Rolling Out GENESIS/SciFlo in the ESIP Federation's *Earth Information Exchange*

Thomas Yunck, Brian Wilson, Eric Fetzer, Amy Braverman, Annmarie Eldering
Michael Garay, Gerald Manipon, Elaine Dobinson and Benyang Tang

Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109

Abstract. GENESIS/SciFlo is a Grid-based automated workflow execution system for conducting atmospheric and climate studies that integrate data sets from EOS and other instruments. Many components of SciFlo are in experimental operation and several scientific studies making use of these tools are underway. These include validation studies of AIRS, MODIS, and GPS-based atmospheric temperature and moisture data, atmospheric aerosol studies with MISR data, and cloud model validation with multiple datasets. The first public version of SciFlo will be rolled out in late 2006 as part of the ESIP Federation's new portal, the Earth Information Exchange (EIE). The EIE will be a "community" within the government-wide Geospatial One-Stop (GOS). SciFlo will enable the chaining of data services from multiple providers into an automated end-to-end investigation. Services will include data discovery, subsetting, co-registration, data mining, and analysis. At present its functions are restricted to the developmental scenarios of the GENESIS project.

INTRODUCTION

Harnessing modern Grid computing and data system technologies, GENESIS (the General Earth Science Investigation Suite) is developing an interactive tool to enable multi-sensor measurement-based science and applications to be conducted with ease on large datasets from Aura, Aqua, Terra, and GPS sensors. This year an early prototype of GENESIS will be deployed within the Goddard, Langley, and JPL DAACs, and be accessible for demonstration purposes through the Foundation for Earth Science's new Earth data portal, the Earth Information Exchange (EIE). The EIE will be housed within the government-wide Geospatial One-Stop (GOS) portal, operated by the USGS [1]. GENESIS is designed to serve for public education and discovery and for leading-edge atmospheric and climate research. Functions to be introduced in the initial roll-out on the EIE include:

- Integrated space/time query and URL access services for AIRS [2], MODIS [3], MISR [4], and GPS [5]
- On-demand multivariate statistics distilled from AIRS and MISR datasets
- Specific climate model and data comparisons
- On-demand match-up, comparison, and visualization of atmospheric temperature and moisture data from AIRS, MODIS, and GPS
- Service chaining using data from WMS/WCS servers and possibly GRASS operators

When fully deployed in 2008, GENESIS will:

- Facilitate seamless interoperability and data flow over broadly distributed systems
- Use smart web services for efficient data access, processing, and exchange
- Incorporate advanced data subsetting and a variety of built-in analysis tools
- Streamline and automate the flow of information among science and application elements
- Accelerate the use and validation of EOS and other Earth science data for science and applications

GENESIS will offer a palette of data selections along with web services for subsetting, transforming, co-registering, mining, and analyzing the data. Both internal and external registries will track available datasets and services enabling the system to locate and engage them. GENESIS is built on the SciFlo workflow engine [6,7] and will exploit the Earth Observing System Clearing House (ECHO) [8] and the Global Change Master Directory (GCMD) [9] for data discovery and resource brokering, and OpenDAP [10] for data subsetting. Developed under NASA's REASoN program, GENESIS, is a prototype now being used in science investigations on atmospheric data from three EOS sensors (AIRS, MODIS, and MISR) and from GPS occultation limb sounders.

GENESIS will provide a measurement-based environment enabling Earth scientists and environmental practitioners to access diverse datasets and employ them in a variety of scientific investigations or environmental analyses. Studies that today may entail months or years of effort will be compressed into weeks or even days, and may be iterated dozens of times in the time now required to complete a single study.

This expansion of productivity will enable new approaches in Earth system science and applications. Rather than having to focus on specialized tool-building for one-off analyses with restricted measurement sets, analysts will be able to fuse data at a touch from many instruments and execute expansive ensemble studies providing deeper insight into the behavior of the Earth system. By accelerating the analysis process many-fold, GENESIS will advance the vision of Earth System Science benefiting society—the foundation idea behind NASA's Earth Observing System.

