radar is expected to collect upwards of 4.9TB (tera-byte) of radar signal data a day, which in turn generate approximately some 16PB (peta-byte) of data products per year. These volumes represent two and three orders of magnitude increase, respectively, in comparison with SMAP. A versatile data system architecture and an efficient operations concept are needed to effectively handle and process this amount of data consistently and without backlog. To maximize the utility of the data set, DESDynI plans to work closely with the Earth Science Data System Working Groups (ESDSWG) in its selection and

adoption of data product standards and design, and in its selection of components for reuse (as we will elaborate later in the paper). The goal is to maximize to the greatest extent DESDynI data products' utility to the science and applications communities in concert with science data products from other DS missions.

In the following Sections, we will describe the highlights of the DESDynI mission, the DESDynI data system challenges, the proposed data system architectural concept and design, the DESDynI data system operations concept, and some of the implementation specifics and considerations.

II. THE DESDYNI MISSION DESCRIPTION

A. The Mission

A brief DESDynI mission overview is included in Figure 1. DESDynI is a 3 to 5-year mission targeted for launch in the 2018 time frame. DESDynI is dedicated to support scientific research and applications in three major disciplines relating to geosphere (deformation), biosphere (ecosystem structure), and cryosphere (dynamics of ice). DESDynI employs two principal instruments, a high-resolution L-band InSAR operating at a 761 km 8-day repeat orbit and a Lidar operating at a 400 km 91-day repeat orbit. The two instruments are flown independently on their respective spacecraft.

An overview of major characteristics of the SAR instrument is depicted in Table I. A novel approach in radar design and its operations called SweepSAR promises to allow routine polarimetric radar data acquisition in wide swath. The

Figure 1. DESDynI Mission Overview

SweepSAR allows steering of the radar beam in the range dimension to cover a wide 350 km swath. Within the swath, multiple modes of imaging are possible, enabling individual sub-swath to be acquired in single, dual, or quad-polarizations. The spacecraft is maneuverable to achieve Zero Doppler as well as left-/right-looking antenna pointing, thereby enabling 3-D deformation mapping.

TABLE I. DESDYNI SAR INSTRUMENT CHARACTERISTICS

Altitude	761 km
Elevation angle	27~43° (Mechanical bore sight: 35°)
Antenna	15 m diameter mesh reflector
Operating Frequency	L-band / 1250 MHz
Bandwidth	25 MHz (5MHz for RFI mitigation)
Field of Regard	350 km
Instance swath	Dual-pol: up to 350 km / Quad-pol: ≤ 220 km
# of beams	32 per pol
PRF	3~3.2 KHz (Quad) / 1.5~1.7 KHz (Dual)
Nominal resolution	8m x 10m
Number of looks	≥ 100 at a 100 m resolution
8-day repeat cycle	
Zero Doppler steering & Right-/Left-Looking operations	

B. The DESDynI Science Data Products

There are currently 28 DESDynI radar and radar/lidar fusion products identified based on the EOS data product level definition [4]. They are:

- (1) Level 0b product consisting of formatted signal data
- (9) Level 1 products including single-look (complex & detected), multi-look (complex, detected & browse), interferogram, and coherence data
- (13) Level 3 products including deformation maps, ice and glacier velocity maps, and geocoded Stoke’s matrix

products

- (5) Level 4 radar/lidar data fusion products including sea-ice thickness, canopy closure, forest structure, and biomass maps

All data products are expected to be produced and stored in a swath-based format. Validated data is released to designated DAAC’s (Distributed Active Archive Centers) for long-term archive and distribution. Amongst the DESDynI data products, it is anticipated that geocoded detected images and quick-look images are of most interest to prospective applications/operations users.

III. DESDYNI SCIENCE DATA SYSTEM

A. Science Data System Challenges

The main responsibility of the DESDynI Science Data System (SDS) is the reduction of L0 radar signal data into the 28 different types of data products. Its primary functions are essentially data ingestion, data products generation, data storage, and data distribution within the project. Its main external interfaces are with the ground data system that provides L0 raw instrument data and any ancillary data such as orbit information, with the DAAC(s) that receives the SDS produced data products for long-term archive and distribution to the public, with the instrument team to collaborate on instrument and processor calibration, and with the science team for providing data product access and supporting product verification and validation (V&V) activities.

In addition to handling and producing large volumes of data and products, significant challenges to the DESDynI SDS include designing and adopting an architecture that can readily accommodate participation of various expert institutions, an operations concept that facilitates SDS in meeting its goals and objectives reliably and cost effectively, and data format and standards that promote utilities of the DESDynI data sets.

