Enabling Computational Geoscience

Enabling Computational Geoscience

Sameera Jayaratna, Harsha Kumara, Nipuni Perera, Ishara Ranatunga,
Shahani Markus Weerawarana

Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka

ABSTRACT

The bond between science and computing has become stronger with the advent of massively parallel computers and the resultant high performance computational (HPC) capabilities. Geoscience is one of the most prominent research areas involved in modeling and analyzing the evolvement of Earth systems for the betterment of mankind. Scientific computing is emerging as a powerful tool for enhancing predictions and decision making related to geoscience as well. Experiments in geoscience are often long running, data intensive and consist of heterogeneous geospatial data. Yet development of effective tools to deal with these complex experiments and processes remain a major challenge. This challenge, along with the lack of low-level expertise amongst geoscientists in consuming the underlying computational infrastructure, is a significant issue in conducting complex geoscience experiments. Reflecting on this problem scenario, it is apparent that a common and convenient platform for geoscientists around the world is a crucial need. This paper discusses the current state-of-the-art computational software infrastructure that can be leveraged to address this need in the geoscience domain. The paper then presents an analysis of the architectural concerns of an ideal computational geoscience enabling solution.

1.0 INTRODUCTION

Geoscience experiments have been a dominant research area in recent years. One of the major characteristics in geoscience experiments is processing large scale environmental data to reach timely and important decisions. Scientific computing has laid the foundation for the assistance of geoscientists with the design and implementation of relevant automated tools and web services. Nevertheless the gap between scientists and computational resources continue to exist due to the complexity of the underlying technologies.

This paper presents a framework to distill the requisite capabilities, tools and technologies in order to enable effective computational geoscience by analyzing the current state-of-the art in computational geoscience. Section 2 gives an overview of geoscience and geoscience experiments. Section 3 identifies the current state of technological support and infrastructure for geoscience research, workflow management systems, and evaluates their capabilities and issues. Section 4 contains an overview of geoscience standards and their usage by geoscience data providers and web services. Based on these discussions, section 5 gives a high level gap analysis of the available geoscience capabilities and features. Architectural concerns in computational geoscience are identified and discussed in section 6, followed by the presentation of an analysis framework that synthesizes the architectural perspectives of a computational geoscience enabling platform in section 7.

2.0 COMPUTATIONAL GEOSCIENCE

Geoscience includes all the sciences such as geology, geophysics and geochemistry that study the structure, evolution and dynamics of the planet Earth and its natural mineral and energy resources [1]. In the recent past, the use of cyber infrastructure and other computing technologies to enhance geoscience research has increased exponentially. The effect of the industrial revolution during past decades and the rapid growth of world population have strongly impacted the nature and behavior of earth systems. Thus geoscience has become a prominent research area which mainly focuses on providing a better and more protective environment for all living beings.

Geoscientists engage in a vast number of research experiments to analyze the behavior of earth systems in order to predict the future evolvement of the earth. Geoscience experiments have many distinct characteristics; they are often long running; they involve massive scientific computations; they are data intensive; and they consist of heterogeneous, multidisciplinary and distributed geospatial processing workflows [2]. Geospatial processing workflows often include a sequence of operations that use a range of tools to collect, translate, and analyze distributed heterogeneous geospatial data.

3.0 ENABLING TECHNOLOGIES

3.1 Geographic Information Systems (GIS)

Geographic Information Systems (GIS) is specialized software which is essential for geospatial analysis, managing and interpreting geospatial data [3]. These systems can also centralize available distributed geospatial data and make them accessible in several forms such as image maps, digital maps and raw data files [4]. Modeling, simulation and visualization can be considered as major functionalities required in geospatial applications. GISs cater to the needs of scientists, policy makers and general public in geographic analysis and geographic modeling [3]. GRASS GIS is a popular open source GIS while ArcGIS is an example of a proprietary GIS.

