

# e-ScienceTalk

## GRIDGUIDE UPGRADED INTEGRATION WITH THE RTM

### EU DELIVERABLE: D2.1

---

Document identifier:	e- ScienceTalk_D2_1_GridGuide_Integratio n_With_RTM
Date:	<b>31/08/2011</b>
Work package:	<b>WP2</b>
Lead Partner:	<b>Imperial</b>
Document Status:	<b>FINAL</b>
Dissemination Level:	<b>PUBLIC</b>
Document Link:	<a href="https://documents.egi.eu/document/766">https://documents.egi.eu/document/766</a>

---

#### Abstract

The GridGuide and Real Time Monitor have been upgraded during the course of the first year of e-ScienceTalk and new versions launched to meet D2.1 GridGuide Upgraded Integration with the RTM. This document describes the work carried out on the upgraded versions during PY1 as part of this Deliverable.

## COPYRIGHT NOTICE

Copyright © Members of the e-ScienceTalk collaboration, 2010-2013. See [www.e-sciencetalk.eu](http://www.e-sciencetalk.eu) for details of the e-ScienceTalk project and the collaboration. E-ScienceTalk is a project co-funded by the European Commission as an Support Action within the 7th Framework Programme. E-ScienceTalk began in September 2010 and will run for 33 months. This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/3.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, and USA. The work must be attributed by attaching the following reference to the copied elements: "Copyright © Members of the e-ScienceTalk Collaboration, 2010-2013. See [www.e-sciencetalk.eu](http://www.e-sciencetalk.eu) for details of the e-ScienceTalk project and the collaboration". Using this document in a way and/or for purposes not foreseen in the licence, requires the prior written permission of the copyright holders. The information contained in this document represents the views of the copyright holders as of the date such views are published.

## I. DELIVERY SLIP

	Name	Partner/Activity	Date
From	J. Martyniak	Imperial/WP2	4/08/2011
Reviewed by	Moderator: Reviewers:	Various	29/08/2011
Approved by	PMB:		29/08/2011

## II. DOCUMENT LOG

Issue	Date	Comment	Author/Partner
1	29/08/2011	First draft	J Martyniak & N O'Neill / Imperial & QMUL
2			
3			

## III. APPLICATION AREA

This document is a formal deliverable for the European Commission, applicable to all members of the e-ScienceTalk project and its beneficiaries and collaborating projects.

## IV. DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the authors.

## V. PROJECT SUMMARY

Over the last 10 years, the European Commission and governments have invested substantial funds in distributed computing infrastructures. Scientists have access to state-of-the-art computational and data resources located around the world, putting European research into a leading position to address the greatest challenges facing us today, such as climate change, pandemics and sustainable energy. The advent of the European Grid Infrastructure, combined with the blurring of boundaries between grids, clouds, supercomputing networks and volunteer grids, means that a clear consistent source of information aimed at non-experts is now more important than ever, through dissemination projects such as e-ScienceTalk, that cross national boundaries.

### Objectives:

- e-ScienceTalk will build on the achievements of the GridTalk project in bringing the success stories of Europe's e-Infrastructure to policy makers in government and business, to the scientific community and to the general public.
- e-ScienceTalk will work with EGI-InSPIRE and other collaborating projects to expand the scope of the existing GridTalk outputs, and to report on the interactions of grids with e-Infrastructures such as cloud computing and supercomputing.
- The project will explore options for the sustainability of e-ScienceTalk's products.
- e-ScienceTalk will produce a series of reports aimed at policy makers to disseminate key policy issues underpinning grid and e-Infrastructure development in Europe. The project will also coordinate e-concertation activities.
- The GridCafé, GridCast and GridGuide suite of websites will cover new topics and explore novel web technologies; they will integrate closely with GridPP's Real Time Monitor, combining live views of grid activity with the human aspects of computing.
- The growing weekly publication, International Science Grid This Week (iSGTW) will bring news and events to the existing and potential e-Science community under a new name of The Digital Scientist.

