The Goddard Earth Sciences Data and Information Services Center (GES DISC) at NASA has over the years developed and honed several reusable architectural components for supporting large-scale data centers with a larger customer base. These include a processing system (S4PM) and an archive system (S4PA) based upon a web and an online data visualization and analysis system (Giovanni). Some of these subsystems (S4PM and S4PA) have also been reused by other data centers for operations

S4PM

The Simple, Scalable, Script-based, Science Processor for Measurements (S4PM) is a processing system conceived and developed in 2001 to support the processing of data from MODIS in the GES DISC. By running the proven S4PM core, S4PM developers quickly developed a capable processing system that was optimized to support MODIS algorithms and meet the reprocessing schedule. In 2002, S4PM was enhanced to process data from the Atmospheric Infrared Sounder (AIRS). The influx of new, instrument-unique requirements forced an S4PM refactoring towards a more modular architecture, particularly of the interfaces where algorithms are plugged in. A subsequent requirement to support multiple platforms (Sun, Solaris) provided the impetus to enhance portability. And the need for S4PM to be installed, deployed, and configured quickly by multiple users resulted in better packaging and more automated configuration. This led to public release of S4PM under the NASA Open Source Agreement in 2004. S4PM was then reused by Langley Research Center to process data from CALIPSO, MISR, and FLASHFire, and by EROS Data Center to process ASTER data.

S4PA

The Simple, Scalable, Script-based, Science Product Archive (S4PA) was designed to handle the migration of legacy data from robotic tape archives to disk. Beginning in 2006, S4PA was tapped to replace the massive EOSDIS Core System (ECS) archive data with a hard deadline at the end of 2007. The schedule pressure to support multiple disk-based instances of S4PA to accommodate the vast data volume (~400 TB), each with its own integration engineer, drove the development of more sophisticated packaging and easier, more automated installation mechanisms.

Giovanni

Giovanni® is an online data visualization and analysis system that started out in 2002 as a Web-based CGI application. It provided an interactive session via HTTP through which users could select data parameters, regions of interest and time ranges and generate a variety of plots and images. Data supporting Giovanni were on the machine hosting it and could be downloaded by the user via FTP. As Giovanni grew in popularity, science users provided feedback, which drove enhancements which, in turn, expanded Giovanni’s popularity. Scientists seeking collaborations and interoperability with their systems followed. In 2005, Giovanni was then refactored from a CGI-based application to a services-oriented architecture supported by a synchronous workflow management system. Data no longer had to be on the host machine and could access remote data using an expanding list of protocols as well as multiple output protocols and data formats.

Our experience has been that development and utilization of robust, interoperable, and reusable software systems can actually flourish in environments defined by heterogeneous commodity hardware systems, the emphasis on value-added customer service, and the continual goal for achieving higher cost efficiencies. The repeated internal reuse that is fostered by such an environment encourages and even forces changes to the software that make it more reusable and adaptable. Allowing and even encouraging such selective pressures to software development has been a key factor in the success of S4P and S4PA, which are now available to the open source community under the NASA Open Source Agreement.