3.2 Scientific Workflows

A workflow is an abstract model of a series of steps connected to represent a real world process where each step defines a specific task or functionality. Workflows can be categorized as business workflows and scientific workflows. One of the major differences is that, business workflows are control flow

	Apache Airavata	**Taverna**	**Triana**	**Kepler**	**Pegasus**
Workflow engine	Airavata Workflow Engine	Freefluo	Triana engine	Ptolemy II framework	DAGMan
Workflow description language	WS-BPEL	Scufl (language using XML)	BPELL, Triana specific language	Ptolemy's Modeling Markup Language	DAX - based on XML
User Interface	XBaya GUI	Taverna UI	Triana UI	UI from Ptolemy system	Pegasus web portals, WINGS
Features	Scalability, performance, monitoring, interoperability and on demand service creation	Scalability, reuse, interoperability, error handling, security and concurrency	Scalability, performance, error handling and support automating repetitive tasks	Scalability, reuse, process and data monitoring and provenance tracking	Scalability, performance, reuse, provenance tracking, error recovery and reliability
Involvement in geoscience	OLAM project	Used in PML	GEO 600 project	Support EarthGrid access	CyberShake project
Support for geoscience standards	-	WPS	-	OpenDAP	-
Specific science application areas	Biology, chemistry and oceanography	Bioinformatics, geoscience, astronomy and chemistry	Gravitational wave detection	Biology, ecology, geology, astrophysics and chemistry	Astronomy, bioinformatics and physics

Table 1: Comparative Summary of Workflow Management Systems [5],[6],[7],[8],[9]

oriented while scientific processes are data flow oriented. This paper considers scientific workflows.

Scientific workflow is the procedure of combining data and processes into a structured set of steps which can operate as a solution to a scientific problem. It includes a declarative description about each component and its input and output[10]. Workflows utilize distributed resources in order to access, manage and process a large amount of data from a higher-level [11]. The workflow concept is very important in e-science since; some complex e-science applications require creation of a collaborative workflow, many e-scientists lack the necessary low-level expertise to utilize the underlying computing infrastructure, and workflow specifications can be reused and shared once defined [12].

3.3 Scientific Workflow Management Systems

Scientific workflow management systems have evolved with the aim of producing a high level platform for the end-user scientists to create, manage, compose, execute and test scientific workflows while shielding them from the low-level computational technology. In the paper [10] scientific workflow management systems are recognized as providing means to; model and specify processes with design primitives, re-engineer developed processes such as verification and optimization, automating the execution of processes by scheduling, controlling and monitoring the tasks.

Currently there are many workflow management systems such as Apache Airavata, Triana, Pegusas, Taverna and Kepler, developed targeting various scientific communities [13]. Some significant aspects of a set of popular workflow management systems are categorized in Table 1.

3.4 Science Gateways

A Science gateway is a collection of tools, applications and data that allows scientists to run their applications with minimum concern of where the actual processing takes place or the underlying complexity of computational resources. Science gateways open up opportunities to access all services in one place and reuse services implemented by other scientists. The scope of the science gateways extends to fields such as bioinformatics, astrophysics, oceanography and geoscience.

Computational Science research is an amalgamation of computer architecture and scientific algorithms. Therefore it consumes considerable processing power with a large amount of data storage and transfer as in DNA sequencing computations. Cutting-edge technologies are required to cater to these requirements. Currently in the USA, there are several options available to scientists in terms of high performance computing resources. They are regional grids such as NWICG, Open Science Grid, TeraGrid (now XSEDE) and DOE etc.[14]. Science gateways mainly consist of the application layer, middleware layer and resource layer in high performance computing (HPC) infrastructure.

The major problems with science gateways are that they require significant effort and time to build, they need expertise in HPC and they tend to have a high rate of failure after implementation.

4.0 GEOSCIENCE RELATED STANDARDS

Geoscience standards span many areas and are centered around specific communities. The Open Geospatial Consortium (OGC) is the biggest and most eminent organization among other geo-related standard issuing organizations. OGC standards are the largest set of geo-related standards formulated thus far.

4.1 OGC Standards

OGC [15] is a non-profit organization which defines geospatial standards. The OGC standards “empower technology developers to make complex spatial information and services accessible and useful with all kinds of applications” [15]. WMS (Web Map Service), WFS (Web Feature Service), WPS (Web Processing Service) and WCS (Web Coverage

Service) are some of the OGC standards that are exposed as web services [16]. These services can be obtained by different communities where OGC's power of interoperability among the services gives a valuable output.

Two categorizations have imposed different levels of recognition on geo-software: OGC-compliant and OGC-implementing. OGC-compliant means that a specific software product has passed the relevant CITE (Compliance & Interoperability Testing & Evaluation) tests [17]. On the other hand OGC implementing means that an organization has developed a software product and may or may not have passed the tests [17]. Table 2 delivers a list of OGC compliant tools sorting them by means of their service with the details on supported standards.

