CONCURRENCY AND COMPUTATION: PRACTICE AND EXPERIENCE
Concurrency Computat.: Pract. Exper. 2000; **00**:1–7 *Prepared using* cpeauth.cls *[Version: 2002/09/19 v2.02]*

An Application portal for Collaborative Coastal Modeling

Chongjie Zhang[1,2,*], Chirag Dekate[1,2],
Gabrielle Allen[1,2], Ian Kelley[1], Jon MacLaren[1]

[1] *Center for Computation & Technology, 302 Johnston Hall, Louisiana State University, Baton Rouge, LA 70803, USA.,*
[2] *Department of Computer Science, Louisiana State University, Baton Rouge, LA 70803, USA,*

SUMMARY

We describe the background, architecture and implementation of a user portal for the SCOOP coastal ocean observing and modeling community. SCOOP is engaged in real time prediction of severe weather events, including tropical storms and hurricanes, and provides operational information including wind, storm surge and resulting inundation, which are important for emergency management. The SCOOP portal, built with the GridSphere Framework, currently integrates customized Grid portlet components for data access, job submission, resource management and notification.

KEY WORDS: Grid Computing, Portal, Coastal Modeling, SCOOP, GridSphere, GridPortlets

1. Introduction

The SURA Coastal Ocean Observing Program (SCOOP) [1] is representative of a growing class of geographically distributed collaborators who have realized the need for new infrastructures, such as Grid computing, to support their research work with complex applications requiring sharing of expertise, software, and data across multiple institutions. Building the necessary infrastructures for collaborative projects such as SCOOP involves integrating multiple Grid middleware packages to provide a holistic approach to collaborative problem solving.

Portals have become a popular way to integrate applications and content, providing groups of users (or *virtual organizations*) with a single entry point to interact with their applications, data, colleagues and services, all the while maintaining a uniform interface. As new applications

*Correspondence to: CCT, 302 Johnston Hall, Louisiana State University, Baton Rouge, LA 70803, USA.
Contract/grant sponsor: ONR; contract/grant number: N00014-04-1-0721
Contract/grant sponsor: NOAA; contract/grant number: NA04NOS4730254

Received 29 November 2005
Revised 15 December 2005

and technologies, such as Grid computing, become increasing complex and difficult to configure and use, *Grid portals* have come to be recognized as a useful tool to enable the work of scientists and engineers without burdening them with the low-level details of underlying technologies.

This paper illustrates how portal technologies can complement Grid middleware to provide SCOOP scientists with an easy-to-use collaborative infrastructure that is tailored to their particular needs and has the ability to incrementally introduce and test new capabilities and services. Section 2 of this paper describes the coastal modeling scenarios and requirements, Section 3 introduces the design and architecture of the portal, Section 4 provides implementation details and information about the different services the portal provides and Section 5 looks to future development. Finally, Section 6 presents the conclusions of our work.

2. Collaborative Coastal Modeling

The recent devastation to coastal Louisiana by Hurricanes Katrina and Rita, which cost the lives of over 1000 people and severely damaged the economy and the environment, has emphasised the need for accurate models of hurricanes and other severe weather events. Such accurate models are needed to predict the path and effect of impending hurricanes for evacuation and preparation, to design better coastal defense systems, and to understand the physics and trends of hurricanes.

While model fidelity is improving through the inclusion of more physical features and better algorithms, as well as increased computational power, the experience of researchers is still essential to tune models and interpret their output. SCOOP aims to integrate distributed real-time ocean observing stations and regional coastal modeling entities, to run ensembles of numerical hydrodynamic models for the prediction, verification and visualization of critical storm surge and wave behaviour during severe storms and hurricanes. SCOOP is addressing several needs in reaching their goals including: ubiquitous and easy access to all data, e.g. sensor, satellite, model, visualization data; automated deployment of models across heterogeneous resources, including complex workflows and ensembles; creation of data standards and interoperability of model codes; capabilities for coupled and multi-scale models; operational procedures which can provide GIS visualization and notification to emergency management personnel. Building an infrastructure to meet these needs and supply timely information about severe events requires attention to reliability, fault tolerance, scheduling, as well as end user presentation and interaction.