The GENESIS team consists equally of IT specialists and Earth scientists. The scientists help specify, design, and in

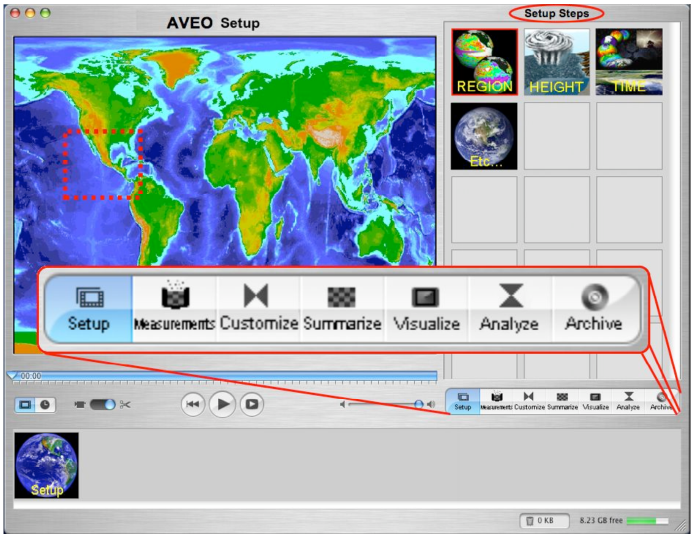

Fig. 1. Conceptual layout for the GENESIS visual workspace. Icons (right panels) are used to represent datasets and operators and to construct analysis workflows in stages that include problem definition (Setup), measurement selection, customization, visualization, analysis, and archive.

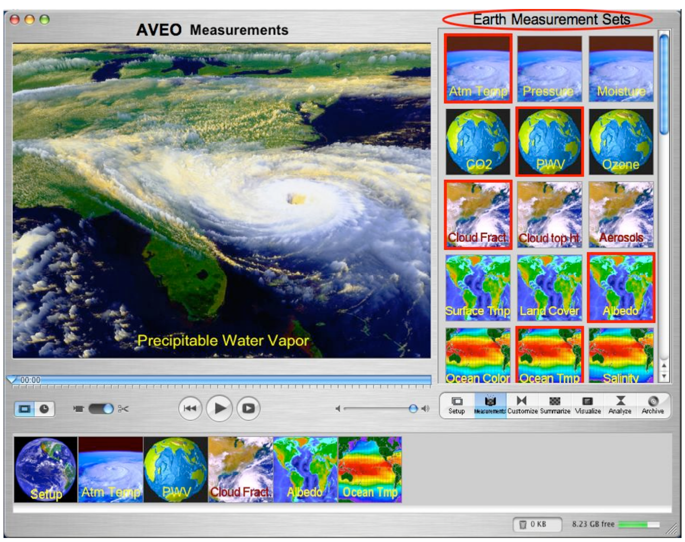

Fig. 2. After the Setup is completed, the analyst will select data products from the palette in the right-hand panel. GENESIS will offer data through Earth Information Exchange from the Federation archives and other sources.

some cases implement the palette of GENESIS functions and, as they are built, exercise them in their ongoing research, providing feedback to the developers. Current investigations with GENESIS include comparison and validation of atmospheric temperature and moisture data from AIRS, MISR, MODIS, and GPS occultation sensors, validation of cloud climate models, aerosol studies with MISR data, and study of moisture transport between the troposphere and stratosphere.

GENESIS is designed to further NASA's theme, "from missions to measurements"—focusing on the physical quantities that must be measured to address defined science or application objectives, rather than on particular notions of missions or instruments to be flown. Related objectives are to:

- Enable greater interoperability and data flow between distributed sources, processing assets, and distribution portals
- Use intelligent web-enabled services for handling routine data access, processing or exchange tasks
- Incorporate data reduction tools into delivery systems allowing scientists to obtain only the data required

GENESIS WALKTHROUGH

Here we walk through an example of the GENESIS concept of operation as depicted in the hypothetical screen shots in Figs. 1-3. [This section lays out the "high concept"—the grand vision. For the prototype we will implement only those functions needed to execute specific science use cases. The delivered open-source system will be owned, in effect, by the community, so that a growing user base may elaborate it at will for additional datasets and operations.]

Suppose Alice wants to perform a complex study involving five different data types—atmospheric temperature, precipitable water, cloud fraction, albedo, and sea surface temperature—within a specified region. The measurements are stored in distant archives in mutually incompatible formats. Alice has no in-depth knowledge of the formats, data centers, or other particulars concerning acquisition and fusion of the data, only what measurement types to use. To begin, she opens her browser to GENESIS within the GOS/EIE portal and activates the Setup screen (Fig. 1). The enlarged inset lays out the steps she can choose from in defining her investigation. (Disclosure: The layout for these conceptual panels is inspired by the *iMovie* application on the Macintosh.)