5.0 CHALLENGES IN ENABLING COMPUTATIONAL GEOSCIENCE

Though there has been a significant technological contribution in geoscience,

Tool category		Available Products		OGC support				Other
				WMS	WFS	WPS	WCS	
Desktop GIS - Desktop GIS software provides complete and powerful set of GIS capabilities to assist in performing complex spatial analysis, spatial data creation, and visualizing data mostly in maps		Open source	GRASS GIS	✓	✓	✓	✓	Raster Data formats/Vector Data format
			uDig	✓	✓	✓		KML, GeoRSS
			QGIS		✓		✓	WCS-T, WFS-T, GML
Web map servers - Provide capabilities for viewing, editing and sharing geospatial data		Open source	GeoServer	✓	✓	✓	✓	Only implements WPS
			MapServer	✓	✓		✓	WMC, Filter Encoding, SLD,GML,SOS,OM
			OpenMap	✓				Plugin support for WMS
Spatial database management systems - Address the issue of processing and analyzing spatial data, providing convenient front-end for visualizing and manipulation of them		Open source	PostGIS					SFS
			SpatiaLite					SFS
			TerraLib		✓			Supports WMS and WCS with the TerraOGC extension
Software development frameworks and libraries - Support building geoscience applications with features such as visualizing, processing and analyzing	*a) Web*	Open source	MapFish	✓	✓			WMC, KML, GML
			OpenLayers	✓	✓			-
	b) Non-web	Open source	GeoTools	✓	✓			WFS plugin, WMS plugin
			GDAL	✓	✓		✓	WMS file format recognition, Separate drivers for WFS & WCS
c) OGC implementation frameworks - ***Serve user written functionality in OGC standards***		Open source	PyWPS			✓		-
			ZOO			✓		-
			52° North WPS			✓		-

Table 2: OGC Support in Geo-tools

``certain features are yet to be fully realized. Less focus on end-user interactions has caused negative impacts in achieving the expected outcome of these modern tools.

5.1 Need for Computational Expertise

Geoscience experiments often deal with extreme technologies involving complex computations running on HPC infrastructure. This makes end users reluctant to use the available sophisticated tools, either due to a lack of required geoscience knowledge or the lack of expertise in the underlying computational infrastructure. An ideal solution should focus on delivering a user friendly environment by hiding underlying complex computational resources that enable users to get familiar with the tool set with low computational expertise.

5.2 Regeneration of Common Geoscience Computations

Several popular workflow management systems are used by science communities according to their respective requirements. Some of these workflow management systems differ in several aspects including workflow engines and workflow description languages. Therefore a workflow created for a certain workflow management system is not compatible to be used in another system [18]. For an example, Pegasus and Taverna have been used as workflow management tools in many bioinformatics projects[9] [6]. Scientists in each of these projects are likely to construct similar workflows to run on respective systems as reuse is not possible due to above mentioned incompatibilities [18]. Lack of reuse may cause significant drawbacks in advancements of scientific research and experiments, especially considering the effort the scientists often require dealing with low level computational infrastructure details [18].

5.3 Handling Massive Data-Growth

Similar to the other sciences, in geoscience also, most research experiments involve massive data sets. As computations grow in complexity and reliability the analysis of the data they generate become more challenging[19]. Unavailability of software and technological support to handle data intensity often delay the progress of experiments. More comprehensive analysis of this data would help in the discovery and identification of unanticipated phenomena, and also help to expose shortcomings in the simulation methodologies and software.

5.4 Effective Tool Selection

There are a number of tools used in geoscience experiments relating to various research requirements as shown in Table 1. Some of these tools are publicly available and some are not, while the majority is in the form of a downloadable client. There are separate tools or components to handle different aspects of computational geoscience. Some of these are databases, web services based tools using desktop clients, web services based tools for browser-based clients, geo-processing applications, spatial tools which focus on co-ordinate handling, navigation and mapping tools, crisis management tools, map data and geospatial libraries.

6.0 SOFTWARE ARCHITECTURAL CONCERNS IN COMPUTATIONAL GEOSCIENCE

As we identified the limitations of existing technological support in geoscience domain, it is also essential to find remedies to overcome these limitations and challenges. In this section we explore the architectural capabilities that should be present in computational geoscience applications in order to handle a growth in data analysis and understanding proportional to the exponential growth in computing, data storage and other performance elements.