The current SCOOP members include a combination of research institutions, university programs and national agencies such as the National Oceanic and Atmospheric Administration (NOAA). This myriad collaboration engages researchers with diverse skill sets and varying degrees of technical expertise.

2.1. Infrastructure

Currently, coastal researchers typically access data from multiple sources (e.g. wind fields from NCEP or USGODAE, hurricane tracks from NHC, observation data from coastal observatories like WAVCIS or SEACOOS) using HTTP, FTP or more recently the LDM [2] protocols.

Prepared using cpeauth.cls

Operational workflows are deployed using "cron" type scripts, which are hard to adapt to address unreliable file arrival or fault tolerance. Usually the models involved in SCOOP (e.g. ADCIRC, WWIII, SWAN, WAM, CH3D, ELCIRC) are run only at local sites, and may require many different configuration and input files. The various institutions deploy their own web servers, delivering results at different times in varying data formats and data descriptions.

Activities in SCOOP and other projects are addressing the complexity of dealing with different data sources and formats. A prime need is to develop data standards to facilitate sharing and collaboration. In lieu of a formal standard, SCOOP has developed a file-naming convention throughout the project that encodes enough information to serve as primary metadata. LSU has established an advanced Grid-enabled data storage archive service [3], providing essential features such as digestion and retrieval of data via multiple protocols (including GridFTP and HTTP), a logical file catalog, and general event-based notification. Compute resources for researchers are available in the form of the SCOOP Grid which is comprised of resources distributed across multiple institutions (Louisiana State University, University of North Carolina, MCNC, University of Florida, Virginia Institute of Marine Sciences). Basic Grid middleware such as Globus Toolkit 3 (GT3) is deployed across the resources. Condor [4] is deployed across the LSU-SCOOP Grid for prototyping scheduling and job management scenarios. An ongoing goal is to be able to coordinate the deployment and scheduling of operational and research simulations across the SCOOP Grid.

3. SCOOP Portal Design

The SCOOP Portal provides the SCOOP user community with a centralized gateway mechanism to submit and manage coastal model simulations and keep track of a large number of data files. To better understand portal requirements, we worked closely with coastal researchers and developed use-case scenarios which have driven the design of the portal.

3.1. Use-Case Scenarios

After consulting with the SCOOP coastal researchers and modelers, we identified different requirements for the SCOOP portal development. One notable requirement was that although the scientists wanted to restrict access to data to those in the collaboration (for one reason, to address potential problems with casual interpretation of severe storm data), no finer grained access control was required. Additionally, all machines in the SCOOP Grid are shared resources among scientists, simplifying authorization needs considerably. In implementing this first version of the SCOOP portal, we concentrated on the following two user scenarios.

Archive Access The SCOOP project has set up an archive service to store source atmospheric data, wave/surge data from model simulations, and other observed data for model accuracy verification. The SCOOP Portal should provide functionality to facilitate modelers and researchers in querying and retrieving datasets. The steps for accessing data files are as follows:

1. A user selects a data class and specifies corresponding metadata to query a metadata catalog service to discover datafiles of interest, e.g. "Output datafiles from ADCIRC

model simulations performed at LSU for the Gulf of Mexico region, during August 2005". A list of Logical File Names (LFNs) are returned from the query to the user.
2. The user can select one or more LFNs of interest, and then the portal contacts the archive's logical file service to return the physical file locations to the user.
3. The users can choose either to download the data file to the local machine or to perform a third-party transfer via a range of protocols.