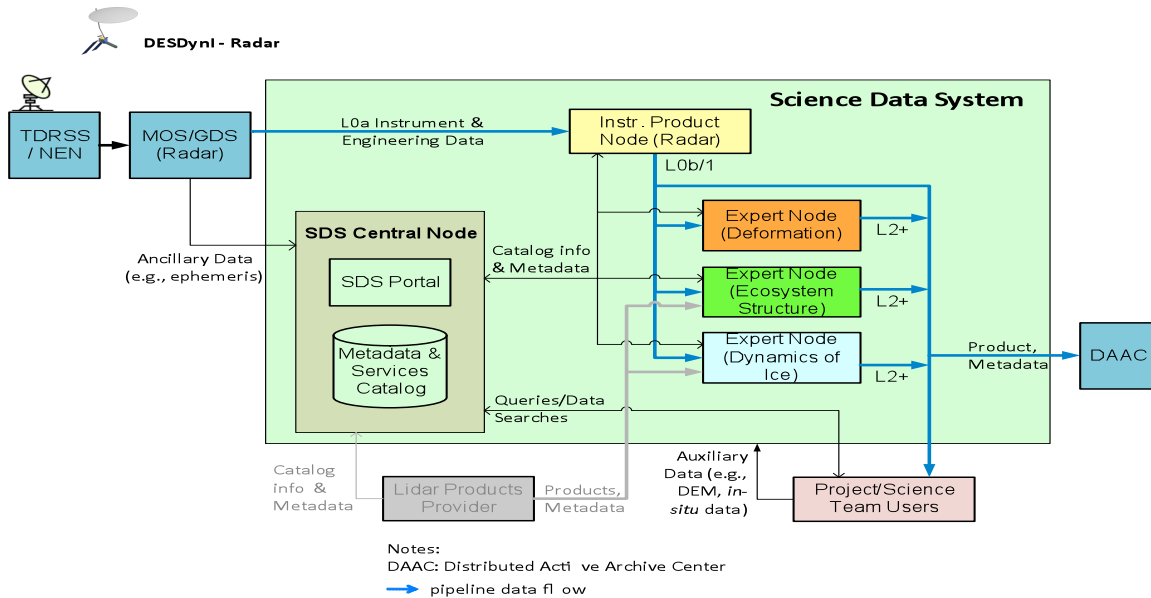


Figure 2. SDS Functional Diagram depicting Modular Architecture

B. DESDynI Science Data System Conceptual Design

The DESDynI SDS has adopted a modular architecture as depicted in Figure 2. The main attributes of the architecture are the dedicated functional nodes for supporting specialized discipline science processing, the distributed data storage that minimizes movement of high volume data within the SDS, and the centralized information access that allows single-point data and info queries and requests. This modular architecture can be implemented in a variety of ways from fully distributed to fully centralized and any hybrid combinations in between. Furthermore, each node of the SDS shares a common architectural framework that is adapted from the evolving OODT-based (Object Oriented Data Technology) [6, 7] PCS (Process Control System) product line [5] which allows for deployment of an SDS over a data grid implementation. This same architectural framework has been successfully adapted and deployed on a number of JPL missions and has recently been demonstrated on the SMAP Testbed [8]. The PCS is earmarked for deployment on all JPL-led Earth DS missions to reduce risks and costs.

In DESDynI's modular architecture (see Figure 2), the Central Node (CN) is responsible for managing information and resources such as information on data, computing, services, etc. across the entire SDS. Product Generation Nodes (PGNs) are composed of the Instrument Product Node(s), the Deformation Node(s), the Ecosystem Structure Node(s), and the Dynamics of Ice Node(s). These PGNs are responsible for data product generation, maintaining and providing catalog information to the CN, hosting a local "life-of-mission" data storage, and executing product delivery as requested via the CN. Although the current architecture accommodates spawning multiple instances of the PGNs, all catalog and metadata information residing at each individual node is made assessable via the CN. In essence, project users can access all data and data information residing in the SDS via a single portal.

Each PGN shares a common architecture depicted in Figure 3. With the exception of the CN, all PGNs are equipped with PGEs (Product Generation Executables) and Testbeds. PGEs are 'processing' units that accomplish data product generation or major processing steps. The Testbeds support algorithm and software development, and are configured to possess the same identical processing environment as the operational PGEs and processing pipeline. Based on early experience with SMAP, this Testbed approach has been found to be extremely effective in reducing costs and risks associated with the process of morphing science algorithms into operational codes [8]. In addition, the Testbeds also serve as platforms for conducting system integration and testing (I&T), product V&V as well as product and data analysis.

C. DESDynI Science Data System Operations Concept

The DESDynI SDS is envisioned to operate in a 'lights-dim' environment; with 'unattended' automatic data processing occurring 24 X 7, augmented by 8 X 5 normal office-hour-like data analyst and operator support. The dedicated processing pipeline operates autonomously, executing data/policy-driven processing steps to carry out

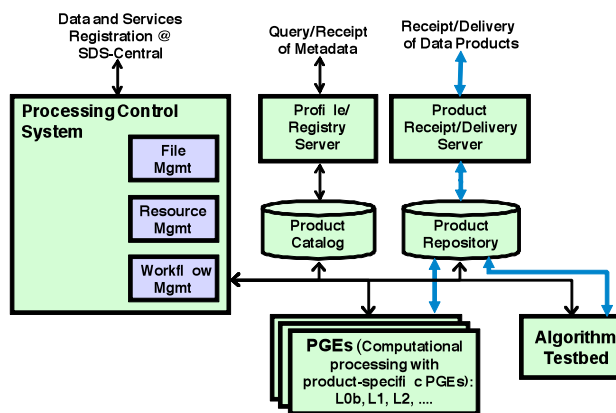


Figure 3. SDS Functional Diagram of Product Generation Nodes

normal processing and to conduct reprocessing campaigns. Product quality assurance is achieved with 'in-line' data quality checks performed on each product, and augmented by selective in-depth data quality analysis performed by data analysts. Pipeline performance, status, and operations statistics are automatically gathered for monitoring the system's health and anticipating potential operations issues and problems.

