



Portal User Interfaces for Metadata Management ¹

| | |
|-------------------|------------------------------|
| Deliverable | 7.6 |
| Authors | Oliver Wehrens, Thomas Radke |
| Date | August 5, 2008 |
| Document Version | 1.0.0 |
| Current Version | 1.0.0 |
| Previous Versions | 0.2.6 |

A: Status of this Document

Officially approved document for project deliverable D7.6.

B: Reference to project plan

This deliverable refers to the task TA VII-5 "*Erstellung von Benutzerschnittstellen fuer die Metadaten- und Datenverwaltung*".

C: Abstract

This document describes the portal interfaces for metadata management as implemented for different AstroGrid-D use cases.

¹This work is part of the AstroGrid-D project and D-Grid. The project is funded by the German Federal Ministry of Education and Research (BMBF).

D: Changes History

| Version | Date | Name | Brief summary |
|----------------|-----------------|----------------|--|
| 0.1.0 | January 23 2008 | Oliver Wehrens | Working Draft Creation |
| 0.2.0 | February 5 2008 | Oliver Wehrens | Changes to style and some corrections |
| 0.2.1 | May 14 2008 | Thomas Radke | Added GEO600 portlet description |
| 0.2.2 | June 6 2008 | Thomas Radke | Added ClusterFinder portlet description |
| 0.2.3 | June 9 2008 | Thomas Radke | More details about the Cactus portlet |
| 0.2.4 | June 12 2008 | Thomas Radke | Finished portlet descriptions |
| 0.2.5 | June 13 2008 | Thomas Radke | Added summary |
| 0.2.6 | June 30 2008 | Thomas Radke | Announced as draft version |
| 1.0.0 | August 5 2008 | Thomas Radke | Publication as official AstroGrid-D Deliverable Document |

E:

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 4 |
| 2 | Metadata Management in AstroGrid-D Portlets | 5 |
| 2.1 | ClusterFinder Portlet | 5 |
| 2.2 | Cactus Simulations Portlet | 6 |
| 2.3 | GEO600 Portlet | 9 |
| 3 | Summary | 10 |
| | References | 11 |

1 Introduction

The majority of AstroGrid-D services and use cases produce and process metadata in some standardised or application-specific form: the *AstroGrid-D Data Management (ADM)* service[1] stores the locations of logical and physical files; AstroGrid-D's job monitoring service[2] keeps a list of all jobs submitted to AstroGrid-D computing resources via Globus WS-GRAM; application-specific monitoring/steering methods built into Cactus[3] generate runtime information about an ongoing parallel simulation; a script periodically collects GEO600[4] job statistics metadata. Most of the AstroGrid-D metadata is stored in RDF format as a W3C standard to uniquely describe arbitrary types of metadata. Such RDF metadata can then be generically maintained by *Stellaris*[5], an RDF metadata management service developed in AstroGrid-D, or other standard-compliant RDF frameworks.

The Stellaris software package already provides a web-based user interface which can be used to browse arbitrary RDF metadata and submit user queries in a query language called SPARQL, however such a generic visual interface can in most cases not satisfy the specific requirements needed by a particular use case or AstroGrid-D application. For practical metadata searches and the presentation of the query results a tailored user interface is often preferred by the actual astronomers and astrophysicists as end users.

Several such graphical user interfaces to Stellaris have been developed in AstroGrid-D for particular use cases:

- *AGD Resource Map* showing the location of computing resources
<http://www.gac-grid.org/project-products/grid-status/ResourceMap.html>
- *AGD Timeline* to display usage records of computers and telescopes
<http://www.gac-grid.org/project-products/grid-status/Timeline.html>