From the initial setup menus GENESIS allows Alice to define the nature and basic parameters of her investigation, offering point-and-click icons, graphical region specification, and text entry for more precise definition. When the setup is complete, a Setup icon appears in the visual program along the bottom. Alice can modify the setup by clicking open the icon at any time to return to the Setup menus.

Alice then moves to the Measurements screen to select from the vast ESIP Federation canon [11] and other sources (Fig. 2). These will be stored at many locations, including possibly at sites of users who wish to share with peers. Each measurement selected appears as an icon in the visual program, which can be clicked open to specify particulars.

Next Alice calls up a series of screens offering operators to subset, transform, and fuse the measurements, perform statistical "summarization" and other data mining tasks, create visualizations, and conduct a range of statistical analyses (Fig. 3). GENESIS will also let her create custom operators in the form of executables generated from virtually any coding language and publish them as web services. Thus any code she may have developed in the past can be infused into GENESIS and, if desired, exposed for use by others.

As Alice proceeds, GENESIS extends its visual program representing each step in the investigation (bottom panel). These can be clicked open for editing or debug execution. When the workflow is ready, Alice hits the Execute button

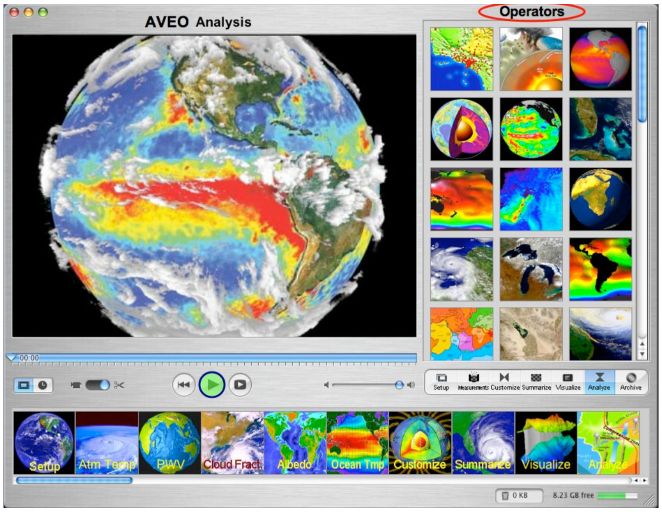

Fig. 3. The analyst then selects other modes to customize, mine, visualize, and analyze the data. Each mode will offer a menu of operators and services for these purposes. Many will offer multiple options.

and SciFlo carries out the investigation from end to end, creating an efficient execution plan, engaging the required resources, managing execution, and reporting results.

A complex analysis might be set up and executed within days or hours and rerun with modification within minutes. The published workflow could then become a general tool for environmental research. The expert knowledge needed to fuse and manipulate the data will be built in; all such steps will be documented at every stage.

Engaging the Global Grid

To execute a complex study quickly, Alice may have to engage resources on an extended computing Grid. While GENESIS in its full generality will be realized only as a true national computing Grid emerges, that day is not far off. The TeraGrid [12], National Lambda Rail [13], Earth System Grid [14], and like networks are quickly expanding, and broadband fibers now span the nation. Key data centers for GENESIS at Scripps, Langley, Goddard, and JPL are connected by 10 Gig Ethernet links. GENESIS will exploit such links where they exist (and the Internet where they don't), and provide middleware at several levels to forge these elements into a practical scientific research and applications tool.

Fig. 4 illustrates what will occur when Alice hits the Execute button. (For our demonstration system we will use a very limited Grid consisting of our portal testbed and well-tried links to specific archives.) Residing within the GOS/EIE portal, GENESIS will parse the workflow and convert it to an XML [15] description document. SciFlo will then interpret this document and consult resource registries (e.g., GCMD or ECHO) within the GENESIS domain. SciFlo will devise an execution plan based on current resource loadings and, employing SOAP (Simple Object Access Protocol) [16] web services along with the facilities of Version 4 (the web services version) of the Globus toolkit [17], negotiate participation by the resources, engage them, parcel out the tasks in a parallel scheme, schedule execution, monitor progress, restart subflows when required, and return the results.

