Information Services for Smart Decision Making: An eSI Event Theme Prototype Final Evaluation

e-Science Institute

Jennifer M. Schopf,
National e-Science Centre and Argonne National Laboratory

26-01-2006

Abstract:

This article reports on a series of events, public lectures, and publications organised around the topic of Grid information services and their use in decision making. These elements constituted a prototype theme in the area of information services, led by Dr Jennifer M. Schopf as part of her visitor program agenda. Part of the goal of this prototype was to help define the structural aspects of a theme, as well as to encourage research in the area.

This theme confirmed that the basic expected functionality in this area is finally being met by the various information service tools. However, it also revealed many cases of mismatches in expectations between application scientists and tool developers. A clearer analysis of user requirements is needed to help the tool builders in their plans.

The main focus of this theme became tools and how they are used to serve information to the community. Several ongoing collaborations emerged from this work. Furthermore, the theme strongly enabled research, opening up many new questions and encouraging interactions between several disparate groups.

UK e-Science Technical Report Series

Report UKeS-2006-04
Available from http://www.nesc.ac.uk/technical_papers/UKeS-2006-04.pdf

1. Theme Overview

Grid computing resources and services can advertise a large amount of data for many different use cases. These include providing data so that resource brokers can locate computing elements appropriate for a job, streaming data so application steering adjustments can be made during runtime, and notifying system administrators when changes in system load or disk space availability occur in order to identify possible performance anomalies.

We proposed to organize a series of events, public lectures, and publications around the topic of Grid information services and their use in decision making. Specifically, this involved interacting with people building monitoring systems and trying to better understand how they can be used with schedulers and by operations and system administrators, and to better understand performance faults in the context of Grid applications.

2. About the Theme Leader

Jennifer M. Schopf is a Scientist at the Distributed Systems Laboratory, part of the Mathematics and Computer Science Division at Argonne National Laboratory. During the theme, Schopf spent the year as a half-time researcher at the e-Science Institute (eSI) in Edinburgh, UK. She is a member of the Globus Alliance, and technology coordinator for the Monitoring and Discovery Service, and spent 5 years helping to establish the Global Grid Forum standards body. She is currently also employed by the UK National e-Science Centre (NeSC) and JISC as the e-Infrastructure Policy Advisor.

In addition to her work with the Globus MDS, in the past ten years Schopf has co-edited a book and co-authored over 30 journal papers, book chapters, and refereed conference submissions, in large part related to information services, predictions, and their use in scheduling in Grid environments. She previously organized two workshops on the topic of Grid performance, and has received multiple grants in support of this work.

Schopf received a BA in Computer Science and Mathematics from Vassar College, and MS and PhD degrees from the University of California, San Diego in Computer Science and Engineering.

3. Prototype Theme

This theme, unlike the ones that will follow, grew out of discussions in the planning phase for the new eSI program. An evaluation of past events was done to determine what possible themes had already, informally, been done. Looking at the past 12 months and the next 5 months of planned events (approximately 100 meetings), certain common topics were present, including:

- Data management

- Biomedical work
- Ontologies
- Networks/communications
- Information services

It was decided that a prototype theme could be done in the area of information services, to be led by Schopf as part of her visitor program agenda. Part of the goal of this prototype was to help define the structural aspects of a theme, as well as to encourage research in the area.

4. Theme Attributes

Themes are meant to consist of a series of events, publications, and visitors. Initially, there were many possible ways to pursue the range of items that could be included. This section discusses the items that were part of the prototype theme and the coverage metrics used to evaluate it.

4.1 Events

The set of events related to this theme included:

- eSI lecture: "Distributed Monitoring and Information Services for the Grid", Jennifer Schopf, Jan 10, 2005 (19 participants)
- Grid Performance Workshop, June 22-23, 2005 (35 participants)
- Globus Toolkit GridFTP Days, January 27, 2005 (20 participants, limited by room size)
- GLUE Schema 1.2 Workshop (RAL, Didcot, Oxford), February 2005
- UK Globus Week, April 04-08, 2005 (51 participants)
- Networks for non-Networkers 2, June 20-21, 2005 (associated, not directed by TL, 68 participants)

In addition, the research leader gave a set of invited talks and visits, within and beyond the UK, including:

- Invited Seminar, Cambridge e-Science Centre, Cambridge University, December 6, 2005.
- Invited Seminar, University of Bath, September 30, 2005
- Visit, Rutherford Appleton Laboratory (January 2005)
- Visit, Warwick University (May 2005, Dec 2005)
- Visit, Newcastle University (June 2005)
- Invited talk: "Grid Monitoring and Information Services: Globus Toolkit MDS4 & TeraGrid Inca", LCG Operations Workshop, CERN November 2-4 2004
- Invited talk: "Globus Toolkit Monitoring and Discovery System: MDS4", Technology Review Talk, Argonne Booth, SuperComputing Nov 8-11, 2004,
- Invited talk: "Performance Inside: Performance Monitoring and Diagnosis for NMI Software and Applications", NMI Meeting December 2004

- Invited talk: “Distributed Monitoring and Information Services for the Grid”, MSc in High Performance Computing Seminar, EPCC Edinburgh, January 28, 2004
- Invited talk: “Distributed Monitoring and Information Services for the Grid”, Master Class, University of Amsterdam, April 2005

The outreach component to this theme is larger than we expect other themes to have, partly because of the nature of the work already underway by the research leader in the visiting program. Several events that were initially planned did not come to fruition due to time constraints. These include a workshop examining current scheduling issues (originally planned by the research leader), and several more theoretical workshops by other associated researchers that were cancelled.

