

# ADAPTING THE DEVELOPMENT MODEL OF THE GRID ANATOMY TO MEET THE NEEDS OF VARIOUS APPLICATION DOMAINS

Soha Maad, Brian Coghlan, Gabrielle Pierantoni, Eamonn Kenny, John Ryan, Ronan Watson  
Trinity College Dublin, Ireland  
{maads, coghlan, pierantg, ekenny, john.p.ryan, watsonr}@cs.tcd.ie

## Abstract

The grid is penetrating various application domains. Two implications arise: the grid is reshaping our programming models for various application domains; and the development and deployment of grid middleware is reshaped by the needs of various application domains. Whereas the implementation of the grid anatomy [1, 2] succeeded so far, to some extent, in realizing the Virtual Organization (VO) concept concerning the security, the workload management, and the discovery and monitoring of grid resources [3, 4], work remains to be done in adapting the implementation of the grid anatomy to meet the needs of various application domains. The key issue to be addressed is the software engineering practice adopted in grid middleware development. This paper surveys the penetration of the grid in nine application domains and discusses relevant issues related to the software engineering model for grid middleware development.

## 1 Introduction

Various grid projects have been launched so far. The focus of these projects varies depending on the deployed technology and the target application domains. These domains are classified as compute and data intensive, knowledge based, or service based [1, 5]. This paper surveys the penetration of the grid in nine application domains and discusses grid middleware from a software engineering perspective. The paper is divided into four sections including this one. The second section presents an overview of the grid penetration in nine application domains. Based on this survey, the third section highlights the impact of the application domain on the development of grid middleware and discusses grid middleware development from a software engineering perspective. The paper concludes with a summary.

## 2 The Grid Penetration in Various Application Domains

This section surveys the grid penetration in nine application domains: physics, medicine, astronomy, environment, engineering, media, chemistry, finance, and government. This survey highlights, for each application domain: (a) the challenges faced; (b) the grid solution; and (c) the limitations of the grid solution.

### 2.1 Physics

The major *challenge* facing the physics domain is conducting experiments that need to process tera-byte scales of data. Examples of these experiments are the Large Hadron collider LHC [6], A Large Ion Collider Experiment ALICE [7] and ATLAS [8]. The *grid solution* drove the development of the EGEE middleware [9]. The major *challenge facing the grid solution* is meeting the deadline of the start date for conducting the planned physics experiments. The major *limitation facing the grid solution* may well turn out to be the non-determinism of the grid.

### 2.2 Medicine

Three distinct *challenges* are facing the field of medicine: (a) producing interactive medical simulations (e.g. heart simulation [10, 11]); (b) analyzing and managing medical images [12]; and (c) supporting virtual collaboration in e-Hospitals [13] (a virtual network of hospitals that delivers medical training, e-surgery, and medical analysis services). The *grid solutions* to address the above three challenges consist of: (a) the development of various special-purpose middleware (e.g. innerGrid Nitya [14] developed by GridSystems, and CrossGrid [15]); (b) the development of resource brokers for the discovery of suitable clusters for the execution of parallel image reconstruction algorithms [12]; and (c) the proposal of a metagrid engine that provides a superset of functionalities across different Grid Engines with privacy and quality of service QoS management [13]. Acknowledged *limitations of the grid solutions* on offer are the lack of capabilities needed for communication and knowledge sharing (especially for e-Hospital applications).

### 2.3 Astronomy

The major *challenge* facing the field of astronomy is the analysis of tera-bytes of astronomical image data generated by telescopes. Moreover, astronomical image capturing devices can generate several images, each of hundreds of Mbytes, per single shot [16]. This necessitates: data intensive computation; scalable file I/O in the order of GB/s; replica management; and parallel / distributed processing of files. As a *grid solution*, the GFARM Grid file system [17] offers a special purpose Grid

middleware for data intensive computation. A major *limitation* is the need for manual intervention to move the peta-byte scale archive data to the GFARM system. Currently, there is no direct connection between the GFARM system and the telescope devices.

## **2.4 Environment**

For environment applications a major *challenge* is the parallel execution of hundreds of programs corresponding to large scale air pollution, nuclear waste storage, pollution and climate models. Various *grid solutions* have been offered. For a particular large scale air pollution model, a national VO was formed [18] involving researchers and resources from 7 UK institutions. At an international level, a similar VO was formed involving 21 EU institutions within the CrossGrid [15] consortium. The major *limitations* were the lack of effective workflow tools and the immaturity of interactivity tools.