The technical tools needed to engage and manipulate the Earth science data from five NASA sensor types exist in part already within GENESIS, sufficient for a limited GUI overlay. Many more such tools will be built during the remaining two years of this development.

Realizing GENESIS Objectives

Returning to our three principal objectives, we can now say more specifically how GENESIS will address them.

1) *Enable greater interoperability and data flow between distributed data sources.* SciFlo will bring together the specialized services of GENESIS, ECHO, and OpenDAP to discover, access, subset, retrieve, customize, fuse, visualize, and analyze diversely formatted, incompatible products from NASA instruments and automate their execution—chaining and managing the flow of data and services among sources and assets—on a national computing Grid. GENESIS will do this for data from AIRS, MODIS, MISR, and GPS, residing in four dispersed data centers (the GSFC, LaRC, and JPL DAACs, and the JPL GPS archive [18-21]), while adding a host of new science operators. It will wrap all of these tools into a simple, intuitive, point-and-click environment where they can be applied to create complex science workflows.

2) *Use intelligent web services for handling routine data access, processing or exchange tasks.* The SciFlo server is an expert web services engine that knows the whereabouts of all assets within the GENESIS Grid, can parse the investigation

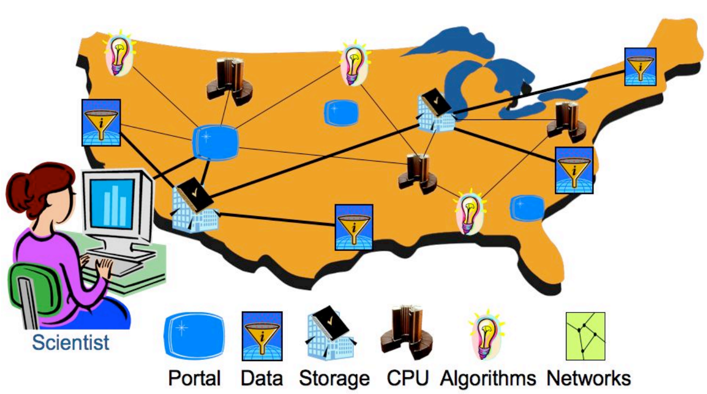

Fig. 4. When Alice executes a GENESIS visual workflow it is first translated by the local SciFlo client into an XML document and then sent to the SciFlo server for interpretation by the SciFlo execution engine. The execution engine consults a registry or registries of the computing resources within the GENESIS domain. Registries to be used by SciFlo include NASA's ECHO and GCMD, as well as an internal registry maintained by SciFlo. The execution engine devises an efficient parallel workflow execution plan, negotiates with and engages the necessary data and computing resources, initiates and monitors execution, and notifies Alice upon completion.

workflow, assess the network and resource loadings, devise an efficient parallel execution plan, negotiate its execution with the different assets, and properly manage the execution as it proceeds. It will incorporate "semantic web" features via special Earth science "ontologies" devised by the NASA Earth Science Data System Working Groups for enhanced understanding of queries [22,23], and will carry out all of its functions through SOAP web services—that is, by exchanging XML messages among nodes via SOAP.

3) *Incorporate data reduction tools into delivery systems that allow scientists to obtain only the data required.* Central to GENESIS and OpenDAP are specialized subsetting operators tailored to the particular focus archives and data formats. At present the GENESIS subsetting operators function only within the GENESIS testbed at JPL, where extensive samples of the instrument data have been installed; that is, they have not yet been infused into the DAACs, where they will ultimately be most effective. However, three DAACs (JPL, GSFC, and LaRC) are participants in GENESIS and will infuse the operators when they are released. The first prototype SciFlo server will be installed on the DAACs in the summer of 2006. These installations will be employed in the initial GENESIS roll-out on GOS/EIE later in 2006.

Enabling Technologies

GENESIS will provide a graphical interface and new data operators to weave several key Grid technologies into a seamless Grid service for Earth science and applications. To achieve this, GENESIS and the core tools it brings together tap many elements of the Grid revolution, including:

- Nationally distributed parallel execution
- Automated machine-machine transactions
- Simple XML messaging services with SOAP
- Intelligent network load balancing at run time
- Multi-scale resource integration, from desktop to national supercomputer centers
- Peer-to-peer resource sharing: Users can post resources (operators, workflows, data products and services, even cpu cycles) to the GENESIS registry for access by all users.