## VI. EXECUTIVE SUMMARY

The GridGuide and Real Time Monitor have been upgraded during the course of the first year of e-ScienceTalk and new versions have been launched as D2.1 *GridGuide Upgraded Integration with the RTM*. This document describes the work achieved during PY1 as part of this Deliverable.

The RTM is a real time visualisation of activity on the grid computing infrastructure. The latest version utilises the NASA World Wind virtual globe, which is based on OpenGL and Java. The RTM overlays the movement of site activity and job transfers onto the 3D globe, giving users the ability to see the current state of the grid infrastructure. Since becoming a part of the e-ScienceTalk project the RTM development has focused on four areas; the website, maintenance, user support and extending the application's functionality. This has included input from other areas of the project to help give a fresh look to the application. In PY1, 64 countries are included in the RTM and the team has visited 16 events where the RTM has been demonstrated.

GridGuide ([www.gridguide.org](http://www.gridguide.org)) is the youngest of the e-ScienceTalk products and gives a human face to the grid, showing the sites and sights of grid computing. Users can listen to podcasts from grid sites worldwide, read about the ongoing work and watch interviews with researchers. As well as giving a visual overview of current grid work, GridGuide enables users to drill down to more detail about an individual scientist's work and how the grid has produced results. There are currently 38 sites on the GridGuide, including 28 EU sites and 11 non-EU in the US and the Asia-Pacific region. Of these, 34 are also currently included in the Real Time Monitor.

During PY1, the e-ScienceTalk team has worked on extending the RTM website to make it easier to navigate and to improve the design. For the RTM developers, a lot of the work on the non-grid parts are looked after by the WorldWind community. However, the grid technology can lead to issues for the RTM as grid users start using different ways of interacting with the infrastructure or sharing data. Contact with users has been improved, and the RTM is now easier to use by non-experts.

By the start of the e-ScienceTalk project the RTM was running on the latest Worldwind SDK. During the move to the new SDK the project had also moved to using Java 1.6. This now offers the functionality of calling the computer's default browser. This allowed the developers to include a button that could redirect the user to the GridGuide through their preferred browser. This has now been implemented and over 30 sites on the RTM now have a GridGuide entry.

Jobs from the LHC experiments, including ATLAS, represent more than 90% of the traffic on the grid, and it is important to display the movement of data associated with these experiments. Due to ATLAS's move to PANDA from the WMS/L&B combinations service provided by gLite, the RTM team was unable to get access to the PANDA system to query it directly and so needed to go via the CERN dashboard, which monitors and publishes data from all the LHC experiments. This required the RTM to be extended. This work has led to some important improvements to the RTM, including increased integration with the CERN dashboard. PANDA and "normal" jobs are displayed on separate animation layers of the RTM and can be turned on and off easily by the user. As a result, there is now a better workflow for understanding what needs to be done as more and more communities join the grid and possibly bypass the gLite tools.

## TABLE OF CONTENTS

<b>1</b>	<b>INTRODUCTION</b> .....	<b>6</b>
1.1	The Real Time Monitor.....	6
1.2	The GridGuide.....	7
<b>2</b>	<b>PROGRESS IN YEAR ONE</b> .....	<b>8</b>
2.1	Website.....	8
2.2	Maintenance.....	8
2.3	User Support.....	8
2.4	Extending the RTM Functionality .....	9
2.4.1	GridGuide Integration .....	9
2.4.2	Support for new data sources .....	10
<b>3</b>	<b>CONCLUSION</b> .....	<b>13</b>
<b>4</b>	<b>REFERENCES</b> .....	<b>14</b>

## 1 INTRODUCTION

The GridGuide and Real Time Monitor have been upgraded during the course of the first year of e-ScienceTalk and new versions launched as D2.1 *GridGuide Upgraded Integration with the RTM*. The current version of the GridGuide is available at [www.gridguide.org](http://www.gridguide.org) and the Real Time Monitor can be downloaded as a standalone application from <http://rtm.hep.ph.ic.ac.uk/>, or launched as a Java webstart version. This document describes the work achieved during PY1 as part of this Deliverable.