## **2.5 Engineering**

A major *challenge* facing engineering work is the virtual collaboration on the design, production and maintenance of products that are described in complex structured product model databases [19, 20, 21]. Existing web service paradigms can weakly support such collaboration. The potential *grid solution* lies in an appropriate implementation of the VO concept. The latter provides a robust framework for engineering collaboration due to its flexibility, adaptability, and security. An acknowledged *limitation of the grid solution* [19, 20] is the lack of awareness of the business concepts. *The future research agendas* involve: (a) the development of generic business-object-aware extensions to grid middleware. This involves grid enabling present applications on existing grids using toolkits, and extending the grid architecture with semantics and ontologies beyond current work on metadata and heterogeneous data formats; (b) making typical server-side applications (or components of applications) grid-computing compatible; and (c) enabling the mostly client-side Computer Aided Design (CAD) applications to interface with the grid.

## **2.6 Media**

The major *challenge* facing media applications is the production, broadcasting, delivery and playout of interactive media content (audio, video, image) in real time. To address this challenge various *grid solutions* were developed. These include, amongst others: the Grid Visualization Kernel (GVK) [22] which allows the visualization pipeline [23] (data enrichment/reduction, followed by mapping of the data to an abstract form, and finally composing the visual image) to be ported to grid resources, and also handles the communication between the simulation generating the

data and the visualization of the simulated data; and G-Vid [24] which is based on GVK which allow the production of real time interactive MPEG4 compliant video content on the grid. *Limitations of the grid solutions* are manifested in the difficulty to split out the media production pipeline and distribute it to the grid resources and to guarantee reliable and secure interaction with the pipeline. The grid has succeeded in providing batch processing, scheduling, data storage, and transfer services, but providing reliable and secure interaction with the grid is not yet mature. Various efforts are currently underway to enhance interactivity with the grid [25, 26, 39].

## 2.7 Finance

The major *challenges* facing the financial industry are: (a) solving real-world large scale investment problems (a sub area of financial engineering); (b) producing more realistic financial modeling based on the processing of large financial datasets; and (c) complying with the regulatory constraints imposed by the financial industry authority (Basle II), which is forcing large scale financial companies to create their own grid applications. Various *grid solutions* were offered including, among others, the implementation of the Open Grid Service Environment [27] that provides an abstract service stack used to model large scale computational financial problems as abstract workflows. *Limitations of the grid solutions* are that they have yet to converge to a web service implementation.

## 2.8 Chemistry

A major *challenge* facing the field of chemistry is molecular design and engineering [28]. This typically involves the implementation of the QSAR/QSPR methodology which involves 3D structure generation, semi-empirical calculations, descriptor calculation, and model building. A *grid solution* involved the development of OpenMolGRID [29] system that provides the necessary infrastructure for molecular design and engineering. OpenMolGRID is based on UNICORE grid middleware.

## 2.9 E-government

In e-government, the major *challenge* is the development of a an e-government infrastructure that supports the shift to a service-oriented e-government model [30]. The VO concept and the Web Service Resource Framework architecture of the grid are potentially viable *grid solutions*. This is basically due to the flexibility of the VO concept and the added security and state control of the Web Service Resource Framework. However a *major limitation of the grid solution* is the ability to develop appropriate VO models capturing all forms of government interaction [31] including: Government-to-Government (G2G) interaction which involves public agencies

interacting with each other in order to implement processes; Government-to-Citizen (G2C) interaction which involves interaction between citizens and public agencies; and Government-to-Business (G2B) interaction which is the interface between companies and public agencies. Another *limitation* is the inability of grid data management solutions in handling the semantic content of the huge amount of government data.

### **3 Grid middleware from a software engineering perspective**

Based on the above overview of the grid penetration in various application domains, we notice that grid middleware is being extensively and continuously reshaped by the needs of various application domains. This takes the form of developing special purpose compute and data intensive grid middleware (physics and astronomy), introducing new metagrid engines (medicine), developing resource brokers (medicine), developing grid service APIs (environment), raising grid awareness of business concepts (collaborative engineering), developing new middleware kernels (media), adding new middleware services (finance), developing grid application toolkits (chemistry), and advising semantic-aware solutions (e-government). It is notable how many of the problems relate to user interactions, data I/O and workflow. The continuous demand for change in existing grid middleware can be attributed to the adoption of either a *bottom-up* approach or a *top-down* approach for grid middleware development.

#### **3.1 Bottom-up approach**

In a bottom-up approach for software development [32] individual parts of the system are specified in detail and the parts are then composed to form larger components, which are in turn composed until a complete system is formed. This approach has been utilised in most of the prevalent grid middleware. For instance, the Globus Alliance “Ecosystem” of Grid Components [33] framed the exercise of building a Grid system or application as a software integration problem, hopefully leveraging existing grid components to reduce the development cost.