Much of its technical substance lies in the proven elements that GENESIS brings together. To illustrate how they will operate, we describe several of them here in greater detail.

A. The SOAP/XML Substrate

The Extensible Markup Language, XML, has now become a universal standard, the lingua franca of commerce and communication on the web. SOAP web services transacted via XML messaging have both enabled and transformed e-business and are rapidly invading the science arena. The alliance of XML metadata, messaging, and service execution with advanced tools for semantic understanding (ontologies) offers vastly simplified means of chaining services on the web, and thus of building elaborate, versatile, yet easy-to-use service systems. The universality of these systems has come about through the wide acceptance of basic standards and proliferating software tools for applying them.

B. The Earth Information Exchange

GENESIS will build upon a sound in-place technology base. The cornerstone is the Federation's new environmental data marketplace, the **Earth Information Exchange** (EIE) [24], soon to be installed within the multi-agency Geospatial One-Stop (GOS) portal. The EIE is being built in partnership with GOS (managed by USGS), ESRI, and NOAA's National Geophysical Data Center (NGDC) [25]. It receives Financial support from the Foundation for Earth Science and contributed effort from the partners of the ESIP Federation.

The EIE will appear as a "Community" within GOS and offer access to the full suite of GOS products, which focus on 2D mapping and imaging data and GIS functions. When the EIE debuts in late 2006 it will expand GOS to a wealth of new science products from NOAA, NASA, EPA, and other providers. These include all NOAA and NASA space-based data, as well as the multi-sponsor GPS occultation products.

The EIE will provide automated search and discovery across its products and services through the Global Change Master Directory and EOS Clearing House. The GENESIS team will register with GCMD and ECHO all new science products generated by its science use cases.

C. Web Services

A web service is (for our purposes) any function performed at a web-accessible site upon request by a remote user. Of interest for GENESIS are such data services as Discover, Subset, Retrieve, Fuse, Mine, and Analyze. Web services are invoked by a user application, often a browser, by sending an XML message to the host site asking that the service be executed. This process is invisible to the user who typically clicks a browser button or fills in a form. The XML message is, in effect, a "loosely-coupled" remote subroutine call.

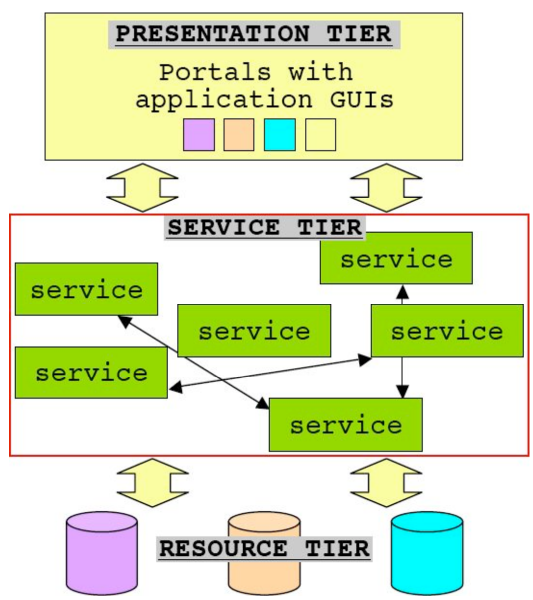

Fig. 5. Standard 3-tier structure for a Grid application. The AVEO prototype will include a new top tier – GUI and portals – and a number of critical services on distributed computing resources.

Powerful new applications, such as configurable virtual observatories, can now be created simply from the scripting needed to choreograph available services, to chain them into a productive sequence. Domain experts now provide essential functionality in the form of web services. The XML control code can be generated automatically with the aid of a service registry maintained at a portal, such as GOS/EIE. The tandem of user interface and portal forms the top layer of the standard three-tier "Grid application" (Fig. 5).

This "service-oriented architecture" (SOA) is the enabler and foundation of modern e-commerce. In science it offers a means to propagate individual advances—to broadcast them instantly to the world—and foster opportunistic remote collaborations. Many scientists develop unique analysis tools for their work but have little means for offering them widely. Exposing them as web services, a seamless process within GENESIS, will make them available to a world of prospective users and provide the essential components for fashioning virtual observatories on demand.