D. Data Storage and Backup Strategies

A project life-of-mission data storage made up of a federation of local storage at each PGN is maintained to provide on-site staging of intermediate products and ready access of data products by the project teams. However, with the current estimated downlink volume of 4.9 TB per day, storing all data products becomes cost prohibitive. L0a data alone can reach 2 PB per year and single-look complex (SLC) products would add another 7 PB per year. To address the high costs associated with staging and storing DESDynI's large data volume, innovative data storage and backup strategies will be needed. By adopting a rolling storage scheme in combination with judicious reprocessing/regeneration of products and utilization of DAAC(s) as an integral component of data backup, the data storage and backup volume and therefore the associated data storage costs can be markedly reduced. Table II summarizes the storage and backup strategies that are in consideration for the DESDynI SDS [9].

TABLE II. DATA STORAGE AND BACKUP STRATEGIES

Data Product		Primary Storage	Backup Storage	Archived @ DAAC
L0a	Depacketized raw bit stream; no processing done ¹	Rolling storage for most recent 6 months	Most recent 6-month	No
L0b	Formatted ² raw data suitable for distribution to public users	Full volume LOM storage	Most recent 30-day	Yes
L1B	SLC products	Rolling storage of 1 year; on-	Most recent	No

		demand product re-generation	30-day	
L1B	MLC products	Full volume LOM storage	Most recent 30-day	Yes
L1C	Interferograms	Full volume LOM storage	Most recent 30-day	Yes
L3+	High-level science products	Full volume LOM storage	Most recent 30-day	Yes

- 1) Generally not usable by the public without proper L0a-to-L0b conversion software.
2) This is a reformatted version of L0a with raw data artifacts cleaned-up or removed.

IV. SUMMARY

We have described a Science Data System approach for the DESDynI mission that is modular in functions, and scalable and flexible in implementation. It's an approach based on the proven OODT framework for reducing risks and costs in system development, allowing for deployment across a highly distributed environment, and enhancing reliability and efficiency in system operations. We have outlined an Operations Concept designed with the goal of effectively and efficiently meeting the high demands in day-to-day processing operations. We have also highlighted a Testbed approach that can significantly reduce the costs and risks associated with transforming the many DESDynI science algorithms into operational codes. We believe this SDS approach forms a strong basis for helping DESDynI achieve its many science goals and objectives.

ACKNOWLEDGMENT

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] "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.
- [2] D. Cuddy, D. Freeborn, and P. Callahan, "DESDynI Science Data System (SDS) pre-MCR Peer Review," *internal document*, NASA JPL, Nov. 2009.
- [3] W. T. Johnson, "DESDynI Radar pre-MCR Peer Review: Radar Introduction", *internal document*, NASA JPL, Nov. 2009
- [4] GSFC/NASA, "EOS Data Product Levels", <http://observer.gsfc.nasa.gov/sec3/ProductLevels.html>
- [5] C. Mattmann, D. Freeborn, D. Crichton, B. Foster, A. Hart, D. Woollard, S. Hardman, P. Ramirez, S. Kelly, A. Y. Chang, C. E. Miller. A Reusable Process Control System Framework for the Orbiting Carbon Observatory and NPP Sounder PEATE missions. In *Proceedings of the 3rd IEEE Intl' Conference on Space Mission Challenges for Information Technology (SMC-IT 2009)*, pp. 165-172, July 19 - 23, 2009.
- [6] C. Mattmann, D. Crichton, N. Medvidovic and S. Hughes. A Software Architecture-Based Framework for Highly Distributed and Data Intensive Scientific Applications. In *Proceedings of the 28th International Conference on Software Engineering (ICSE06)*, pp. 721-730, Shanghai, China, May 20th-28th, 2006.
- [7] D. Crichton, J.S. Hughes and S. Kelly. A Science Data System Architecture for Information Retrieval. In *Clustering and Information Retrieval*. Kluwer Academic Publishers, December 2003.
- [8] D. Woollard, O. Kwoun, T. Bicknell, R. West, and K. Leung, "A Science Data System Approach For The SMAP Mission", presented at RadarCon09, Pasadena, CA, May 2009.
- [9] D. Cuddy, O. Kwoun, and K. Leung, "DESDynI Operations Concept pre-MCR Peer Review: Science Data Processing and Distribution," *internal document*, NASA JPL, Oct. 2009.