Model Simulations One of the scientific objectives of SCOOP is to run an ensemble of hydrodynamic models driven by input conditions from a range of different atmospheric models. The steps for running hydrodynamic model simulations are as follows:

1. A user is required to retrieve a proxy credential to authenticate to Grid resources.
2. The user specifies metadata describing atmospheric data and a hydrodynamic model. The SCOOP Portal contacts a metadata catalog service and the archive's logical file service to get atmospheric data files of interest. Based on the data files, the SCOOP Portal constructs a list of possible simulations, depending on the available input files in the archive. Each of these simulations will then be submitted to a job scheduler.
3. The user then can track the progress of each simulations via the SCOOP Portal or use the portal's notification services, which include AIM and email.
4. Upon successful completion of each simulation the results are pushed into the archive for dissemination and further processing.

3.2. Portal Development Toolkits

The requirements from the SCOOP community are still evolving. The use of a mature portal framework and a well-designed Grid portal toolkit was necessary to be able to focus on the business logic of the SCOOP use-case scenarios and to allow for extensibility to future requirements. There are a number of portal frameworks including IBM WebSphere, BEA WebLogic, uPortal and GridSphere [5], and Grid portal toolkits including GridPortlets [6] , GridPort [7], OGCE [8], and Java COG [9]. Based on our comparative analysis of GridPortlets and OGCE [10], we chose GridSphere and GridPortlets as our main toolkits to speed up the process of developing and deploying an application portal for SCOOP modelers and researchers.

GridSphere is a free, open-source portal framework developed by the European GridLab project [11], which focused on developing Grid application tools and middleware. GridSphere provides a well documented set of functionality, including portlet management, user management, layout management, and role-based access control. Its portlet-based architecture offers flexibility and extensibility for portal development and facilitates software component sharing and code reuse. GridSphere is compliant with the JSR-168 portlet specification [12] which allows portlets to be developed independently of a specific portal framework. GridSphere's portlet service model provides developers with a way to encapsulate reusable business logic into services that may be shared between many portlets.

The advantages of using GridSphere come not only from its core functionalities, but also from its accompanying Grid portal toolkit, *GridPortlets*. GridPortlets abstracts the details of underlying Grid technologies and offers a consistent and uniform high-level service API,

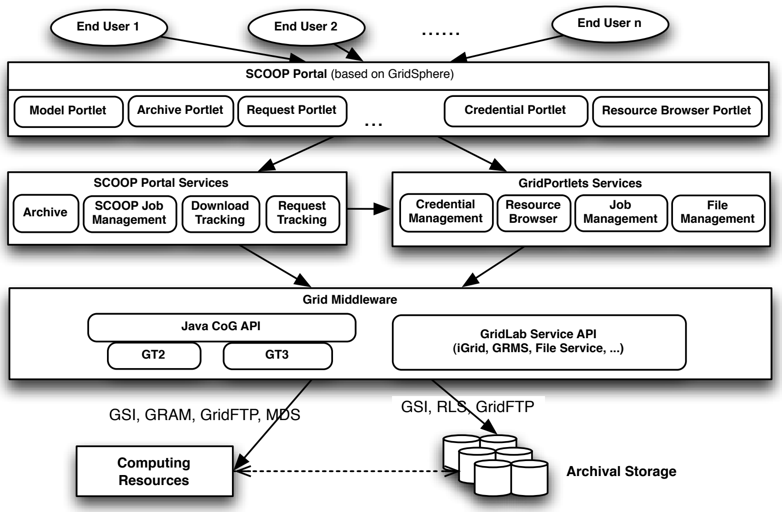

Figure 1. SCOOP Portal architecture based on GridSphere.

enabling developers to easily create custom Grid portal web-applications. The GridPortlets services provide functionalities for managing proxy credentials, resources, jobs, and remote files, and supports persisting information about credentials, resources, and jobs submitted by users. The GridPortlets service API currently supports both GT2 and GT3. In addition, GridPortlets delivers five well-designed, easy-to-use portlets, which include: resource registry, resource browser, credential management, job submission, and file management.