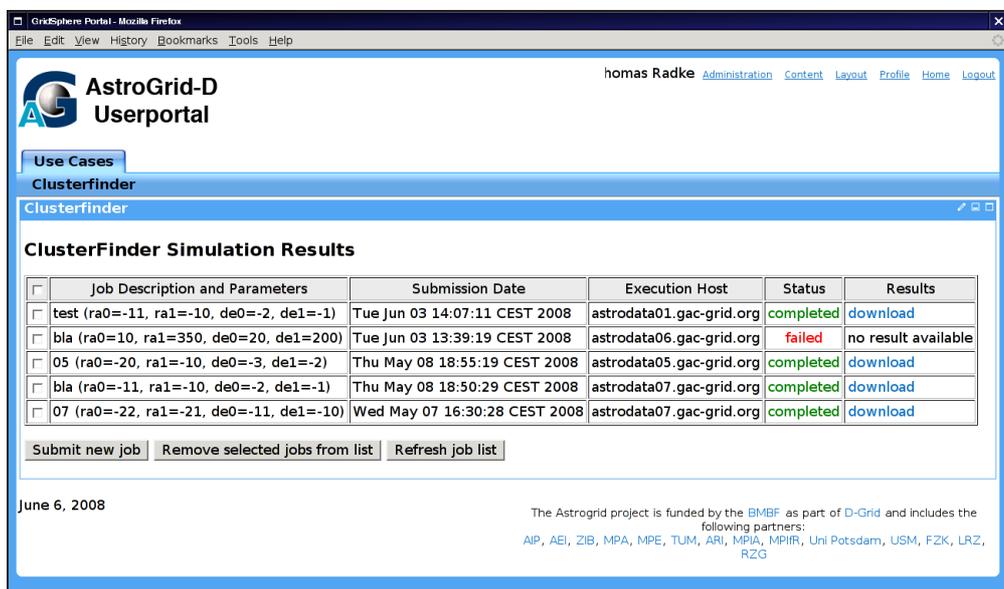
In this deliverable D7.6 we present graphical user interfaces for metadata management which have been integrated for various use cases in the AstroGrid-D user portal <http://cashmere.aip.de:8080> which is based on the GridSphere portal framework[9].

2 Metadata Management in AstroGrid-D Portlets

The management of metadata in the AstroGrid-D portal was implemented in application-specific GridSphere portlets. This section describes the commonalities and differences between these portlets with respect to how metadata are stored, queried, and retrieved.

2.1 ClusterFinder Portlet

The ClusterFinder portlet was written for the *ClusterFinder* use case[6], a cosmology simulation code. The motivation for writing a ClusterFinder portlet was that astronomers can submit ClusterFinder jobs via a standard graphical user interface from any web client, with the AstroGrid-D user portal on the server side acting as a single point of contact.



The screenshot shows the 'ClusterFinder Simulation Results' page in the AstroGrid-D Userportal. The page features a table with columns for Job Description and Parameters, Submission Date, Execution Host, Status, and Results. Below the table are buttons for 'Submit new job', 'Remove selected jobs from list', and 'Refresh job list'. The date 'June 6, 2008' is displayed at the bottom left, and funding information is at the bottom right.

| <input type="checkbox"/> | Job Description and Parameters | Submission Date | Execution Host | Status | Results |
|--------------------------|---|-------------------------------|--------------------------|-----------|--------------------------|
| <input type="checkbox"/> | test (ra0=-11, ra1=-10, de0=-2, de1=-1) | Tue Jun 03 14:07:11 CEST 2008 | astrodata01.gac-grid.org | completed | download |
| <input type="checkbox"/> | bla (ra0=10, ra1=350, de0=20, de1=200) | Tue Jun 03 13:39:19 CEST 2008 | astrodata06.gac-grid.org | failed | no result available |
| <input type="checkbox"/> | 05 (ra0=-20, ra1=-10, de0=-3, de1=-2) | Thu May 08 18:55:19 CEST 2008 | astrodata05.gac-grid.org | completed | download |
| <input type="checkbox"/> | bla (ra0=-11, ra1=-10, de0=-2, de1=-1) | Thu May 08 18:50:29 CEST 2008 | astrodata07.gac-grid.org | completed | download |
| <input type="checkbox"/> | 07 (ra0=-22, ra1=-21, de0=-11, de1=-10) | Wed May 07 16:30:28 CEST 2008 | astrodata07.gac-grid.org | completed | download |

Submit new job Remove selected jobs from list Refresh job list

June 6, 2008

The Astrogrid project is funded by the BMBF as part of D-Grid and includes the following partners:
AIP, AEI, ZIB, MPA, MPE, TUM, ARI, MPIA, MPIR, Uni Potsdam, USM, FZK, LRZ, RZG

Figure 1: *Simulations Results* page in the ClusterFinder portlet

When logged into the AstroGrid-D user portal scientists can

1. submit a ClusterFinder simulation job to a given Grid resource, with the simulation parameters conveniently specified by the user in an HTML form
2. view the list of previous ClusterFinder simulations along with their metadata (see figure 1 for an example)
 - job description and parameters
 - submission date
 - execution host
 - job status
 - simulation results
3. download the result (a datafile) generated by a successful ClusterFinder job

