Editorial Introduction: Special Issue on Grids and Geospatial Information Systems

Marlon E. Pierce
Community Grids Laboratory
Indiana University
mpierce@cs.indiana.edu

Abstract: Grids and Geospatial Information Systems (GIS) are both based on distributed service architectures and have complementary capabilities. GIS systems provide a comprehensive set of services for managing maps, geospatial data sets, and geospatial information that can be applied to areas ranging from access to scientific data by researchers to disaster planning and emergency management. The data and information focus of GIS is being augmented with the computational and virtual organization capabilities of Grid computing by many projects, including the contributors to this special issue. This editorial introduction serves as an overview of the issues discussed at the GIS-Grid Workshop in the Open Grid Forum and the follow-on papers of this special issue.

Introduction

Grids [1] and Geospatial Information Systems (GIS) follow very similar distributed, service-based computing architectures, and their open, community-led approach to standards and governance should enable mutually beneficial collaborations. Many projects in the Grid community have recognized this and have adopted GIS technologies in their cyberinfrastructure. Recognizing this common approach, members of this community organized a workshop to present their work. This special issue is an outgrowth of the GIS-Grid Workshop at Open Grid Forum 15, October 3-6 2005 held in Boston, Massachusetts [2]. Presentations from the workshop were as follows:

1. *GEON: Ashraf Memon, San Diego Supercomputing Center.* The focus of this presentation was the development of distributed service infrastructure and science portals to support online mapping of semantically described geological information.
2. *LAITS: Wenli Yang, George Mason University.* This presentation focused on the integration of GIS and Globus services to provide access to NASA satellite data sets.
3. *LEAD: Beth Plale, Indiana University.* This presentation discussed the integration of real-time weather data sources with computational models and Web portals.
4. *SERVOGrid: Marlon Pierce, Indiana University.* The focus of this presentation was the development and application of GIS services, which were integrated with computational methods for earthquake modeling and forecasting.
5. *GISolve: Shaowen Wang, University of Iowa.* This presentation focused on the high performance computing aspects of GIS and how these can be integrated with Grid job management services.

This follow-on special issue includes extended papers by the LAITS, SERVOGrid (now known as QuakeSim), and GISolve teams. In addition, the Southeastern Coastal Ocean Observing and Prediction (SCOOP) team at Louisiana State University has contributed a paper on weather and storm forecasting infrastructure that combines GIS data and information systems with Grid-style computational modeling. As can be seen, and as we discuss more below, these papers provide a reasonably comprehensive overview of the

potential interactions between Grids and GIS services and more generally collaborations between these communities.

Parallels of GIS and Grids

Open Geospatial Consortium Services

In the introduction, we claimed the existence of obvious parallels between GIS and Grid systems, as well as the mutual benefits for integrating the two. We will expand upon these assertions in this section.

Like Grids, modern GIS systems are typically built as service-oriented distributed systems. Although there are many proprietary commercial offerings in this area (most notably from the vendor ESRI [3]), the community also has a strong open standard activity, the Open Geospatial Consortium (OGC) [4]. The OGC defines an abstract (or reference) service model architecture and several practical standards. The most prominent of these include the following:

- The Web Feature Service (WFS): this service provides access to XML encoded information about geospatial features. Information can range from locations and drawing instructions (vector data) to non-visual metadata.

- The Web Map Service (WMS): this service renders XML encoded features into maps using various encoding standards (SVG, JPEG, PNG, etc).

- The Web Coverage Service (WCS): this service provides access to raster data, which could be both images and binary-encoded observational data. Data can be both regularly and irregularly arrayed.

- The Web Catalog Service (CWS): this service provides an information and metadata directory for other OGC services.

In addition to their primary service interfaces (for returning maps, features, data, etc), OGC services also have metadata and capability query interfaces that allow invoking agents to learn more about the specific data sets and capabilities of a given service installation. These also enable virtualization: for example a Web Map Server can aggregate capabilities of other map servers and act as a proxy server through a process sometimes referred to as cascading. Further information on these services is available from [4].

These services are unified by the use of the very extensive Geographic Markup Language (GML) as an underlying data model. GML is a suite of XML specifications that covers a range of relevant topics, such as how to express mapping primitives (lines, points, and polygons), how to express capabilities of services, how to express observations and measured values, how to express abstract (non-visual) information about map features, and so on. The papers by Di *et al* and Aydin *et al* in this special issue provide more information about these standards.

