

INTEGRATING SEISMIC EVENTS FOCAL MECHANISMS WITH IMAGE SERVICES IN KEPLER

**Efrat Jaeger-Frank¹, Ashraf Memon¹, Ilkay Altintas¹, Ghulam Memon¹, Dogan Seber¹,
Bertram Ludaescher^{1,2}, Chaitan Baru¹**

¹San Diego Supercomputer Center, University of California, San Diego

²Department of Computer Science, University of California, Davis
{efrat, amemon, altintas, ludaesch, gmemon, seber, baru}@sdsc.edu

Geosciences Network (GEON) is a grand-challenge project to prototype cyberinfrastructure for the Earth Sciences. GEON is building a mechanism for integration and visualization of seismic events and their related fault orientations--referred by seismologists as focal mechanisms (or "beach balls") with other image (map) services using a service oriented architecture (SOA). The Kepler scientific workflow system provides capabilities to harvest and query datasets and services and process data using local and remote services. Using a workflow, we can combine the harvested datasets (such as seismic event locations) or map services of various formats with matching beach ball symbols, retrieved from a beach ball images database, for seismic events. The result is processed into a new dataset, using an ASCII to shapefile service, and integrated and visualized with other image services using ArcIMS. The new dataset can also be registered and searched within the GEONsearch registry.

1. Introduction

The Geosciences Network (GEON) [1] is an NSF-funded large Information Technology Research (ITR) project to facilitate collaborative, inter-disciplinary science efforts. GEON is developing an infrastructure that supports advanced semantic-based discovery and integration of data and tools via *portals* (the GEON portal [2]), to provide a unified authenticated access to a wide range of resources to conduct comprehensive analysis using emerging web and grid-based technologies to enable the next generation of science and education.

One of the many challenging problems GEON is focusing on is the integration of multiple datasets to understand the lithospheric structure of North America. One critical dataset that provides valuable information about the nature of the lithosphere is seismic data representing locations of earthquakes as well as earthquakes' source parameters or focal mechanisms. This paper addresses development of a workflow-based map generation tool for earthquake locations and their focal mechanisms to be used in integrative research efforts. Plotting focal mechanism on the fly in a GIS environment is a cumbersome task. A large number of earth scientists use GIS system in their research. Although currently available GIS tools are quite sophisticated and allow representation of so many different types of datasets, they fail to provide tools for easy representation of earthquake focal mechanisms (sometimes referred as "beach balls").

Solving this problem requires adaptable interfaces and tools for accessing heterogeneous, distributed, scientific data and computational analyses, e.g., access to and querying spatial resources, on the fly extraction of datasets and interactive viewers to provide a neat visualization of the results. ArcIMS services are used to achieve these functionalities. These services,

wrapped as web services, can be composed in a meaningful order, resulting in a scientific workflow in which the flow of data from one analytical step to another and the overall execution semantics is defined by a formal workflow language. While traditional business workflows are oriented towards document processing, task management, and control-flow, scientific workflows typically are data-intensive and/or computationally expensive, dataflow-oriented, and often involve data transformations, analysis, and simulations.

In this paper we present a scientific workflow approach to model the problem using the Kepler [3,4] scientific workflow system. Kepler is a multi-disciplinary scientific workflow system that provides domain scientists with an easy-to-use yet powerful system for capturing and automating the ad-hoc processes that a scientist may go through to get from raw data to publishable results. Kepler along with the service oriented architecture allows coordination of distributed data and process components and supports reusability and interoperability of components on the Web.

The rest of this paper is organized as follows. Section 2 provides a brief overview of the Kepler workflow system and describes the feasibility of integrating ArcIMS services through web services. Section 3 depicts the ArcIMS services used to address the problem. A coordination of these services to a scientific workflow solution is presented in Section 4. Finally, Section 5 provides a summary and a discussion of additional support for ArcIMS services utilized in Kepler.

2. Kepler and the Service Oriented Architecture

Kepler [3] is a system for the design and execution of scientific workflows. It is built on top of the PtolemyII system, a modeling and design tool for assembling concurrent components by means of various models of computations [5]. Kepler provides the scientists with a graphical user interface to register and discover resources, and interactively design and execute scientific workflows using emerging Web and Grid-based technologies to distributed computations. Kepler is unique in that it seamlessly combines high-level workflow design with execution and runtime interaction, access to local and remote data and legacy applications, and local and remote service invocation along with a built-in concurrency control and job scheduling mechanism.