The web revolution has spawned a vibrant economy in web services: competing service providers, discovery services and registries, value-added consolidators, reviews and ratings of providers, discussion forums, etc., many of which spring up spontaneously within participating communities. This market model is a potent engine that can multiply scientific productivity by fostering open commerce in scientific web services.

D. *Creating Web Services with GENESIS*

Any remote executable or script that employs a standard command line (e.g., switches and positional arguments) and reads/writes one or more files can be exposed as a web service and become a clickable GENESIS operator. To invoke the operator one simply selects it in the GUI and specifies the calling parameters by, for example, filling in a pop-up form. The resulting workflow document will contain a call to the service, with the calling parameters described in XML. In executing the workflow, SciFlo will send the XML call to the remote host, which will convert it back to its original command-line form and run the executable.

Publishing a new web service based on a functioning SciFlo document is as simple as asking SciFlo to do it. One will be able to create new SOAP services simply by building a workflow from within GENESIS and declaring it. To expose finer-grained functionality in a complex executable—e.g., different run modes or individual sub-functions—one must first create separate configurations that perform those functions and expose each one individually.

In addition, one can expose Java methods (rather than executables) as services through the Apache Axis framework [26]. For IDL and MATLAB codes, one can automatically wrap a routine that reads and writes supported formats. A few lines of IDL or MATLAB must be added to pick up the inputs and call the function. The biggest chore in this process may be cleaning up the original code. Depending on the case, that could take from a few minutes to a few weeks.

E. *SciFlo Architecture*

SciFlo's architecture combines four core ideas: loosely-coupled distributed computing using SOAP; exposing scientific operators as SOAP web services; specifying a data processing stream as an XML document; and a dataflow execution engine for parallel execution and load balancing.

Loosely-coupled distributed computing. SOAP remote procedures are invoked through XML messages without regard to the implementation details of the procedures or services being engaged. SOAP is lightweight, language-independent, and ideal for loosely-coupled distributed computing.

Analysis operators as SOAP web services. Scientific operators are often highly complex, with many configuration parameters, and may involve vast and diverse inputs and outputs. SciFlo couples diverse operators and other components that can be either local or remote in a simple, declarative manner. A SciFlo operator accepts one or more complex inputs, performs tasks tailored through configuration parameters, and yields one or more complex outputs. Inputs and outputs can be local files and metadata, pointers to remote files or array slices, or the actual numbers.

Declarative programming. SciFlo provides XML standards that describe analysis operators (verbs) and groups of operations (processing flows) to serve as building blocks in an arbitrary processing flow. By encapsulating code within well-defined operator interfaces described in XML, new programs can be assembled with a simple XML document. The user "declares" the new processing flow and the execution engine does the rest. SciFlo will provide the "glue" allowing any user to expose analysis operators as SOAP-callable web services and to invoke remote operations as part of a data processing flow without writing new code. SciFlo adopts a standard web description language and catalog services, but specialized for science—i.e., compatible with science standards (e.g., WCS, DODS) employing simple XML flows, and offering hooks for semantics from the science realm.

Dataflow Execution Engine. SciFlo extends the idea of an execution engine by engaging open, interoperable web services with on-demand integration of operators. It contains a dataflow engine that parses the XML description of the process flow, creates a parallel execution plan, distributes and balances the computing load, and coordinates execution. Operators can be local or remote; if the latter, SciFlo either invokes a SOAP service, makes a WMS/WCS call, submits an http POST request, or submits a one-line http GET request. The engine creates the code to move data inputs and outputs between nodes, sequence operations, execute operators in parallel, update a status log, and deliver results.

SciFlo can also exploit elements of Globus toolkit Version 4 (via SOAP) to allocate resources and submit and monitor remote jobs. However, SciFlo has its own capabilities in these areas and complies with the WS-Resource Framework (WSRF) standard. Thus, distributed dataflows can utilize a network of lightweight SciFlo nodes with only selected nodes having to install the more cumbersome Globus toolkit.

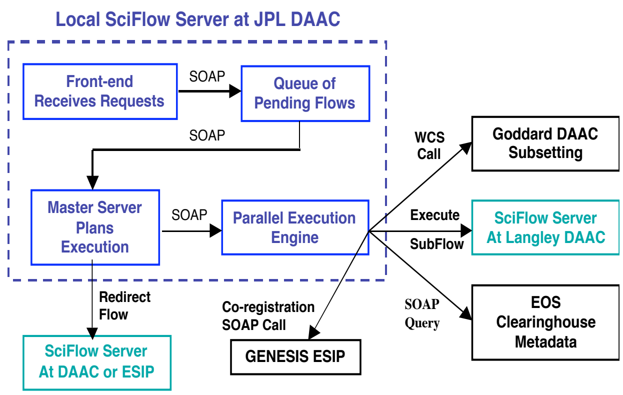

Fig. 6. Block diagram of SciFlo operation.