### 1.1 The Real Time Monitor

The RTM is a real time visualisation of activity on the grid computing infrastructure. The High Energy Physics e-Science group at Imperial College London has been developing it since 2002. Initially funded by the UK-based GridPP collaboration, work on the RTM became a part of the e-ScienceTalk project in September 2010. The latest version utilises the NASA World Wind virtual globe, which is based on OpenGL and Java. The RTM overlays the movement of site activity and job transfers onto the 3D globe, giving users the ability to see the current state of the grid infrastructure. It is modularly designed, making it easy to add and change various aspects of the application depending on the demands placed on it.

The application also includes more detailed information about the status of the individual sites on the grid including:

- Current number of jobs running/queued at a site
- The load on a site's computing elements and workload management systems
- Graphs detailing the work done at the site over various time periods
- Information from the GridGuide about that site, if available

Running the RTM is also relatively straightforward, with no special installation required, on computers with a modern up-to-date operating system. This has widened the use beyond a small community of experts to anyone wanting to look at or discuss the grid infrastructure.

The tools on which the RTM is based are used for more than generating the live visualisations. They also generate important statistics about sites, jobs and virtual organisations, which can be used to help in diagnosing problems or monitoring performance of the infrastructure [R1,2,3,4].

Since becoming a part of the e-ScienceTalk project the RTM development has focused on four areas: the website, maintenance, user support and extending the application's functionality. This has included input from other areas of the project to help give a fresh look to the application.

In PY1, 64 countries are included in the RTM and the team has visited 16 events where the RTM has been demonstrated. These included the EGI Technical and User Forums in Amsterdam and Vilnius, and ICT2010 in Brussels, which attracted 10,000 delegates. Researchers at the University of Birmingham and at Cambridge University asked for the demo to be presented at the Royal Society Summer Science 2011 exhibition.



## 1.2 The GridGuide

GridGuide ([www.gridguide.org](http://www.gridguide.org)) is the youngest of the e-ScienceTalk products and gives a human face to the grid, showing the sites and sights of grid computing. Users can listen to podcasts from grid sites worldwide, read about the ongoing work and watch interviews with researchers. As well as giving a visual overview of current grid work, GridGuide enables users to drill down to more detail about an individual scientist's work and how the grid has produced results. For these reasons, the GridGuide is useful for engaging with policy makers who are able to find out more detail about work going on in their local regions or areas of responsibility, as well as the general public and other scientists.

The GridGuide complements the GridCafé by providing a more in-depth guide to institutions across the globe that are involved in grids and distributed computing. GridGuide has become increasingly interactive and accessible through co-development with the Real Time Monitor (RTM), which shows traffic on the worldwide grid in real time. The current integration with GridGuide allows a visitor to click on a site and view both the technical statistics from the RTM as well as the pages from GridGuide. The RTM is widely used for demonstrating the grid at conferences and events and is an accessible and engaging way to understand more about the grid.

Feedback on GridGuide was gathered through email and one-to-one interviews with science communicators and scientists. The GridGuide map has been cited as a good way of demonstrating what the grid does, and provides a draw for people researching grids and e-Infrastructure in their own locality. E-ScienceTalk has focused its efforts to encourage more interest in the GridGuide site. One suggestion would be to provide the information in different formats and to cross-promote by including a link from the GridCafé website. Reviewers have also suggested creating a short video clip showing a few minutes of grid activity on the RTM for marketing and demo purposes. This could include a human aspect with a young scientist uploading data to the grid.

There are currently 38 sites on the GridGuide, including 28 EU sites and 11 non-EU in the Americas, Africa and the Asia-Pacific region. Of these, 34 are also currently included in the Real Time Monitor.