6.1 Spatial Data Visualization

Spatial data visualization is often tightly coupled with geoscience experiments. Distributed geo visualization systems are capable of collaborative synchronous and asynchronous visual exploration and analysis of geospatial data via the Web, Internet, and large-screen group-enabled displays [20]. Results of high data intensive geoscience experiments are meaningless until output data is mapped into a visualization tool for viewing and further analysis. Thus geo scientists often use visualization tools in their experiments where they render large amounts of data sets into geographic maps to visually analyze geospatial data.

6.2 Collaboration

Geosciences research demands not only heterogeneous data and models but also includes a range of experts. This requires the use of new architectural styles which support distributed processing and remote communications for successful collaborative and multidisciplinary research. Hence geoscience collaboration focuses on distributed computing, reviews on geospatial services, integration and reuse [21]. Amalgamation of several disciplines over distributed computing has emerged as the need for simultaneous analysis. Common cyber infrastructure is the best solution for such collaboration. Even in the same domain people have diverse software experience and expectations for a common cyber infrastructure. SOA is the best architectural style with loosely coupled concepts involved with simple user interface flows.

6.3 Reuse

Sharing information about research and experiments has been in practice among scientists for a long time. Today in the SOA domain, service computation is a key mechanism which allows for creation of value-added services by integrating and reusing existing services [21]. This mechanism exposes multiple reusable services to a wider community to be consumed for different purposes. Workflow sharing offers reuse in another level. Workflows provide capabilities to manage computations and allow capturing a full process electronically for sharing and reuse. In order to gain the full benefit of this mechanism, science communities require agreeing on process semantics and making the workflow representations as explicit as possible. Furthermore the workflows should be able to deal with different underlying infrastructures in different systems [22].

6.4 Ease of use

Nowadays sophisticated high level tools to deal with complex computations and experiments are available for scientists. One shortcoming of scientific research tools, which needs to be addressed, is its complexity and the difficulty of adapting. Generally the resource sharing, scalable facilities and accessing of Grid infrastructures have steep learning curves. Therefore much effort is required to make them more user-friendly. Thus while expanding the facilities and capabilities of the applications, users should be provided with easy interaction mechanisms along with clear assistive instructions and intuitive graphical user interfaces [23]. In an ideal solution researchers should be able to focus on their scientific research work with little concern about the underlying technologies. To accomplish this, the applications need to hide the complexity and expose required features to the user in a convenient manner.

6.5 Scalability

Scalability is a fundamental requirement in large-scale scientific experiments and has a great impact on application performance and completion. Recently reasonable efforts have been made in modeling and visualization technologies for utilizing resource pools for on-demand and scalable scientific computing. The nature of geoscience experiments includes variation of resource quantities and characteristics at runtime. Thus cloud computing enabled workflow engines would be a desirable feature for applications to facilitate dynamic scaling up and down. In addition, frameworks excessively need to access a range of computational resources in order to manage distributed geoscience applications.

6.6 Security

Security is one of the key features in a framework which supports secure access to cyber infrastructure by providing advanced interfaces for collaboration, analysis, data management, and other tools for students and researchers [24]. Systems need to access resources, computing cycles, instruments and data from cloud and grid environments on the researcher's behalf. Resource access often requires the use of researcher's security credentials which leads to potential compromise of the researcher's password [24].

Workflow management systems must capture and generate provenance information as a part of workflow security. Users demand that their research and personal information are kept protected, and resource providers demand that their computing and storage resources are used appropriately by authorized users [24].

6.7 Reproducibility

Reproducibility is a key factor in scientific analyses and processes. It enables scientists to evaluate and finalize the validity of each