Computational units in Kepler are called *actors*, which are reusable components that communicate with each other via input and output ports. Actors are linked to each other to compose a scientific *workflow*. The workflow execution is orchestrated by a *director* that schedules component interaction. In this paper we utilize the Kepler environment to discover seismic events data, compose and execute ArcIMS services in a scientific workflow to automate and visualize the integration between seismic events and their related beach balls symbols.

Spatial data are served by a variety of distributed heterogeneous servers that include Oracle Spatial, DB2, ArcIMS and WMS servers. All data sources are wrapped in WSDL/SOAP wrappers, so that they can exchange and process SOAP messages. The source output, an XML document, possibly referencing to a map image or a compressed feature set, is returned wrapped in SOAP response.

Several generic web services actors have been implemented in Kepler that serve as clients for accessing distributed resources within Kepler workflows. Specifically, the *WebService* actor provides the user with a simple plug-in mechanism to execute any SOAP Service. The user can instantiate the generic web service actor by providing the WSDL URL and choosing the desired web service operation. The actor then automatically specializes itself and adds ports with the inputs and outputs as described by the WSDL. The so instantiated actor acts as a proxy for the web service being executed and links to the other actors through its ports. Using this component, any application that can be deployed as a remote service, can be used as a Kepler component.

For plugging-in a larger set of, possibly related, services, Kepler also provides a web service *harvester* capability for importing web services from a repository. Another means to import web services is to use the Kepler search functionality to discover and access web services registered in heterogeneous, distributed, repositories, for example, those accessible through the GEON portal.

3. Spatial Information Processing

GIS on the Internet is gaining popularity in the scientific community as it enables a wide audience to share and analyze spatial data. Even though there is a wide variety of products to serve the scientific spatial data on the Internet, such as ESRI's ArcIMS, Minnesota Map Server and GeoServer, ArcIMS is becoming the de-facto standard. Along with a server-side interface, ArcIMS also provides a programming API, called Java Connector API, so that all the functionality which is available through the application can also be exploited programmatically. This API can be used to generate dynamic maps and can be combined with other APIs and tools to perform interesting analysis and science applications as a part of a process in a workflow environment like Kepler.

Following the patterns of service oriented architecture, ArcIMS introduced the concept of an *Image Service*, which is composed of one or more shapefiles. It also includes the rendering information for each dataset. This service provides a flexible interface to support critical operations, such as map creation and query of data and metadata. The communication between client and server is achieved using an XML based query language called ArcXML.

GEON provides web service wrappers around ArcXML for certain common tasks, thus simplifying their use. Some of this functionality that we have been using with Kepler includes:

1. **GET_SERVICE_INFO:** This request returns metadata information, list of layers, their schemas, bounding box and rendering information in an ArcIMS Service.
2. **GET_FEATURES:** This request is used to query the features for extracting the vector data in XML format from an ArcIMS Image Service, within a given envelope along with an attribute query.
3. **GET_IMAGE:** This service responds to the request for rendering the image map for the layers the user has selected from the image service.

Spatial data is often in different formats, in order to support on the fly mapping of various formats we have created web services for format conversion and map generation. Two such

format conversion services are *ASCIIToMap* and *XMLToMap*. Another service, which is a classic example of a JConnector Wrapper web service, is *ArcIMS_Image_Service_Creator*.

The details of the data conversion services are as follows:

ASCIIToMap. This service is designed to consume spatial data in ASCII format. A typical example would be the point vector dataset that expresses each point as a combination of X,Y (in some cases X,Y,Z) along with a set of attributes. This service generates a shapefile as an output along with the creation of an Image Service and a mapping interface on the fly.

Example ASCII File Format:

```
Longitude Latitude Magnitude ...
143.002 37.2850 2.25000E-02
-106.081 -4.33800 2.50000E-02
151.750 45.3590 3.75000E-02
```

Each line in the file, except for the first, represents a spatial point information. The first line contains space separated strings, representing column names. The first two strings are always expected to be “Longitude” and “Latitude” respectively, indicating that the first two columns in the file represent location. All the remaining column names are considered as attribute names and their values are considered as attribute values. All values, that is, Longitude, Latitude and attributes are space separated.