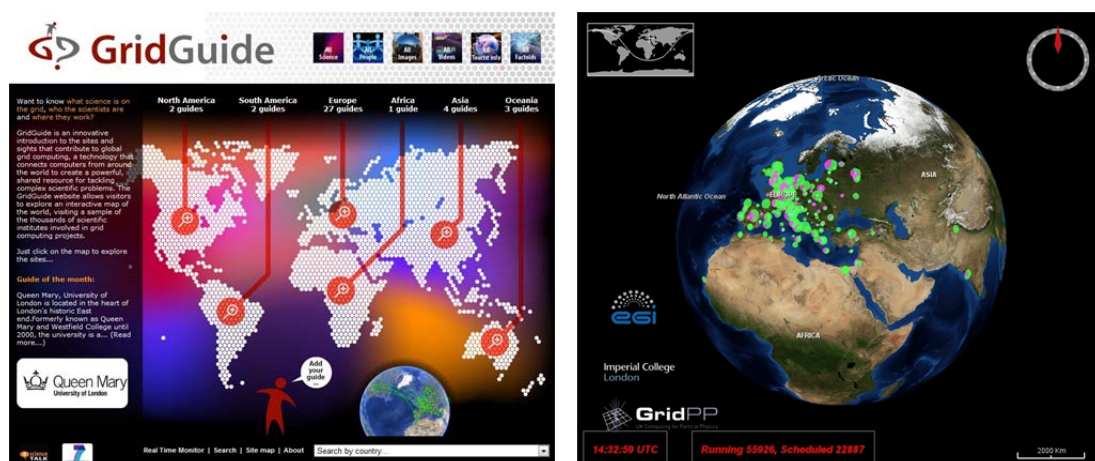


Fig. 1: Screenshot of the GridGuide (left) and Real Time Monitor (right)

## 2 PROGRESS IN YEAR ONE

### 2.1 Website

In preparation for its move to an e-ScienceTalk funded project, the RTM website (<http://rtm.hep.ph.ic.ac.uk>) was updated by the GridPP dissemination team. Since then, the e-ScienceTalk team has worked on extending the site to make it easier to navigate and to improve the design. The major improvements have been:

- Making the “run” button easier to find and more prominent
- Including detailed information for all major operating systems (Windows, MacOSX and Linux)
- A section for site admins detailing how to enable their site on the RTM
- Solutions to frequently encountered problems
- Links to the raw information used by the RTM

The RTM team will continue to work on the design and content of the website based on comments from users and other team members and changes to the RTM application. It will also change to reflect the new logo designed by the team at APO Design, leader of WP2.

### 2.2 Maintenance

The RTM relies on slowly but constantly evolving technologies. These include the maturing grid technology, the Java the RTM is based on, updates to the WorldWind application and changes to the operating systems it can be run on. For the RTM developers a lot of the work on the non-grid parts are looked after by the WorldWind community and just requires keeping abreast of developments there. This required a major overhaul of the underlying software that runs the RTM just prior to the start of e-ScienceTalk.

However the grid technology presents a slightly different issue, as the grid responds to the needs of its users, not those of the RTM. This can lead to issues for the RTM as grid users start using different ways of interacting with the infrastructure or sharing data (see Section 2.4.1 *Extending the RTM functionality - Grid Guide integration*). The RTM team is involved in supporting grid users in various disciplines so is aware of the changing demands of the users and the tools being offered by the grid. This allows them to know about and understand any changes that may affect the RTM.

### 2.3 User Support

Currently the RTM has a wide range of users, with varying levels of computer literacy. This is in stark contrast to early versions, which were used by a small number of technical users. Since the move to the new WorldWind version the application became easier to run on a wide range of computers



without any specialist knowledge. To help these new users the RTM website was overhauled (see above) and the team at Imperial College set up a mailing list to inform users of new developments.

Users may also contact the team through a dedicated email address to ask for new features, highlight bugs or ask for support. This has led to new and improved functionality as well as revealing a few flaws and bugs. For example one user who was running the RTM for days at a time discovered a serious bug that was easily fixed but would only be apparent in such a situation.

## **2.4 Extending the RTM Functionality**

### **2.4.1 GridGuide Integration**

The RTM uses recognised standard data formats to automatically collect information about grid resources registered on the European grid's BDI<sup>1</sup> (Berkeley Database Information Index) database. Before moving to the e-ScienceTalk project this had been beneficial, allowing non-European Grid Infrastructure communities who do not use this central system (like EUMEDGRID, for example), to include their resources on the RTM. This also means that while the RTM undergoes cosmetic changes or extends its functionality the key underlying data collection has remained consistent throughout the life of the project.

