A Reusable Process Control System Framework for the Orbiting Carbon Observatory and NPP Sounder PEATE missions

Chris A. Mattmann, *Member, IEEE*, Dana Freeborn, Dan Crichton, Sean Hardman, Paul Ramirez, Sean Kelly, David Woollard, Andrew Hart, Brian Foster, Albert Y. Chang, Charles E. Miller

***Abstract*—We describe a reusable architecture and implementation framework for managing science processing pipelines for mission ground data systems. Our system, dubbed "PCS", for Process Control System, improves upon an existing software component, the OODT Catalog and Archive (CAS), which has already supported the QuikSCAT, SeaWinds and AMT earth science missions. This paper focuses on PCS within the context of two current earth science missions: the Orbiting Carbon Observatory (OCO), and NPP Sounder PEATE projects.**

***Index Terms*— PCS, OODT, software architecture, grid.**

I. Introduction and objectives

To support the stringent data management and processing requirements (schedule 10s of 1000s of jobs/day, manage terabytes of data) levied upon us by the Orbiting Carbon Observatory (OCO) and NPP Sounder PEATE missions, we were forced to re-think our existing software implementation and design of the OODT Catalog and Archive Service (CAS) software component [1], a proven software solution which has been used on prior science missions incl. Seawinds/QuikSCAT, and the Active Mirror Telescope (AMT). Our main objective was to split CAS into two separate components representing its canonical capabilities: *file manage*r, which ingests files and catalogs their metadata, and *workflow manager*, which executes sequences of Product Generation Executives (PGEs) and models their control flow and data flow. We also augmented CAS with a new component, *resource manager* that allows CAS to manage and use heterogeneous computing resources including high performance clusters, grids and commodity hardware.

II. Approach

We upgraded the CAS by leveraging open source technologies from the Apache Software Foundation (ASF) to support our new interfaces, and design decisions. When open source components did not exist for particular functionality that we needed, we wrote small instances of glue code, tying together the open source components into a unified, easily maintainable and evolvable system. XML-RPC, a lightweight version of the popular SOAP/Web Services implementation substrate, was used as the communication mechanism between our three canonical CAS components and client interfaces.

Manuscript received November 3, 2008. This work was supported by the Jet Propulsion Laboratory, managed by the California Institute of Technology.

C. A. Mattmann, D. Crichton, S. Hardman, D. Freeborn, P. Ramirez, S. Kelly, D. Woollard, A. Hart, B. Foster, A. Y. Chang and C. E. Miller are with the Jet Propulsion Laboratory, Pasadena, CA 91109 USA (corresponding author: C. Mattmann, hone: 818-354-8810; fax: 818-393-1370; e-mail: chris.a.mattmann@jpl.naa.gov).

III. Results

Our work has directly benefited OCO and NPP Sounder PEATE. NPP Sounder PEATE has already used the file manager and workflow manager components to ingest and process hundreds of gigabytes of IASI data (and is in preparation to accept CRIMS data). On OCO, the mission is using the file manager to ingest MODIS, CloudSat and other ancillary data products for use in the high performance Level 2 Science Algorithm. To date, OCO has already used new CAS software to process over four terabytes of Fourier Transform Spectrometer (FTS) data provided by ground-based instruments located (inter-)nationally, and has used the software to process all data taken during Instrument Thermal Vacuum (TVAC) testing.

IV. Impact and Benefits

Traditional science data processing systems have required extensive development efforts that have built customized software in order to use and execute the science data processing algorithms. Our new CAS componentized architecture provides a framework for plugging in and managing the entire science-processing pipeline, enabling extensive reuse and prescribing an architecture on which future systems can be built. With OCO and NPP Sounder PEATE, we have already seen a large reduction in cost to build and deploy the science data system portion of the GDS.

V. Future Directions

Our future work involves building additional capabilities using the PCS software including complex data provenance and search capabilities for existing and future missions.

VI. References

[1] C. Mattmann et al. A Software Architecture-based Framework for Highly Distributed and Data-Intensive Scientific Applications. Proc. *ICSE*, 2006.