XMLToMap. This service is similar to *ASCIIToMap* described above with a different input format. This service expects an XML file input, in the format described below:

XML File Format:

```
<?xml version="1.0" encoding="UTF-8" ?>
<table>
  <record>
    <spatial>
      <lat>33.42227</lat>
      <lon>-111.82264</lon>
    </spatial>
    <attributes>
      <name>ASU</name>
      <hostname />
      <ip />
      <nodetype />
      <version />
      <clusterinf />
      <status>down</status>
    </attributes>
  </record>
</table>
```

<table> is the root element for the document. Each <record> element represents a spatial point information. The child element <spatial> represents the spatial coordinates of the point. <attributes> represent the attributes attached to this point. Tag-names within <attributes> are attribute names and text nodes are attribute values for respective attribute names.

ArcIMS_Image_Service_Creator. This service creates an ArcIMS Image Service remotely, thus eliminating the need for client applications (e.g. the Kepler system) to execute on the same machine as ArcIMS. The input to this service are unique geon ids of shapefile datasets registered with the GEON portal and stored using the SDSC Storage Resource Broker (SRB) [6], a grid-enabled data management system. These ids are used to download data stored on the Storage Resource Broker (SRB). Then this data is used to create the ArcIMS Image Service. Finally, the name of the created service is returned.

Sample Input:

```
<layers>
  <layer>25dfb3db-e710-11d8-b226-ab22ed7681c0</layer>
  <layer>6d5da44e-e80a-11d8-8e79-6747442646a7</layer>
  <layer>f1ed5760-e80a-11d8-8e79-6747442646a7</layer>
</layers>
```

4. Focal Mechanisms Workflow-Based Modeling

In the following section we present the workflow approach for overlaying seismic events' focal mechanisms using image services. We describe the preparation and processing steps to achieve our goal and provide a discussion on the derived products.

4.1. Workflow Preparation

Two inputs are required for executing the workflow: a dataset resource (e.g. a shapefile, database, ASCII file, netCDF file) that contains seismic events information, that is, location, magnitude, direction, etc., and a beach balls symbols images database identified by a focal mechanism's strike, dip and slip.

In GEON, discovery and access to distributed heterogeneous datasets is available through a GEON portal function called GEONsearch. This assumes that datasets have been "semantically" registered beforehand, through the GEON portal. The registration information includes metadata annotations based on the ADN framework [7]. Once a dataset has been registered, it is searchable based on its metadata or other semantic annotations. For the purpose of earthquake analysis, various seismic events datasets have been registered and made publicly available. Kepler allows to discover and access them through its dataset search tab. For example, a CMT [8] focal mechanism shapefile dataset has been registered. The dataset contains location, time, and focal mechanism solutions for earthquakes that occurred between 01/01/1977 and 08/31/2000 with surface wave magnitude greater than or equal to about 5.5.

Creating the beach balls database was done using the Generic Mapping Tool (GMT) [9]. Specifically, the *psmecha* function is used to plot focal mechanisms with a 1cm diameter, using strike, dip and slip as variables. A script was written to populate a beach balls database table by creating a beach ball symbol per each variable degree. The database is available at: <jdbc:db2://data.sdsc.geongrid.org:60000/FOCALMEC>¹.

4.2. Workflow Model

Figure 1 provides a snapshot of the integration workflow. The workflow composition process is as follows: The user first searches for the desired seismic events datasets using the Kepler dataset tab which accesses the GEON metadata catalog and performs a “semantic” (ontology-aware) match. Discovered datasets (i.e., shapefiles, databases, ASCII or netCDF files) appear in the top left pane of the Kepler graphical user interface, identified by a geonId and the dataset title. The datasets can be dragged and dropped onto the workflow canvas to be accessible within a scientific workflow.

The web service client actor is used to access an ArcIMS service wrapping the discovered dataset with an image service and later query it to getFeatures (for this scenario, we wrapped the ArcIMS GET_FEATURES request as a web service). Query results are returned from the server as ArcXML documents. The interesting features, that is, location, magnitude, direction, strike,