3.3. Architecture

The architecture of the SCOOP Portal is based on the GridSphere framework. Figure 1 shows the SCOOP Portal software components and their interactions and relationships. From the simplified diagram, it can seen that SCOOP portlets use SCOOP services for application-specific functionality and business logic. The SCOOP Portal services themselves use or extend services built into the GridSphere framework, the GridPortlets package, as well as some third-party packages. For example, the SCOOP portal services mainly use the GridPortlet service API to interact with Grid resources, such as submitting jobs and moving remote files. Most portlets are independent from one-another, however, they can communicate with each other via the service layer. For example, the credential portlet calls the credential service to retrieve proxy credentials from a MyProxy [13] server, and later a job submission portlet can get the retrieved credentials to authenticate with Grid resources. This portlet-based, service-oriented, architecture greatly speeds up portal development and exhibits high extensibility.

4. SCOOP Portal Implementation

The SCOOP Portal is implemented with GridSphere version 2.0.3 and GridPortlet version 1.0, which together provide a coherent core set of relevant functionality. When evaluating the SCOOP community requirements, we found several common functionalities had already been implemented in other Grid portal projects. To avoid reinventing these, the fully deployed SCOOP Portal contains not only portlets developed specifically for the SCOOP, but also shared portlets from GridSphere, GridPortlets, and GridLab Testbed [14]. The following list illustrates the functionalities of portlets and services that are specific to the SCOOP project:

- Archive – enables users to retrieve, either by HTTP to the local machine or GridFTP to a remote machine, SCOOP data files from archive storage. The interfaces provides queries using metadata, and custom interfaces to specific data formats such as OpenDAP.
- SCOOP Job Submission – provides custom interfaces and options for users to launch coastal models. The interface matches models to available data files in the archive.
- Simulation Tracking and Notification – allows users to track the progress of active simulations via the SCOOP Portal and to receive notification via email or AIM.
- Request Tracking – coordinates team work by allowing users to manage and track the status of tasks and defects in various SCOOP sub-projects.
- Download Tracking – tracks downloads of software tools distributed through the portal.

The following list illustrates the Grid-related functionalities of portlets and services that are deployed to the SCOOP Portal but developed by other Grid projects. All other portlets are from the GridPortlets distribution except the Grid Resource Status Monitoring portlet that is developed by the GridLab project.

- Credential Management – enables users to retrieve, renew, and delete proxy credentials from a MyProxy server.
- Resource Registry – enable portal administrators to register or unregister Grid resources, such as computing resources or services.
- Resouce Browser – enable users to view available Grid resources, including information about hardware configuration, services, job queues, and accounts on remote machines.
- Grid Resource Status Monitoring – enable users to view the information about whether particular services and software components are installed and available on each machine, and the possible reasons for the services that are not available.
- Physical File Management – enable users to browse and manage files on remote machines.

4.1. Archive

The current archive storage contains three classes of data files: source atmospheric data, simulated wave/surge data, and other observed data to verify the model accuracy. The three class of data are associated with class specific metadata attributes and query interfaces. Figure 2 shows the archive portlet for querying for data files.

The archive portlet gathers metadata information from user requests and retrieves a list of logical file names which have matching metadata from the archive portal service. Currently,