The OGC has also defined a new set of standards for sensor networks, sometimes collectively referred to as SensorML [5]. These specifications are not specifically addressed in this special issue, but sensors are particularly important to storm and earthquake modeling.

The OGC standards in their original form predate Web Service standards (SOAP and WSDL), and in current terminology would be known as Representational State Transfer (REST [6])-style services: they use URLs and HTTP GET/POST. Current versions of the standards can support SOAP messages. Work to align OGC services more closely with Web Services is described by Aydin *et al.*

OGC, Web, and Grid Services

One of the pressures on the OGC has been to determine when to define standards and when to adopt existing, more generalized standards from other communities. The Open Grid Forum has faced similar questions as the Grid community has moved to Web Services [7]. Common examples of generalized services that can be potentially adopted or adapted by the GIS community include the following:

- *Information services:* These are services for finding other services. The Web Catalog Service is specialized to GIS. The obvious general-purpose candidate here is UDDI [8], although WS-Context is another possibility. The use of Globus's Metadata Directory Service (MDS) is another possibility. The papers by Di *et al* and Aydin *et al* examine these requirements.

- *Processing/execution services:* The OGC has put forward the Web Processing Service specification for managing geo-processing on the Web. However, this style of service is a hallmark of Grid services, with the pre-Web Service GRAM and Web service-based WS-GRAM service designed to provide access to computational resources, particularly supercomputers with batch schedulers [9]. The Condor scheduling system (which includes a Web service interface, Birdbath) is another popular Grid execution environment [10]. The papers by Aydin *et al* and Allen *et al* examine the integration of GIS data services with execution services for computational modeling. The paper by Wang *et al* considers this issue from the other perspective: the authors perform computationally intensive geospatial calculations (such as clustering) and need to leverage the advanced, cross-system computational facilities of Grids.

- *Workflow execution services:* Workflow, or service orchestration, is the combination of atomic, general-purpose services into specialized, composite services suitable for a specific task. The Business Process Exchange Language (BPEL) is the workflow standard. Workflow engines from the Grid community include Taverna, Triana, and Kepler. Workflows for Grids are reviewed in [11]. The papers by Allen *et al* and Aydin *et al* provide use cases for combined GIS and Grid workflows.

- *Data movement services:* Data movement services are specialized for the transfer of non-trivial data sets across networks. As both Di *et al* and Aydin *et al* point out, implementations of the OGC standards are not particularly suited to this. GridFTP [12] is a common Grid standard for high performance data transfer. Bittorrent is commonly used by the general Web community.

- *Virtual Organizations, Portals and Gateways:* Grid systems are typically accompanied by Web portals, sometimes called Science Gateways [13]. The OGC Web Map Service specification is explicitly for delivering imagery and is often the basis for user interface components. The papers by Di *et al*, Allen *et al*, and Aydin *et al* discuss issues in providing user interfaces for communities of users.

As we have indicated above, the various papers in this special issue address these concerns. We believe these are important enhancements that the Grid community can bring to GIS.

Future Directions for Grids and GIS

There is current interest by members of the Open Grid Forum and the OGC in collaborative ventures. See for example the proceedings of OGF 23 (http://www.ogf.org/gf/event_schedule/index.php?id=1232). The further coupling of GIS, sensor web, and Grids is inevitable and desirable.

Going beyond immediate tactical issues of this integration, we see several larger issues that need to be considered by both Grids and GIS. Grids, Web services, and GIS standards are all being pressured by the twin concepts of "cloud" computing and Web 2.0. Both of these terms are descriptive, rather than prescriptive, of general trends in distributed computing, so we will provide examples.

Cloud computing is best thought of as providing a service interface for controlling virtual computing images (creating, destroying, modifying, etc). These images can be used in turn as ordinary computing hosts. Amazon's Elastic Cloud Computing (EC2) Service and Simple Storage Service (S3) are popular examples. These compete most directly with Grid infrastructure providers rather than Grid middleware (one can use a virtual image as a host for a Web Map Service, for example). We note also that cloud systems, which build on virtualization technologies like Xen, are well suited for exploiting multicore systems. Grids have not ignored cloud computing: the Workspaces [14] and EUCALYPTUS [15] projects are two examples of cloud-style middleware from the Grid community. However, large-scale deployments of this middleware to support scientific computing are still to come.