4.2 Publications

Several publications resulted from this work, including

- “Monitoring and Discovery in a Web Services Framework: Functionality and Performance of Globus Toolkit MDS4”, Jennifer M. Schopf, Ioan Raicu, Laura Pearlman, Neill Miller, Carl Kesselman, Ian Foster, Mike D’Arcy, *submitted to HPDC 2006*, available as MCS Preprint ANL/MCS-P1315-0106, January 2006.
- “OGSA-DAI Status and Benchmarks”, Mario Antonioletti, Malcolm Atkinson, Rob Baxter, Andrew Borley, Neil P. Chue Hong, Patrick Dantressangle, Alastair C. Hume, Mike Jackson, Kostas Karasavvas, Amy Krause, Simon Laws, Mark Parsons, Norman W. Paton, Jennifer M. Schopf, Tom Sugden, Kostas Tourlas, Paul Watson, and David Vyvyan, Proceedings of the UK eScience All Hands Meeting 2005, September 2005.
- "Monitoring and Discovery in a Web Services Framework: Functionality and Performance of the Globus Toolkit's MDS4", Jennifer M. Schopf, Mike D'Arcy, Neill Miller, Laura Pearlman, Ian Foster, and Carl Kesselman, Argonne National Laboratory Technical Report ANL/MCS-P1248-0405, 2005.
- “Report of the UK Globus Week 2005”, ed. J.M. Schopf, ANL MCS Technical Memorandum ANL/MCS-TM-291, National eScience Centre Technical Report UKeS-2005-06, October 2005, available from http://www.nesc.ac.uk/technical_papers/UKeS-2005-06.pdf.
- “Report of the International Grid Performance Workshop 2005”, ed. J.M. Schopf, Argonne National Laboratory MCS Technical Memorandum ANL/MCS-TM-288.
- “Grid Performance Workshop 2004 Meeting Report”, ed. J.M. Schopf, Mathematics and Computer Science Division Technical Memorandum ANL/MCS-TM-285, Argonne National Laboratory, National eScience Centre Technical Report UKeS-2005-05, August 2005, available from http://www.nesc.ac.uk/technical_papers/UKeS-2005-05.pdf.

In addition, a glossy handout and several posters were developed, for eSI, NeSC, SuperComputing, and the All Hands Meeting.

A special issue on the topic of information services for decision making in e-Science was considered, and the editors of the Journal of Grid Computing expressed an interest in such a proposal, but this did not occur due to time constraints.

4.3 Visitors

One aspect of the current eSI themes that was not taken advantage of was inviting additional visitors to take part in the theme for a period of time outside of just attending meetings. This was mainly due to the fact that the theme was a prototype, and this was not available in a timely enough manner.

4.4 Coverage Metrics

Six metrics have been defined to understand the scope of a given theme:

Participants can be experts to new researchers in a field. This theme had a full spread of participants, but was weighted towards the experts in terms of sheer numbers.

Sector coverage can vary from university to industrial. This theme was evenly covered across the spectrum.

The research area has been defined on a scale of “arts and humanities” to “science”. This theme was weighted towards science.

The aspect metric measures where the work sits in the range from practical to theoretical. This work was heavily weighted towards the pragmatic.

Research issues concerns whether this work is of immediate need or long-term importance. Much of the work in this theme was oriented towards the present and near future.

Engagement for a theme can be local/national to international. This work was international, but weighted towards UK and American, with some EU but very little Asian participation.

5. Evaluation Summary

Because this theme was a prototype, the summary has two aspects to it – structural suggestions and technical results.

5.1 Structural Suggestions

Because this theme was started partway through the allotted time period, and because much of the infrastructure was defined during the course of the theme, it was not able to take full advantage of the infrastructure now in place. The biggest component this theme lacked was additional visitors, which would have strengthened the work significantly. It

is highly recommended that, as part of any theme, a core set of visitors should be defined and invited before starting.

The other component that was not in place was the PR and Web infrastructure. There was no mailing list or formal announcement strategy, and the publicly available materials about the theme lagged significantly behind the main events. This did not hurt any individual event, as all were well attended and received strong comments. However, it did affect the cohesiveness of the theme as a whole.

A year is a very short amount of time to run a set of coordinated events. In addition, there is a tension between running a good event and spending time working on larger research issues, as the more preparation that is possible prior to the event, the better the outcome. That said, having the research leader working on the theme half time and pursuing their own research half time was a logical use of resources.

5.2 Technical Summary

This theme has brought to light several findings in the e-Science Information Services community. Discussions with users attending the events confirmed that the basic expected functionality in this area is finally being met by the various information service tools. However, over the year we saw in many instances that there was a mismatch in expectations between application scientists and tool developers. Application developers often want simple tools for basic problems that work reliably, while tool builders are often funded to supply complex solutions to novel problems instead. A clearer analysis of user requirements is also needed – many of the short term and long term goals for specific users could be clarified to help the tool builders in their plans.

This theme wound up being much more focused on tools and how they were used to serve information to the community. Several ongoing collaborations emerged from this work. Furthermore, while the theme itself was not research per se, it enabled research quite strongly. This work opened up many new questions and encouraged interactions between several disparate groups, which is the strongest metric of success overall.