#### **3.2 Top-down approach**

In a top-down approach for software development [32] a top level view of an application domain is mapped to an appropriate infrastructure. An example of a top down approach for grid middleware development is WebCom-G [34, 35]. In WebComG, applications are specified as Condensed Graphs in a manner which is independent of the execution architecture, thus separating the application and execution environments. Another example of a top down approach is K-Wf Grid [36, 37], the Knowledge-based Workflow system for grid applications that addresses the

need for a better infrastructure for the future grid environment. K-Wf Grid aims at assisting users in composing powerful workflows using a rule-based expert system.

### 3.3 Adapting the development model

Our survey of the grid penetration in various application domains reveals that the cost of developing grid applications using a bottom-up approach is considerable. By adopting a bottom-up approach, most of grid middleware development has evolved in isolation of application domain needs. On the other hand a top-down approach for grid middleware development may lead to narrow solutions and tend to be less generic. In addressing this dilemma a new software development model is needed that adapts the implementation of the grid anatomy to meet the needs of various application domains.

In some way this software development model must either marry or separate the *top-down* approach and the *bottom-up* approach for middleware development. Given the predominance of issues relating to user interactions, data I/O, and workflow, there may be a potential in migrating these functionalities from the traditional bottom-up grid middleware into an independent metagrid environment. Whether they then could be amendable to top-down development is as yet uncertain, but at least they could be dealt with independently.

## 5 Conclusion

This paper surveyed the grid penetration in nine application domains. Based on this survey, the paper discussed the impact of the application domain on grid middleware development. An overview of the software engineering practice adopted in prevalent grid middleware concludes that a new software engineering model is needed for grid middleware development that is more adapted to the needs of application domains.

## References

1. Foster I., Kesselman C. and Tuecke S. 2001. 'The Anatomy of the Grid: Enabling Scalable Virtual Organizations'. International J. Supercomputer Applications, 15(3), 2001.
2. Foster I., Kesselman C., Nick J. and Tuecke S. 2002. 'The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration'. Open Grid Service Infrastructure WG, Global Grid Forum, June 22, 2002. (extended version of Grid Services for Distributed System Integration) [www.globus.org/research/papers/ogsa.pdf](http://www.globus.org/research/papers/ogsa.pdf).
3. Czajkowski K., Ferguson F. D., Foster I., Frey J., Graham S., Sedukhin I., Snelling D., Tuecke S. and Vambenepe W. 2004. 'The WS-Resource Framework', Version 1.0, March, 2004. <http://www.oasis-open.org/committees/download.php/6796/ws-wsrf.pdf>.
4. Globus Toolkit. <http://www.globus.org/toolkit/>