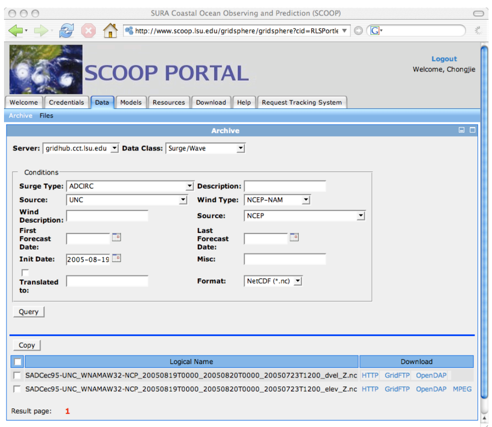

Figure 2. Archive portlet

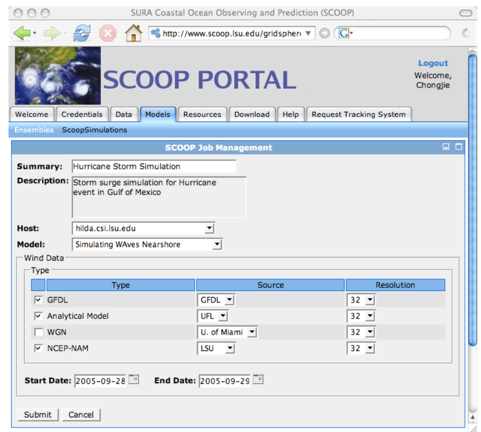

Figure 3. Job submission portlet

SCOOP does not have an appropriate metadata service, the logical file name used in queries is generated from metadata information contained in the SCOOP file-naming convention. The logical file name may contain wildcard characters to accommodate unknown metadata or to select a range of files. The mappings of physical file names and logical file names can be obtained by querying the archive's logical file services, currently provided by an instance of Globus Replica Location Service (RLS) version 2.2 [15]. To provide performance scalability, the query results are shown with dynamically paging techniques, because the RLS API does not support merely returning the size of matched results. Users can retrieve SCOOP data files via HTTPS to their local machine, or perform a third-party transfer via GridFTP. The transfer service is built on the GridPortlet file services and resource services for directory selection and file copy. Currently the logical file entries in the RLS point to the physical file entries available locally. Efforts are underway to incorporate distributed storage resources including SCOOP data store instances at TAMU. The LSU SCOOP archive is also being expanded to leverage SRB based terabyte scale storage at SDSC. Evolving versions of the SCOOP archive access portlet will address issues of accessing federated data stores.

4.2. SCOOP Job Management

From the user scenarios, SCOOP models can run on different input data types (e.g. using wind data generated either by analytic means or by other models). As shown in Figure 3, the SCOOP job submission portlet allows users to select multiple different types of wind data for a particular model. Using specified metadata, the job submission portlet queries the logical file service and generates a file list for each selected data type, and constructs a list of such tasks. The SCOOP job service submits each task to Condor via Globus GRAM to run the model on each file list. Hence one SCOOP simulation job may contain several sub-job runs.

The SCOOP job submission is mainly built on the GridPortlet service API. To store custom job information and the parent-child job hierarchy, we provide a persistent layer for SCOOP job management using Hibernate.

Active proxy credential is required for job submission. Users need to delegate their Grid credentials to the MyProxy server. The SCOOP Portal allows users to retrieve credentials from a MyProxy server via GridPortlets' credential management portlet. The job submission service will automatically use retrieved credentials for authenticating with Grid resources.

The SCOOP job portlet allows users to view the status and output of each sub-job run. Notification of the job status, currently by AIM or email, is implemented by another service: Simulation Tracking and Notification. Each sub-job registers itself with and continuously sends updated information to the notification service via XML-RPC. The notification service collects and sends out this information via email or AIM, depending on the user preference.

Currently, the functionality of the SCOOP job management interfaces is limited by the lack of interoperability of the underlying models and data sources. As ongoing work is completed, more complex workflows and ensemble runs will be implemented.

5. Future Work

In the next year of the SCOOP project, we will add new scenario and usecase driven features to the portal. These enhancements will include advanced model simulation interfaces that will facilitate ensemble simulations by providing modelers the ability to run a spectrum of hydrodynamic models each with different input conditions. Job Management portlets for such scenarios would allow modelers to select priorities of model ensemble, manage watchdog agents, and host of other advanced capabilities.