Web 2.0 provides a more direct challenge to Web and Grid service standards and software. Appropriate to its importance and every-day relevance, OGC services received the first direct challenge from Web 2.0 in the form of Google Maps and Google Earth. Google Maps is notable for its relatively simple JavaScript programming API, built over the top of JSON and AJAX messaging techniques and using remote REST services. The data model for both Google Maps and Google Earth is the Keyhole Markup Language (KML), which is far simpler and easier to work with than GML. There has been some notable reconciliation, as Google has donated KML to the OGC, and Google Maps natively supports the OGC standard GeoRSS. More information on KML and GeoRSS is available from [4].

As has been discussed in [16], this is in fact a general challenge to all Web services and cyberinfrastructure. Several GIS-based Grid applications already provide Google Map and other Web 2.0-style interactivity in their science portal interfaces. However, this is an early effort, and GIS portals must find ways, for instance, to integrate themselves with the mash-up composers of Web 2.0. Workflow composers from the Grid community are one obvious tool for building mash-ups. The influence of social networks on Grids (which support a user-driven alternative to Grids' Virtual Organizations) will also be important.

In any case, the hallmark of Web 2.0 is its support for the "do it yourself" approach to information technology, with relatively low entry barriers for new developers. Web services, Grids and the OGC (in this author's opinion) have been guilty of developing excessively complicated specifications and standards that require specialized knowledge and training (rather than, say, general programming experience) to use and extend. Rather

than continue these trends, it is time for a reevaluation of technical approaches by these communities as a whole.

References

1. Ian Foster and Carl Kesselman, eds, "The Grid: Blueprint for a New Computing Infrastructure," 2nd Edition, Morgan Kaufmann, 2004. ISBN: 1-55860-933-4.
2. "Building Geographical Information System," OGF 15, October 3-6, 2005 Gridshttp://www.crisisgrid.org/GISWorkshop/.
3. ESRI Web Site: http://www.esri.com
4. Open Geospatial Consortium Web Site: http://www.opengeospatial.org/
5. Introduction to SensorML http://vast.uah.edu/index.php?option=com_content&view=article&id=14&Itemid=52/
6. Roy T. Fielding, Richard N. Taylor: Principled design of the modern Web architecture. ACM Trans. Internet Techn. 2(2): 115-150 (2002).
7. Malcolm P. Atkinson, David De Roure, Alistair N. Dunlop, Geoffrey Fox, Peter Henderson, Anthony J. G. Hey, Norman W. Paton, Steven Newhouse, Savas Parastatidis, Anne E. Trefethen, Paul Watson, Jim Webber: Web Service Grids: an evolutionary approach. Concurrency - Practice and Experience 17(2-4): 377-389 (2005) .
8. UDDI Online Community for the Universal Description, Discovery and Integration Standard: http://uddi.xml.org/
9. J. Nabrzyski, J.M. Schopf, J. Weglarz (Eds). Grid Resource Management. Kluwer Publishing, Fall 2003.
10. Douglas Thain, Todd Tannenbaum, Miron Livny: Distributed computing in practice: the Condor experience. Concurrency - Practice and Experience 17(2-4): 323-356 (2005) .
11. Geoffrey Charles Fox, Dennis Gannon: Special Issue: Workflow in Grid Systems. Concurrency and Computation: Practice and Experience 18(10): 1009-1019 (2006).
12. Globus Toolkit Support for Distributed Data-Intensive Science. W. Allcock, A. Chervenak, I. Foster, L. Pearlman, V. Welch, M. Wilde. Proceedings of Computing in High Energy Physics (CHEP '01), September 2001.
13. Nancy Wilkins-Diehr: Special Issue: Science Gateways - Common Community Interfaces to Grid Resources. Concurrency and Computation: Practice and Experience 19(6): 743-749 (2007).
14. Katarzyna Keahey, Ian T. Foster, Timothy Freeman, Xuehai Zhang, Daniel Galron: Virtual Workspaces in the Grid. Euro-Par 2005: 421-431.
15. The EUCALYPTUS project web site: http://eucalyptus.cs.ucsb.edu/
16. Marlon E. Pierce, Geoffrey Fox, Huapeng Yuan, and Yu Deng, "Cyberinfrastructure and Web 2.0" in High Performance Computing and Grids in Action (L. Grandinetti Editor) published by IOS Press, Amsterdam, as the volume no 16 in the series "Advances in Parallel Computing"., Proceedings of HPC2006 July 4 2006 Cetraro Italy.