Architectural Concerns	Type of tool	Approach	Limitations
Spatial data visualization	Desktop GIS	Provide views by rendering different types of images, animations and raster maps	Difficulties in handling heterogeneous computing environments
	Web Map Servers	Publish geospatial data through standards which can be used by visualization software and tools such as Google Maps	Limited OGC support (separate plugins are required in some instances)
	Software development frameworks and libraries	Facilitate the environment for building visualization applications by providing tools, APIs and standards support	Many software frameworks and libraries are specialized in specific areas
Collaboration	Scientific workflow management systems	Provide a platform for scientists to collaboratively create, execute and monitor workflows by using the services provided by various communities	Complex to build, No unified method of generating workflow, Limited to specific domains
Reuse	Scientific workflow management systems	Create interoperable workflows with other software systems by generating generic representations of workflows	Difficulties in service discovery and finding appropriate process
Ease of use	All	Hide complex computational resources and provide user friendly environments, Support universal standards, Provide standard APIs while hiding complex processing algorithm	Complexity of the domain make it difficult to create intuitive interfaces
Security	Scientific Workflow Management systems	Use security frameworks and security protocols	No framework addresses all the security concerns
Reproducibility	Scientific Workflow Management systems	Manage provenance information of workflows	Security concerns may raise when handling provenance data
Scalability	Scientific Workflow Management systems	Scale up and down resources according to runtime requirements with proper usage and handle cloud services	Requires substantial domain knowledge
`Interoperability	All	Support universal standards to access services	Continuous change of standards

Table 3: Analysis Framework for Architectural Concerns

other's hypotheses [22]. Provenance information on data derived and the derivation procedure is essential in order to achieve reproducibility in scientific computations. Provenance data includes important

information on preserving data, determining the data's quality and authorship and reproducing results [25]. In a highly collaborative environment, where many scientists are involved, provenance data is lost, fragmented in e-mails, wikis, journal references etc. [22]. Workflow management systems must capture and generate provenance information as a part of the workflow-generated data. Nevertheless it is critical to share provenance information systematically and explicitly.

6.8 Interoperability

Issues of service and data discovery, composition, communication among end users and distributed services are critical in the application domain. Therefore, adhering to well-known and preferably open standards at the interface levels will minimize restrictions, access issues and additional re-implementation in the future. For instance OGC has provided specific interface descriptions for web services. Such a solution should allow scientists to communicate with service instances, connect to corresponding distributed services and directly invoke available OGC-based services via the OGC standard specifications [21]. Hence it is worth ensuring to serve perceived needs of expert users while maintaining common service accessing principles.

7.0 ANALYSIS FRAMEWORK OF SYNTHESIZED ARCHITECTURAL PERSPECTIVES

Table 3 synthesizes the software architectural concerns that were discussed in the previous section. This synthesis illustrates the capabilities and limitations of enabling technologies for computational geoscience from an architectural perspective.

Table 3 presents a comparative categorization of the currently available state-of-the-art tools and applications in the computational geoscience domain. This provides an overview of their capabilities and support towards addressing the architectural concerns relevant in computational geoscience.

This synthesized architectural perspective, along with the challenges discussed in section 5, provide a sound analysis framework for the design and architecture of a geoscience gateway as a solution that would effectively enable computational geoscience.

8.0 CONCLUSION

In this paper we analyzed the domain of computational geoscience investigating deeply into current limitations and identified the architectural concerns in implementing geoscience related applications. The synthesis of these challenges and concerns is presented as a framework that can be leveraged to customize requirements and to make design decisions in implementing geoscience research applications.

This framework is not limited to application developers. It can also be used by anyone who wishes to choose the best geoscience application based on their requirements. The framework presented considers a set of pre-identified features which substantially cover the characteristics of an ideal geoscience gateway.

The paper mainly focuses on OGC standards as geoscience standards. However, in a general geoscience gateway, consideration should be given towards uniform adoption of standards, protocols and service interfaces by other international initiatives as well.

This framework can be expanded through further research to cover more features that would be relevant for either a subdomain-specific or a generic computational geoscience gateway. As future research work, we intend to create a reference implementation of a generic computational geoscience gateway based on this analysis framework.

9.0 REFERENCES

[1] "What is Geoscience?," *The Department of Geosciences*. [Online]. Available: http://www3.geosc.psu.edu/prospective/whatis.php. [Accessed: 10-Apr-2013].

[2] P. Zhao, L. Di, and G. Yu, "Building asynchronous geospatial processing workflows with web services," *Comput. Geosci.*, vol. 39, pp. 34–41, 2012.

[3] Xiaoyan Li, Liping Di, Weiguo Han, Peisheng Zhao, and Upendra Dadi, "Sharing geoscience algorithms in a Web service-oriented environment (GRASS GIS example)," *Comput. Geosci.*, vol. 36, no. 8, pp. 1060–1068, Aug. 2010.