While implementing the current version of the SCOOP portal, the SCOOP Grid team had agreed to utilize Globus 3.2.1 as a standard middleware for the testbed. Various SCOOP partner institutions have started testing Globus 4.0. Several services, such as the Reliable File Transfer Service, could help provide better infrastructure as a base for the next version of the SCOOP Portal. Given the nascent nature of collaborative coastal modeling, a community driven initiative to agree on common metadata terms is currently under way in the form of Marine Metadata Initiative. Based on the status of such initiatives future metadata driven services could also be considered.

Tracking of these model results would be facilitated by development of new job monitoring interfaces with features such as color coded job lists and more. Post analysis of these ensemble runs could potentially be cumbersome if appropriate interfaces are not designed in close collaboration with the modelers. Currently modelers use standalone GIS Clients to visualize such ensemble modeling data. Integration of GIS technologies into the Portal framework, and designing products that would allow end-users to compare ensembles using time series graphs in addition to geo-referenced interactive interfaces, would greatly enhance the state of the art.

The current notification service will be rewritten to provide a generic service, extending to more communication mediums, and able to inform coastal modelers of impending hurricanes, initiate coastal modeling workflows or provide status updates from initiated workflows. Such generic services could potentially be reused in other projects that rely on notification services to provide value added functionality.

We plan to add several portlets which provide new services, for example integrating smart components that anticipate data requirements and can watch for data availability, or that can assist in the generation and configuration of model input files. One example for this is connecting to a service which can dynamically deliver analytical wind fields matched to the computational mesh for models on user selected geographic regions. Current SCOOP data access user scenarios do not incorporate real time streaming of data, in anticipation of such scenarios, we plan to study similar real-time streaming issues as addressed by projects such as LEAD [17]. Additionally we plan to investigate reusable components from related Earth Science projects that relate to the SCOOP usecases. For example, the LEAD project portal provides portlets such as the GeoGUI that allow users to select regions of interest from an interactive map for running model ensembles.

Advancing on these features would require not only designing and implementing new portlets, but additional Grid services to provide capabilities such as generic notification, model description and metadata. SCOOP is one of several emerging Earth Science collaborations that includes projects such as LEAD, Earth Sciences Grid, GEON, NASA QuakeSim. Each of these portals provides unique means of handling issues including user authentication, data management, workflow management, application management (including legacy applications) and visualization. Investigating the approaches used by these projects could provide insights and reusable portlet solutions resulting in development of an Earth Sciences portlet repository that could provide services that address needs of the research community.

6. Conclusions

The development of the SCOOP Portal has shown how the integration of Grid technologies with portal frameworks can provide a community with new collaborative tools to better access data, resources and information, with the end goal of enabling better science and dissemination of storm related information. While the portal interfaces have thus far been well received by the SCOOP community, the challenge now is to make the portal an essential part of the scientists' usual working environment. This requires attention to several issues: firstly, most scientists find it hard to deal with Grid credentials. For instance our current implementation of the SCOOP portal assumes that a valid credential is held on a MyProxy server, which requires client software to be installed on local machines and of course procedures and policies for issuing Grid certificates. Projects such as PURSE [18] and GAMA [19] are developing mechanisms for authenticating solely through a portal. Incorporating authentication mechanisms inspired by such efforts, would go a long way to improving the usability and adoption of portal frameworks as versatile interfaces to utilize distributed resources.

In general, the development of portlets for new model scenarios needs to be simple enough that computer savvy coastal modelers are able to customize and produce portlets that cater to their demands. Finally, a full range of Grid services should be easily accessible through the portal, the GridPortlet API is a step in this direction, and the GGF SAGA working group is developing a general API for application oriented access to Grid services.

Acknowledgment

The authors wish to thank the GridSphere team for their comments and discussions throughout the design and development of the SCOOP Portal. This study was carried out as a component of the SURA Coastal Ocean Observing and Prediction (SCOOP) Program, an initiative of the Southeastern Universities Research Association (SURA). Funding support for SCOOP has been provided by the Office of Naval Research, Award #N00014-04-1-0721 and by the National Oceanic and Atmospheric Administration?s NOAA Ocean Service, Award #NA04NOS4730254. Additional support was provided by the Center for Computation & Technology at Louisiana State University.

