ORIGAMI AND GIPS: RUNNING A HYPERSPECTRAL SOUNDER PROCESSING SYSTEM ON A LIGHTWEIGHT ON-DEMAND DISTRIBUTED COMPUTING FRAMEWORK

Maciej Smuga-Otto, Raymond Garcia, Graeme Martin, Bruce Flynn, Robert Knuteson
Space Science and Engineering Center University of Wisconsin - Madison

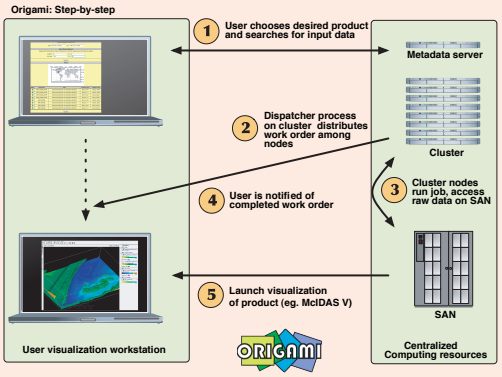
The scientific workflow need we are addressing involves the retrieval, processing and visualization of large volumes of data, often on the order of Terabytes, associated with hyperspectral sounder-imager instruments.

Science workstations, compute clusters and storage area networks are the hardware components available to us. Likewise, certain commodity software packages for visualization, queueing, data storage management and grid resource management are readily available. While there are also off the shelf science workflow management and scientific programming packages out there, these are often proprietary and impose serious architectural constraints on the rest of the system.

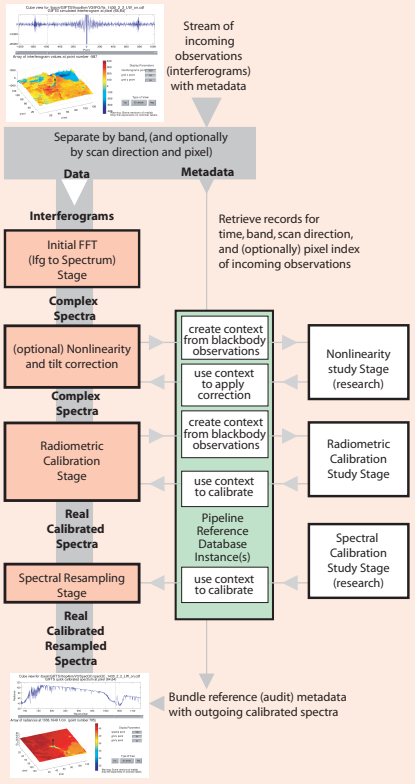
Processing raw data from the Geosynchronous Imaging Fourier Transform Spectrometer (GIFTS) instrument was the principal motivation for this work. Initially the GIFTS Information Processing System (GIPS) was meant to cover various aspects of distributed computing as well as hosting the GIFTS algorithms. Eventually, it was down-scoped to handle provisioning of the algorithms with appropriate data and metadata. This form of GIPS is suitable for running on a single compute node, and the need arose to find a distributed computing solution to match.

Origami is the outcome of a parallel effort in distributed computing at the SSEC - it is an experiment in rapid prototyping of distributed science workflows using lightweight web components. As such, it can be extended to serve as a glue technology between GIPS running on [illegible] of a true distributed computing system.
To make GIPS compatible with an external distribution and workflow manager such as Origami, we found that it was beneficial to convert all algorithm interfaces to a common XML format. This XML interface file would be converted to an API-level interface for compiling into GIPS, and would also be used by the Origami engine to perform workflow and scheduling decisions.

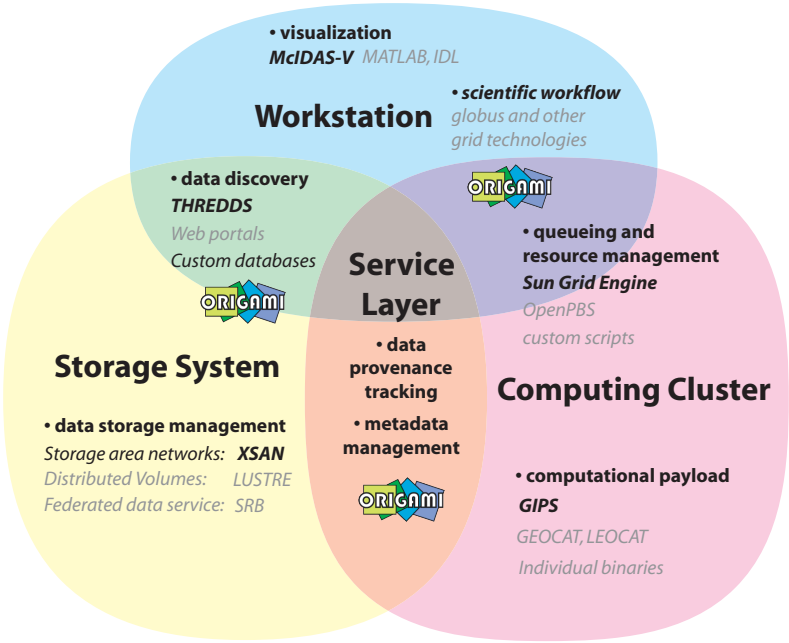
This XML interface style has already spread to GEOCAT and LEOCAT, two other satellite data processing frameworks being developed at the SSEC. These frameworks utilize a fundamentally similar XML interface format as well as the suite of tools for managing it.



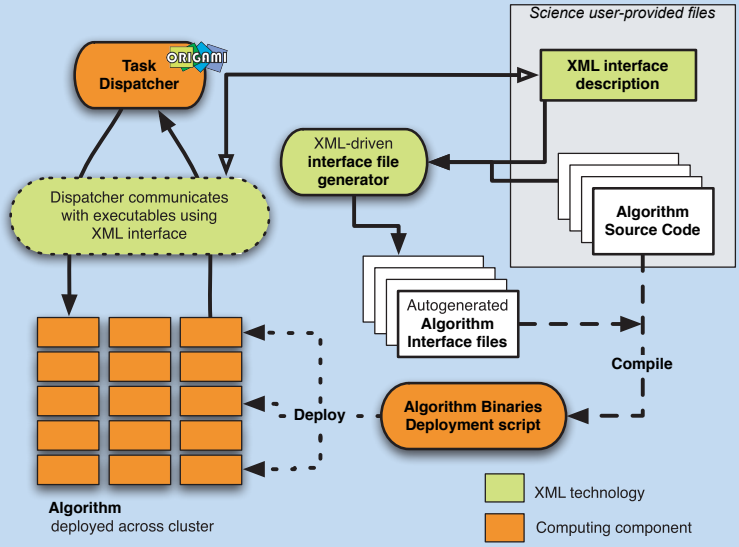
Origami by-the-numbers: Workflow for processing and visualizing large volumes of data



At the heart of GIPS is an algorithm pipeline that produces calibrated radiances from raw imaging interferometer data and is a candidate heavy-weight ground processing task



Component level architecture of the distributed processing system contributes to individual buy vs. build decisions.



Integration of Origami and GIPS drove the creation of an XML interface for algorithms hosted by the system