Server Operation. The execution engine comprises separate processes, executing on one node or many, that talk via SOAP; an XML file specifies the network to employ. The SciFlo server is itself a SOAP service and can register with the UDDI [27] discovery catalog; one can then use UDDI to discover SciFlo nodes to combine.

Parallel Computing. SciFlo will apply parallelism at several levels, including:

- With multiple processes on a single node and parallel subflows on multiple nodes;
- By invoking a remote operator via a SOAP or WCS73 call and waiting for the results;
- By directing flows to servers that can access data locally or can provide computing power;
- By invoking a flow multiple times if a source operator yields multiple files to be processed.

Fig. 6 depicts the SciFlo operational scheme. Front-end nodes field requests and put them in a queue. A master node retrieves the flow and executes it through the planning phase. If flow redirection is needed, the annotated flow document will be sent to the proper server.

SciFlo thus marries multiple tools and servers in one system: a versatile web server supporting data portals, distributed collaboration, SOAP data query and access, metadata query (with the relational database mysql), retrieval of files from archives or cache, subsetting of local input data (with OpenDAP), exposing of output products across the Grid, and parallel flow execution.

F. Semantic Extension

Every physical quantity in SciFlo has a type, unit, default value, and optional semantic kind. SciFlo uses XML schemas to describe object types. The Earth Science Markup Language (ESML) [28] is an XML specification for syntactic metadata, indicating what each variable represents, and for "content" metadata for discovery. SciFlo adopts existing ESML descriptions and readily creates new ones. Units of measure are drawn from both GML and the SWEET ontology. SciFlo also provides a placeholder for "kind" data to allow for later semantic web advances.

Type and kind elements can be formally represented in a semantic ontology using the Resource Description Framework (RDF) [29] or the Ontology Web Language (OWL) [30]. Each input, output, or process tag in a SciFlo document can reside in a domain namespace and be a technical term that points into one or more ontologies. Thus, every SciFlo document contains a mixture of XML namespaces and is semantically extensible simply by changing the name of a tag or its kind attribute. The *signature* of an operator or service consists of the types and kinds of its inputs and outputs and a "semantic kind" for its transformation. Signatures allow tools and services to be found by ontology-enhanced search (synonyms, term narrowing/broadening, etc.). Ontologies for geo-transformations and service interfaces are being created with OWL-S [31].

G. Discovery and Brokering with ECHO

GENESIS will build upon the ECHO system for general purpose metadata discovery and access and will employ ECHO as a central resource for automated cross-provider "service brokering." ECHO offers flexible spatial, temporal, and keyword searching together with interoperable registries allowing diverse datasets and third party services to interact seamlessly. Based on provider specification, any registered service may be applied upon request to any registered dataset; the details of interfacing and execution are handled automatically. This service brokering will be operational in ECHO in late 2006. ECHO is thus tailor-made to expose, discover, and employ web-based Earth science resources.

Conclusion

By uniting a simple graphical interface for building science workflows with powerful data manipulation and Grid computing tools, GENESIS will enable the easy selection, fusion, compression, and analysis of multi-instrument datasets and reduce the time needed for complex Earth system studies by orders of magnitude.

By relieving scientists of the burdens of developing custom code for each step of an investigation, and by removing barriers to access and use of complex EOS datasets, GENESIS will allow them to bring many additional relevant measurements into their studies. And by engaging a national computing Grid to execute those workflows, GENESIS will allow users to perform ensemble studies comprising dozens or even hundreds of variant runs in the time previously required for just one. GENESIS will further allow analysts to preserve and publish their workflows as operators for later use, in follow-on studies or as decision support tools. It will offer scientists and decision-makers a direct entry point to the national computing Grid, enabling more comprehensive and faithful system-level studies of the Earth's evolving condition and its response to forcings both human and natural. With widespread use, GENESIS can accelerate the pace of progress in understanding Earth system behavior many-fold.