During the GridTalk project which preceded e-ScienceTalk, the idea of integrating the GridGuide into the RTM was discussed with Imperial College London and GridTalk. The idea was to extend the use of both products while not losing the core functionality of either. This resulted in a solution with simplified, custom pages being displayed through a Java browser called Lobo within the RTM itself. However during a major update of the RTM through an update of the Worldwind software this became unworkable and the Lobo browser was dropped prior to the start of the e-ScienceTalk project.

By the start of the e-ScienceTalk project in September 2010, the RTM was running on the latest Worldwind SDK and the team could focus on re-integrating the GridGuide. It was decided that this should:

- Allow RTM users to click a link within the RTM to see the GridGuide information for a particular site
- Display the full GridGuide site, not a cut down version for the RTM
- Use a standard browser installed on the computer and not require a browser to be included in the RTM application which could lead to security problems or unsustainable code

During the move to the new SDK the project had also moved to using Java 1.6. This now offered the functionality of calling the computer's default browser. This allowed the developers to include a button that could redirect the user to the GridGuide through their preferred browser. This has now been implemented and over 30 sites on the RTM now have a GridGuide entry.

---

<sup>1</sup> <https://twiki.cern.ch/twiki/bin/view/EGEE/BDII>

## 2.4.2 Support for new data sources

The RTM was designed to use the data produced by the standard gLite middleware. In recent years some experiments have started using alternative tools. In particular those who do not use the WMS/L&B combinations service provided by gLite could not be monitored by the RTM as their job information was not being submitted to the database that the RTM queries. This highlighted the need for the RTM to be able to interact with new data sources and this has been the bulk of the work undertaken by the team at Imperial College London during PY1.

The first major grid user to “disappear” from the RTM was the ATLAS experiment when they moved to their custom workload management system PANDA. Computing tasks, or “jobs” from the four main LHC experiments, which include ATLAS, represent more than 90% of the traffic on the grid, and it is important to display the movement of data associated with these experiments. Due to the move to PANDA, the RTM team was unable to get access to the PANDA system to query it directly and so needed to go via the CERN dashboard, which monitors and publishes data from all the LHC experiments. This required the RTM to be extended by adding two new components:

1. The Enquirer. A program running on the RTM server querying the dashboard for PANDA jobs and writing out digested information in XML for to the storage area served by Apache Web Server
2. A new module in the RTM client that can read the digested information, analyse it and display it

This work has led to some important improvements to the RTM:

1. Increased integration with the CERN dashboard. This will make it easier to add new data sources that use the dashboard or a similar framework. The team are currently working on implementing data transfers by the CMS experiment thorough a similar module
2. PANDA and “normal” jobs are displayed on separate animation layers of the RTM and can be turned on and off easily by the user
3. Better workflow for understanding what needs to be done as more and more communities join the grid and possibly bypass the gLite tools.