[4] M. Zapatero, R. Castro, G. Wainer, and M. Houssein, "Architecture for integrated modeling, simulation and visualization of environmental systems using GIS and cell-devs," in *Proceedings of the Winter Simulation Conference*, 2011, pp. 997–1009.

[5] "Apache Airavata." [Online]. Available: http://airavata.apache.org/. [Accessed: 17-Apr-2013].

[6] "Taverna Official Site." [Online]. Available: http://www.taverna.org.uk/. [Accessed: 02-Apr-2013].

[7] "Triana." [Online]. Available: http://www.trianacode.org/index.html. [Accessed: 17-Apr-2013].

[8] "Kepler." [Online]. Available: https://kepler-project.org/. [Accessed: 17-Apr-2013].

[9] "Pegasus." [Online]. Available: http://pegasus.isi.edu/. [Accessed: 17-Apr-2013].

[10] Ustun Yildiz, Adnene Guabtni, and Anne H. H. Ngu, "Business versus Scientific Workflow A Comparative study," in in *Proceedings of the 2009 Congress on Services - I (SERVICES '09)*, 2009, pp. 340–343.

[11] S. Pandey, D. Karunamoorthy, and R. Buyya, "Workflow engine for clouds," *Cloud Comput. Princ. Paradig. R Buyya J Broberg Goscinski Eds Isbn-13*, pp. 978–0470887998, 2011.

[12] J. Chen and W. M. P. van der Aalst, "On scientific workflow," *Ieee Tech. Comm. Scalable Comput. Newsl.*, vol. 9, no. 1, 2007.

[13] I. Altintas, J. Wang, D. Crawl, and W. Li, "Challenges and approaches for distributed workflow-driven analysis of large-scale biological data: vision paper," in *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, 2012, pp. 73–78.

[14] Preston Smith, "Northwest Indiana_Computational Grid." [Online]. Available: https://indico.fnal.gov/getFile.py/access?contribId=28...10... [Accessed: 26-Apr-2013].

[15] "About OGC." [Online]. Available: http://www.opengeospatial.org/ogc. [Accessed: 21-Apr-2013].

[16] "OGC standards," *OGC standards*. [Online]. Available: http://www.opengeospatial.org/standards/is. [Accessed: 15-Apr-2013].

[17] "FAQs - OGC's Purpose and Structure." [Online]. Available: http://www.opengeospatial.org/ogc/faq#14. [Accessed: 26-Apr-2013].

[18] Tamas Kukla, Tamas Kiss, Gabor Terstyanszky, Peter Kacsuk, and Gergely Sipos, "Enabling the execution of various workflows (Kepler, Taverna, Triana, P-GRADE) on EGEE."

[19] James Ahrens, Bruce Hendrickson, Gabrielle Long, Steve Miller, Rob Ross, and

Dean Williams, "Data-Intensive Science in the US DOE: Case Studies and Future Challenges," *Comput. Sci. Eng.*, vol. 13, no. 6, pp. 14–24, Nov. 2011.

[20] T. M. Rhyne, A. MacEachren, and T.-M. Rhyne, "Visualizing geospatial data," in *ACM SIGGRAPH 2004 Course Notes*, 2004, p. 31.

[21] C. Granell, L. Díaz, and M. Gould, "Service-oriented applications for environmental models: Reusable geospatial services," *Environ. Model. Softw.*, vol. 25, no. 2, pp. 182–198, Feb. 2010.

[22] Yolanda Gil, Ewa Deelman, Mark Ellisman, Thomas Fahringer, Geoffrey Fox, Dennis Gannon, Carole Goble, Miron Livny, Luc Moreau, and Jim Myers, "Examining the Challenges of Scientific Workflows," *Computer*, pp. 24–32, 2007.

[23] R. Barbera, G. La Rocca, R. Rotondo, A. Falzone, P. Maggi, and N. Venuti, "Conjugating Science Gateways and Grid Portals into e-Collaboration environments: the Liferay and GENIUS/EnginFrame use case," in *Proceedings of the 2010 TeraGrid Conference*, 2010, p. 1.

[24] J. Basney, R. Dooley, J. Gaynor, S. Marru, and M. Pierce, "Distributed web security for science gateways," in *Proceedings of the 2011 ACM workshop on Gateway computing environments*, 2011, pp. 13–20.

[25] S. B. Davidson and J. Freire, "Provenance and scientific workflows: challenges and opportunities," in in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, 2008, pp. 1345–1350.