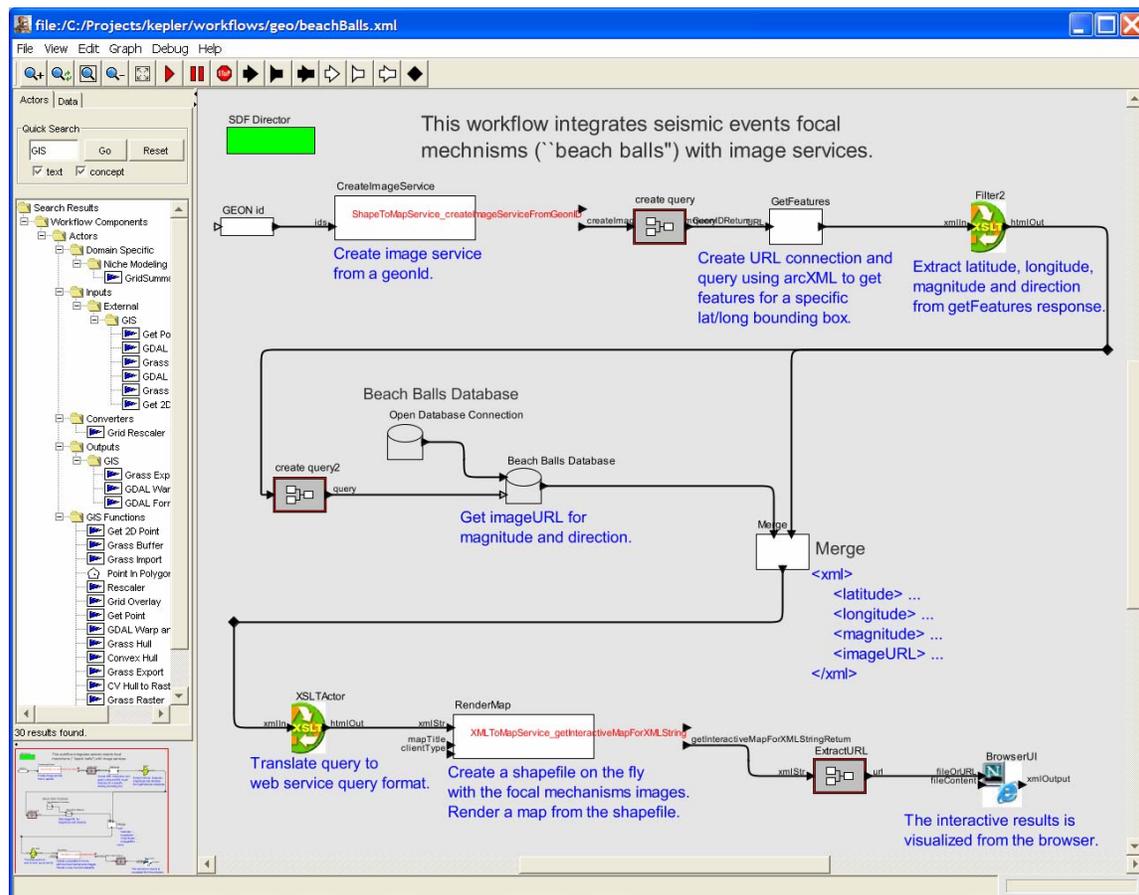


Figure 1. Integrating beach balls with image services in Kepler

¹To request read access to the database please contact efrat@sdsc.edu

dip and slip are extracted from the results. The strike, dip and slip values are used to query against the pre-calculated beach balls images database to get the matching image URL. A merged XML result is then created with a beach ball image link per each event location.

Finally, in order to display the integration between the seismic events location and their balls images using a background map, another ArcIMS service, called XMLToMap, is used. This service takes the XML result of the previous step and renders an interactive map, on the fly, of the beach balls images. The interactive map, retrieved by the client and displayed on a web browser, using a browser actor, allows the user to zoom in or out on events and to query for information on each beach ball icon.