REFERENCES

1. SURA Coastal Ocean Observing Program (SCOOP). September 20, 2005, `http://www1.sura.org/`.
2. Unidata Local Data Manager. Unidata. September 2005. `http://www.unidata.ucar.edu/software/ldm/index.html`
3. J. MacLaren *et al.*. "Shelter from the Storm: Building a Safe Archive in a Hostile World", to appear in *Proceedings of the The Second International Workshop on Grid Computing and its Application to Data Analysis (GADA'05)*, 2005.
4. D. Thain, T. Tannenbaum, and M. Livny. "Condor and the Grid", in *Grid Computing: Making The Global Infrastructure a Reality*, edited by F. Berman and G. Fox and T. Hey. John Wiley, 2002.
5. J. Novotny, M. Russell, and O. Wehrens. "GridSphere: A Portal Framework for Building Collaborations", in *Proceedings of 1st International Workshop on Middleware for Grid Computing*, Rio de Janeiro, 2003.
6. M. Russell. "GridPortlets Overview". February 2005. *http://www.gridsphere.org/gridsphere/html/mardigrasworkshop2005/02_gridportlets.pdf*
7. M. Dahan, M. Thomas, E. Roberts, A. Seth, T. Urban, D. Walling, J.R. Boisseau. "Grid Portal Toolkit 3.0 (GridPort)", in *Proceedings. 13th IEEE International Symposium on High performance Distributed Computing*, 4-6, pp.272 - 273, June 2004
8. Open Grid Computing Environments Collaboratory. *http://www.ogce.org/*. May 2005.
9. Gregor von Laszewski, Ian Foster, Jarek Gawor, and Peter Lane. "A Java Commodity Grid Kit", in *Concurrency and Computation: Practice and Experience*, vol. 13, no. 8-9, pp. 643-662, 2001
10. Chongjie Zhang, Ian Kelley, Gabrielle Allen. "Grid Portal Solutions: A Comparison of GridPortlets and OGCE", to appear in the *special issue GCE05 of Concurrency and Computation: Practice and Experience*, 2006.
11. GridLab: A Grid Application Toolkit and Testbed Project Home Page. August 12, 2005. `http://www.gridlab.org`.
12. The Java Community Process. "JSR 168: Portlet Specification v1.0". 2003. `http://www.jcp.org/en/jsr/detail?id=168`.
13. J. Basney, M. Humphrey, and V. Welch. "The MyProxy Online Credential Repository". In *Software: Practice and Experience*, 2005
14. P. Holub, M. Kuba, L. Matyska, and M. Ruda. "Grid Infrastructure Monitoring as Reliable Information Service", in *The 2nd European Across Grids Conference*, Nicosia, Cyprus, January 2004.
15. Ann L. Chervenak, Naveen Palavalli, Shishir Bharathi, Carl Kesselman, Robert Schwartzkopf. "Performance and Scalability of a Replica Location Service," in *13th IEEE International Symposium on High Performance Distributed Computing (HPDC-13 '04)*, vol. 00, pp. 182-191, 2004.
16. Douglas Thain, Todd Tannenbaum, Miron Livny. Condor and the Grid. In *Grid Computing*, John Wiley & Sons, May 2003
17. Linked Environments for Atmospheric Discovery (LEAD) Project. September 20, 2005, `http://lead.ou.edu/`.
18. PURSe: Portal-Based User Registration Service. The GRIDS Center. September 2005. `http://www.grids-center.org/solutions/purse/`
19. GAMA: Grid Account Management Architecture. San Diego Supercomputer Center. December 2005. `http://grid-devel.sdsc.edu/gama`