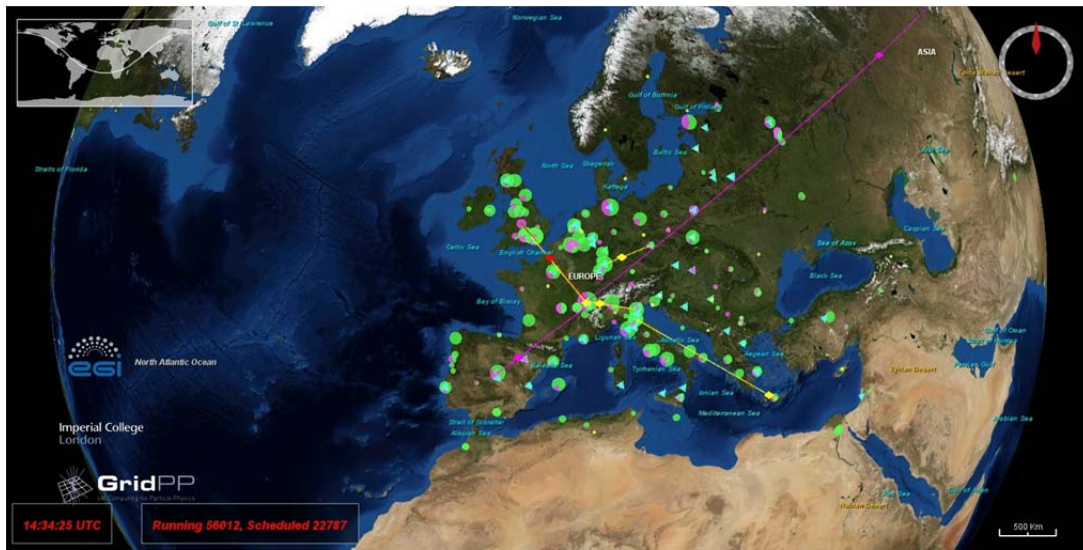


Fig. 2: Close up view of the RTM in operation

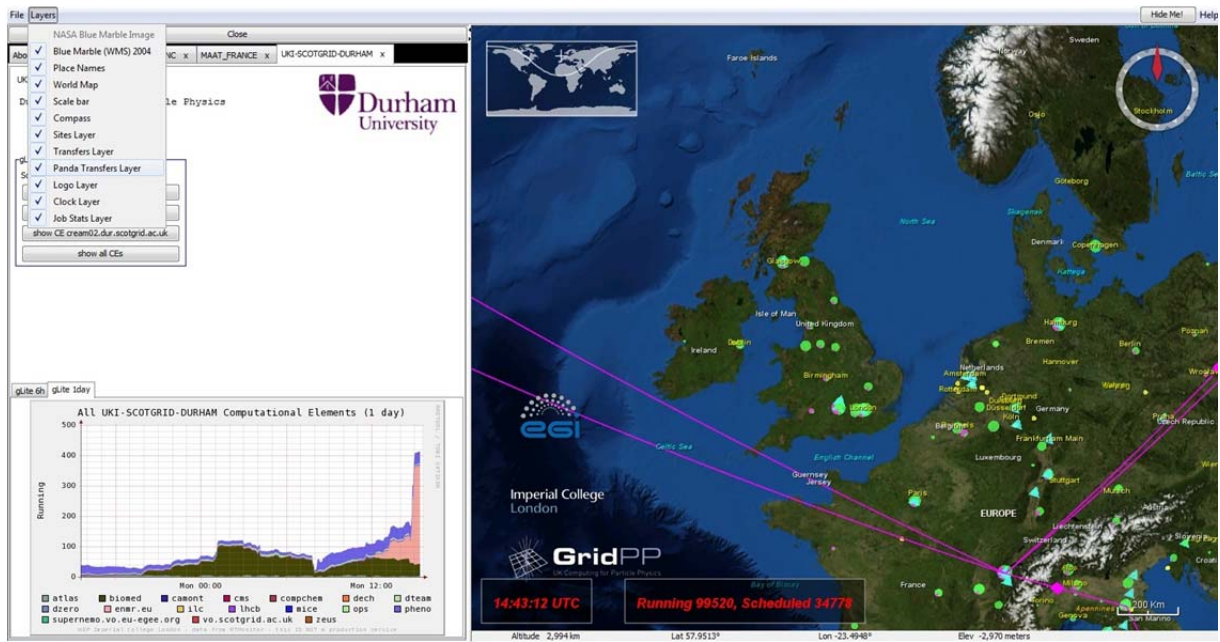


Fig 3. View of the bleeding edge webstart version of the RTM with Panda transfer layer