5. Grid application and deployment projects, <http://www-fp.mcs.anl.gov/~foster/grid-projects/>
6. LHC Computing Grid, <http://lcg.web.cern.ch/LCG/>
7. ALICE grid, <http://www.alice-grid.to.infn.it/>
8. ATLAS grid, <http://atlas.web.cern.ch/Atlas/GROUPS/SOFTWARE/OO/grid/>
9. Enabling Grids For E-science, <http://www.eu-egee.org/>
10. Sloot P., Novel methods and techniques for modeling, simulation and grid-computing applied to vascular reconstruction, Invited lecture, The International Conference on Functional Imaging and Modeling of the Heart 2005 (FIMH 2005), Barcelona, Spain, June 2005.
11. Ordas S., Van Assen H.C., Boisrobert L., Laucelli M., Puente J., Lelieveldt B.P.F., Frangi A. F. 2005, Statistical Modeling and Segmentation in Cardiac MRI using a Grid Computing Approach, In proceeding of European Grid Conference, 2005, The Netherlands.
12. Guarracino M., Lacetti G., Murli A., Application Oriented Brokering in Medical Imaging: Algorithms and Software Architecture, In proceeding of European Grid Conference, 2005, The Netherlands
13. Graszew G., Roelofs T., Rakowsky S., Schlag P., Albayrak S., and Kaiser S. 2005. "Telemedical applications and Grid technology", In proceeding of European Grid Conference, 2005, The Netherlands
14. InnerGrid Nitya, <http://www.gridsystems.com/products/innergrid.html>
15. CROSSGRID, <http://www.crossgrid.org/>
16. Yamamoto N., Tatebe O., and Sekiguchi S. 2004. "Parallel and Distributed Astronomical Data Analysis on Grid Datafarm", Proceedings of 5th IEEE/ACM International Workshop on Grid Computing (Grid 2004), 461-466.
17. Tatebe O., Soda N., Morita Y., Matsuoka S., Sekiguchi S. 2004. "Gfarm v2: A Grid file system that supports high-performance distributed and parallel data computing," Proceedings of the 2004 Computing in High Energy and Nuclear Physics (CHEP04), Interlaken, Switzerland, September 2004.
18. Thandavan A., Sahin C., Alexandrov V. N., Experiences with the Globus Toolkit on AIX and deploying the Large Scale Air Pollution Modelling application as a grid service 27, Proceeding Cracow'04 Grid Workshop, Crakow, Poland.
19. Ž Turk, M. Dolenc, J. Nabrzyski, P. Katranuschkov, E. Balaton, R. Balder, M. Hannus (2004) Towards Engineering on the Grid, in Dikbas A. and Scherer R. (eds.) Proceedings, ECPPM Conference, Istanbul, 2004.
20. Dolenc M., Turk Z., Stankovski V. 2004, InteliGrid Project: A Vision of Engineering on the Grid, Proceeding Cracow'04 Grid Workshop, Crakow, Poland, p.19.
21. InteliGrid, <http://www.inteligrid.com/index2.htm>
22. Kranzlmüller D., Heinzlreiter P. and Volkert J. 2003. "Grid-Enabled Visualization with GVK", In 2003 Annual Crossgrid Project Workshop & 1st European Across Grids Conference.
23. Haber R and McNabb D. 1990. "Visualization Idioms: A Conceptual Model for Scientific Visualization Systems", In G.M. Nielson, B. Shriver, and L.J. Rosenblum, (Eds.): Visualization in Scientific Computing, IEEE Computer Society Press, Los Alamitos, NM, USA, pp. 74-93 (1990).
24. Polak M, Kranzlmüller D, and Volkert J. 2004. 'G-Vid - A Dynamic Grid Videoservice for Advanced Visualization'. Proceeding Cracow'04 Grid Workshop, Crakow, Poland, 471-477.
25. Benedyczak K., Nowinski A., Nowinski K.S., and Bala P.. 2004. 'Interactive visualization using the UNICORE grid middleware'. Proceeding Cracow'04 Grid Workshop, Crakow, Poland, 478.

26. Rosmanith H., Kranzlmuller D., glogin - A Multifunctional, Interactive Tunnel into the Grid, Fifth IEEE/ACM International Workshop on Grid Computing (GRID'04), pp. 266-272
27. Hochreiter R., Wiesinger C., Giczi D. 2005, Large Scale Computational Finance Applications on the Open Grid Service Environment, In proceeding of European Grid Conference, 2005, The Netherlands
28. Schuller B., Romberg M., Kirtchakova L. 2005, Application Driven Grid Developments in the OpenMolGRID project, In proceeding of European Grid Conference, 2005, The Netherlands
29. OpenMolGRID, <http://www.openmolgrid.org/>
30. Maad S., Coghlan B., John R., Eamonn K., Watson R., and Pierantoni G. 2005. 'The Horizon of the Grid For E-Government'. Proceeding eGovernment'05 Workshop, Brunel, United Kingdom.
31. SAGA 2003. 'Standards and Architectures for e-Government Applications Version 1.1'. KBSt Publication Series, ISSN 0179-7263 Volume 56 February 2003.
32. Lethbridge T.C., and Laganière R., Object Oriented Software Engineering: Practical Software Development using UML and Java, published by McGraw Hill, Summer 2001.
33. Globus-web, Globus Alliance, An Ecosystem of Grid Components, [http://www.globus.org/grid\\_software/ecology.php](http://www.globus.org/grid_software/ecology.php)
34. Morrison J., Coghlan B., Shearer A., Foley S., Power D., Perrot R. 2005, WebCom-G: A Candidate Middleware for Grid-Ireland
35. Morrison J., Power D., and Kennedy J. 2002. Load balancing and fault tolerance in a condensed graphs based metacomputer. The Journal of Internet Technologies. Special Issue on Web based Programming, Vol. 3(4), PP. 221-234, December, 2002.
36. Gubala T., Bubak M., Malawski M., Rycerz K., Abstract Workflow Composition in K-WfGrid Project Environment, Proceeding Cracow'04 Grid Workshop, Cracow, Poland.
37. K-WF Grid, <http://www.kwfgird.net/main.asp>
38. Pierantoni, G. and Lyttleton, O. and O'Callaghan, D. and Quigley, G. and Kenny, E. and Coghlan, B., Multi-Grid and Multi-VO Job Submission based on a Unified Computational Model, Cracow Grid Workshop (CGW'05)Cracow, Poland, November 2005
39. EU FP6, [http://www.cordis.lu/fp6/fp6\\_glance.htm](http://www.cordis.lu/fp6/fp6_glance.htm)