ACKNOWLEDGMENT

The work described in this paper was performed by the Jet Propulsion Laboratory under contract with the National Aeronautics and Space Administration. We would like to thank Dominic Mazzoni, formerly of JPL and now at Google, and David Pierce of the University of California, San Diego, for their contributions to GENESIS/SciFlo development.

REFERENCES

[1] GOS homepage: http://gos2.geodata.gov/wps/portal/gos/

[2] AIRS homepage: http://www-airs.jpl.nasa.gov/

[3] MODIS homepage: http://modis.gsfc.nasa.gov/

[4] MISR homepage: http://www-misr.jpl.nasa.gov/

[5] The ESIP Federation's Global Environmental and Earth Science Information System, homepage: http://genesis.jpl.nasa.gov/zope/GENESIS.

[6] Yunck, T., B. Wilson, A. Braverman, E. Dobinson, and E. Fetzer, GENESIS: The General Earth Science Investigation Suite, Proc. NASA Earth Science Technology Conference – ESTC-04, Palo Alto, June 2004. (Available at http://sciflo.jpl.nasa.gov/. This site also contains examples of SciFlo dataflow documents and results for a large-scale Genesis-II analysis comparing temperature and water vapor profiles from AIRS Level 2 retrievals and GPS limb scans.)

[7] Wilson, B., B. Tang, G. Manipon, D. Mazzoni, E. Fetzer, A. Eldering, A. Braverman, E. Dobinson, and T. Yunck, GENESIS SciFlo: Scientific Knowledge Creation on the Grid Using a Semantically-Enabled Dataflow Execution Environment, 17^{th} Scientific and Statistical Data Base Mgt. Conf., Santa Barbara, CA, June 2005 (Available at http://sciflo.jpl.nasa.gov/, see above).

[8] ECHO homepage: http://www.echo.eos.nasa.gov/

[9] Global Change Master Directory homepage: http://gcmd.gsfc.nasa.gov/

[10] OpenDAP homepage: http://www.opendap.org/

[11] ESIP Federation homepage: http://www.esipfed.org/

[12] TeraGrid homepage: http://www.teragrid.org/

[13] National Lambda Rail homepage: http://www.nlr.net/

[14] Earth System Grid, homepage: https://www.earthsystemgrid.org/

[15] The eXtensible Markup Language. See http://www.xml.com/ and http://xml.silmaril.ie/

[16] See, for example: Egan, S., Simple Object Access Protocol: A Step-By-Step Approach. Available at: http://www.vbip.com/xml/soap_syd.asp

[17] Globus toolkit homepage: http://www.globus.org/toolkit/

[18] The Goddard Space Flight Center's Earth Sciences Data and Information Center, homepage: http://daac.gsfc.nasa.gov/.

[19] The NASA Langley Research Center's Atmospheric Science Data Center, homepage: http://eosweb.larc.nasa.gov/.

[20] JPL PO.DAAC homepage: http://podaac.jpl.nasa.gov/

[21] The ESIP Federation's Global Environmental and Earth Science Information System, homepage: http://genesis.jpl.nasa.gov/zope/GENESIS.

[22] Raskin, R. G. and M. J. Pan, Knowledge representation in the semantic web for Earth and environmental terminology, *Computers and Geosciences*, 31, 1119-1125, 2005.

[23] Earth Science Data System Technology Infusion Working Group, "IT Capability Vision," working draft, May 2005, Available at: http://www.sciencedatasystems.org/seeds/wg/infusion/Working%20Documents/Capability%20Vision/capability%20vision%20charts.pdf/.

[24] Yunck, T. P. and R. Wertz, Earth Science and the Global Grid: The ESIP Federation Today, presented at the ESIP Federation Summer Meeting, Washington, DC, June 2005. Available at: http://genesis2.jpl.nasa.gov/archive/200601004/.

[25] NOAA NGDC homepage: http://www.ngdc.noaa.gov/.

[26] See Apache Axis homepage: http://ws.apache.org/axis/.

[27] Universal Description, Discovery and Integration, see http://www.uddi.org/.

[28] See ESML homepage: http://esml.itsc.uah.edu/index.jsp.

[29] See RDF homepage: http://www.w3.org/RDF/.

[30] OWL documentation at http://www.w3.org/TR/owl-features.

[31] OWL-S documentation at http://www.daml.org/services/owl-s.