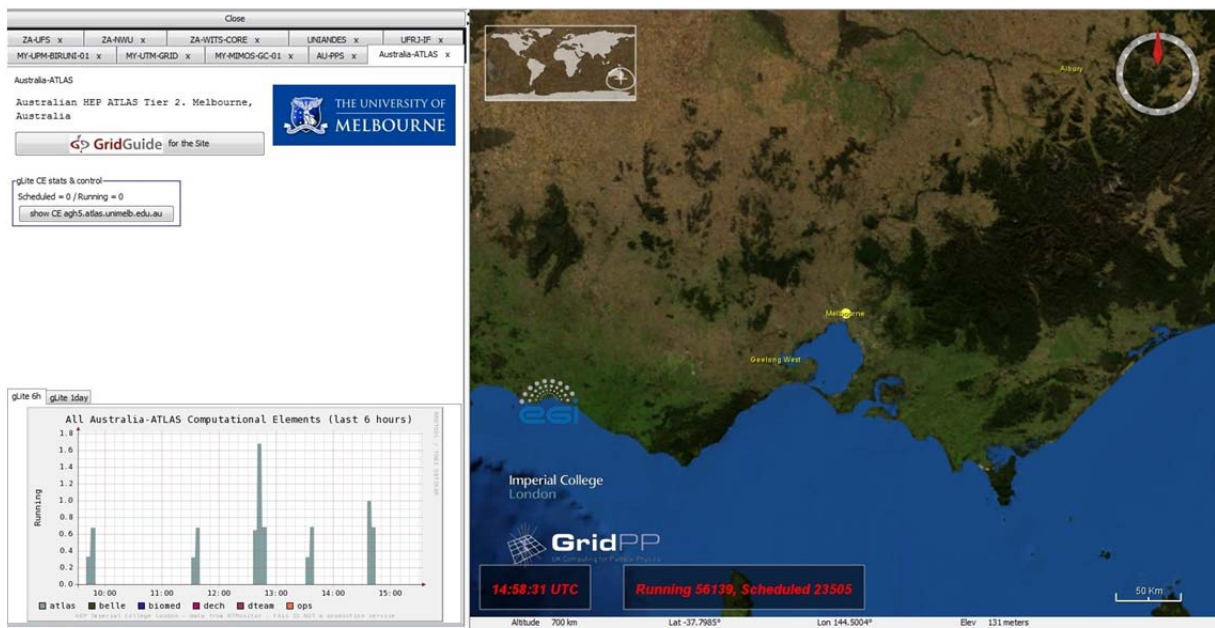


Fig. 4: GuideGuide integrated site guide for University of Melbourne



### 3 CONCLUSION

The work carried out on the GridGuide and RTM integration in PY1 has led to a substantial new source of grid jobs to be displayed on the RTM, the PANDA jobs from ATLAS. The challenge for PY2 will be to increase the visibility of the GridGuide, which is the youngest and least visited e-ScienceTalk website. Currently, the main metrics for GridGuide and RTM are based on usage (i.e. page views and unique visitors), content (i.e. number of sites) and dissemination (i.e. number of demos). Finding out why people visit, or do not visit, the websites could be done by surveying online users. In PY2, it would also be worthwhile examining who the main referrers are and how visitors arrive at the site. Unfortunately, the metrics for page views and unique visitors are low, but this does not reflect 'real' numbers. The RTM has been demo-ed at a number of conferences, and therefore a significantly larger number of individuals have viewed the RTM than is perhaps represented in the figures. The RTM also continues to generate a number of papers in the areas of workflow and improvement of job submission, execution and monitoring.

## 4 REFERENCES

<b>R 1</b>	Real Time Monitor of Grid job executions - D J Colling et al 2010 J. Phys.: Conf. Ser. 219 062020 doi: 10.1088/1742-6596/219/6/062020
<b>R 2</b>	The Grid Observatory – C. Germain-Renaud et al. ccgrid, pp.114-123, 2011 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2011 doi: 10.1109/CCGrid.2011.68
<b>R 3</b>	Optimization of jobs submission on the EGEE production grid: modeling faults using workload – D. Lingrand et al. JOGC, 8 (2), pages 305–321 doi: 10.1007/s10723-010-9151-2
<b>R 4</b>	Analyzing the EGEE production grid workload application to jobs submission optimization - D. Lingrand et al. Lecture Notes in Computer Science, 2009, Volume 5798/2009, 37-58, DOI: 10.1007/978-3-642-04633-9_3