4.3. Results and Evaluation

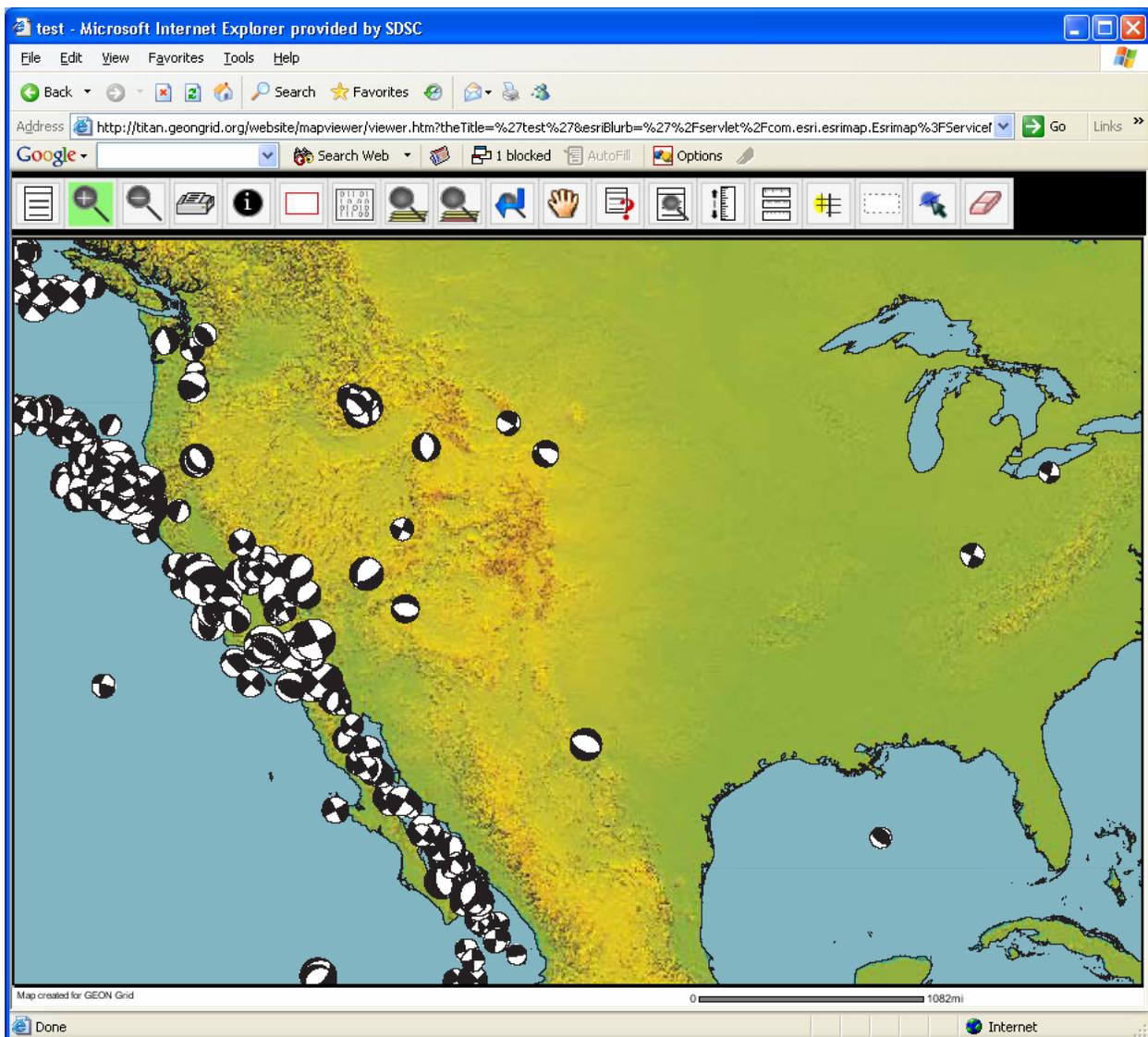


Figure 2. Overlaying seismic events focal mechanism on a map.

The scientific workflow environment provides us with the ability to plot on the fly seismic events' focal mechanisms on a map, accessing image services through adaptable web service interfaces. The execution of the workflow extracts the following:

1. A new dataset, in the form of a shapefile, is created on the fly from the merged XML representation, attaching each seismic event data with its matching beach balls symbol URL. The new dataset can be registered with the GEONsearch and used within other applications like ArcGIS.
2. An interactive map, displayed on a web browser. The map, depicted in Figure 2, displays the beach balls' images overlaid on a background map. The size of a beach ball is proportional to the magnitude of the respective seismic event. The resulting interactive map allows the user to retrieve additional information about each seismic event's attributes, such as, its location, magnitude, time, etc. This information can be accessed by clicking on a beach ball icon, which extracts the relevant data from the service's associated shapefile (as described above).

Kepler focuses on the reuse of existing components (actors) and workflows. This workflow was created using existing actors, such as the web service invocation actor, an XSL transformation actor, and database access actors, without having to create new components. Setting up the beach balls database and wrapping the above mentioned ArcIMS services required little effort. Thus, the workflow environment proves to be beneficial for solving complex tasks involving remote execution environments and generic analyses.

A caveat to this approach is that intermediate results generated by ArcIMS services may consist of huge XML files (ArcXML), which might overwhelm channels with low bandwidth or slow servers. This problem can be alleviated by using ordinary space separated ASCII files instead of XML.

4.4. Execution Environments

GEON supports two types of execution environments: the web-based interface to access various resources and tools, called the GEON portal, and local applications. Kepler serves as a local desktop application; however, various workflows can and have already been deployed as web services and are available for portal environment execution as well. The workflow discussed in this paper was initially designed as a local application that provided access to numerous distributed computations. The workflow was also compiled into a web service that takes as inputs a dataset id (geonId) and the beach balls database authentication information, and returns a URL to the interactive map result. Thus it can also be available for execution from the GEON portal, redirecting the resulting URL back to the client page. The user is then able to search for seismic events datasets using the GEONsearch function and apply matching geonIds as workflow inputs. A future step is to register workflows within the GEON portal and then make them searchable by subject/keyword.

5. Summary

In this paper we have presented a scientific workflow solution to mapping and visualizing seismic events' focal mechanisms using ArcIMS services wrapped as web services and accessed through the Kepler workflow system. Providing scientists an easy-to-use environment enables them to be more flexible and efficient in their work. Via the workflow tools we developed, scientists now can assess a variety of datasets in the GEON network and overlay earthquake focal mechanisms to study spatial and temporal relationships in these datasets.

Other Kepler workflows that utilize ArcIMS services include a workflow for integrating heterogeneous State Geologic Maps using rock and geologic age ontologies. This workflow demonstrates how, using ArcIMS services, data sources of various formats can be queried uniformly and later be assembled together to a map. Heterogeneities include structural and representational differences as well as semantic differences (semantic differences are the result of using different geologic conventions and scales).

Another example is the “on-the-fly” extraction of datasets, that is, querying for information from a database or an ASCII file and creating a shapefile on the fly along with a visual image of the new dataset that can be displayed on a browser. ArcIMS services are also used in a rock classification workflow, integrating between data queried from a relational database and a spatial database, creating a shapefile that contains the combined information and rendering an interactive map that uses the ArcIMS viewer to retrieve additional data about the classification result.

Kepler along with ArcIMS services provides a visual programming environment that can efficiently harvest, compose, execute and monitor GIS functions to automate geoscientists' analyses to enable new earth science discoveries.

Acknowledgment

GEON is an NSF ITR Collaborative Research project (#0225673) to Create Cyberinfrastructure for the Geosciences.

References

- [1] GEON: Cyberinfrastructure for the Geosciences, <http://www.geongrid.org>
- [2] GEON Portal: <https://portal.geongrid.org:8443/gridsphere/gridsphere>
- [3] Kepler: An Extensible System for Scientific Workflows, <http://kepler.ecoinformatics.org>
- [4] B. Ludäscher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger-Frank, M. Jones, E. Lee, J. Tao, Y. Zhao, Scientific Workflow Management and the Kepler System, Concurrency and Computation: Practice & Experience, Special issue on Scientific Workflows, to appear, 2005
- [5] Ptolemy II, <http://ptolemy.eecs.berkeley.edu/ptolemyII>
- [6] The SDSC Storage Resource Broker, C. Baru, R. Moore, A. Rajasekar, M. Wan, in Proc. CASCON'98 Conference , Nov.30-Dec.3, 1998, Toronto, Canada.
- [7] ADN Metadata Framework, <http://www.dlese.org/Metadata/adn-item/index.htm>
- [8] CMT Harvard Seismology, <http://www.seismology.harvard.edu/CMTsearch.html>
- [9] GMT: The Generic Mapping Tools, <http://gmt.soest.hawaii.edu/>

Authors Information

Efrat Jaeger-Frank

San Diego Supercomputer Center, University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0505

Phone: (858) 822-3694, fax: (858) 822-3693

efrat@sdsc.edu

Ashraf Memon

San Diego Supercomputer Center, University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0505

(858) 822-0017

amemon@sdsc.edu

Ilkay Altintas

San Diego Supercomputer Center, University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0505

(858) 822-5453

altintas@sdsc.edu

Ghulam Memon

San Diego Supercomputer Center, University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0505

(858) 822-0918

gmemon@sdsc.edu

Dogan Seber

San Diego Supercomputer Center, University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0505

(858) 822-5409

seber@sdsc.edu

Bertram Ludaescher

Department of Computer Science, University of California, Davis
One Shields Avenue, Davis, CA 95616

(530) 754-8576

ludaesch@ucdavis.edu

Chaitan Baru

San Diego Supercomputer Center, University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0505

(858) 822-5035

baru@sdsc.edu