All metadata about previous ClusterFinder jobs, as well as the user preferences (the list of machines to be regarded for ClusterFinder job submissions, the target location for staging the simulation result datafiles) are stored by the ClusterFinder portlet as per-user properties in the common GridSphere portal database that was configured to be used by the portal administrator (by default an HSQL database, other types of external databases can be selected as well). The data format used to store metadata is naturally dependent on the database backend used. However, it is hidden to the actual portlet by a generic database access API as provided by GridSphere. All metadata store/query operations within the ClusterFinder portlet are performed through this GridSphere-intrinsic portlet container API.

2.2 Cactus Simulations Portlet

Similar to the ClusterFinder portlet, the Cactus portlet (as deployed in the Cactus portal[7, 10] and the NumRel portal[8] for the Cactus use case[3]) provides a job list to scientists showing all their Cactus simulations which are currently running or have finished already. Due to the different requirements of the two AstroGrid-D simulation use cases however, the contents of this list is rather different compared to ClusterFinder simulations though, as can be seen in an example in figure 2.

The screenshot shows the 'Cactus Metadata' portlet interface. At the top, there are navigation tabs: 'Welcome', 'Administration', and 'Cactus Metadata'. Below this is a sub-menu: 'Integration Tests', 'Simulations', 'User-defined Queries', and 'Preferences'. The main area is titled 'Simulations' and contains a search bar with filters for 'by title', 'by user', and 'by parameter file', along with a 'Show only the most recent' dropdown and a 'Query again' button. Below the search bar are two buttons: 'Clear selection' and 'Compare parameter files of selected simulations'. The main content is a table with the following columns: 'simulation', 'by user', 'parameter file path', 'running on URL', 'started TZ Europe/Berlin', and 'last updated TZ Europe/Berlin'. The table lists four simulation jobs, all titled 'BNS inspiral' and run by 'baiotti'. Each row includes a checkbox, the simulation title, the user name, the full parameter file path, the execution host and URL, the start time, and the last update time.

| simulation | by user | parameter file path | running on URL | started TZ Europe/Berlin | last updated TZ Europe/Berlin |
|---------------------------------------|---------|---|---|--|---|
| <input type="checkbox"/> BNS inspiral | baiotti | Meudon-moving-6lev-move3-45km_all-time-refinement_dtfac.35_toNS_short_bitant.par /data9/baiotti/Meudon-moving-6lev-move3-45km_allTref_NoMoveEnd_dtfac.35_finer-grids-in-origin_toNS_short_bitant | ic0164 https://peyote.aei.mpg.de:24002 | 221:51 hours ago Mar 24, 2008 4:13:50 AM | 0:13 hours ago Apr 2, 2008 10:51:27 AM |
| <input type="checkbox"/> BNS inspiral | baiotti | Meudon-moving-6lev-move3-45km_all-time-refinement_dtfac.35_larger-finetest.par /data21/baiotti/Meudon-moving-6lev-move3-45km_allTref_NoMoveEnd_dtfac.35_finer-grids-in-origin_larger-finetest-grids | node0058.admin https://belladonna.aei.mpg.de:24000 | 167:00 hours ago Mar 26, 2008 11:04:17 AM | 0:28 hours ago Apr 2, 2008 10:37:08 AM |
| <input type="checkbox"/> BNS inspiral | baiotti | Meudon-moving-6lev-move3-45km_all-time-refinement_dtfac.35_redo_glt_HR.par /data21/baiotti/Meudon-moving-6lev-move3-45km_allTref_NoMoveEnd_dtfac.35_finer-grids-in-origin_redo_glt_HR | node0064.admin https://belladonna.aei.mpg.de:24007 | 163:14 hours ago Mar 26, 2008 2:50:49 PM | 0:32 hours ago Apr 2, 2008 10:32:50 AM |
| <input type="checkbox"/> BNS inspiral | baiotti | Meudon-moving-6lev-move3-45km_all-time-refinement_dtfac.35_redo_glt.par /data21/baiotti/Meudon-moving-6lev-move3-45km_allTref_NoMoveEnd_dtfac.35_finer-grids-in-origin_redo_glt | node0048.admin https://belladonna.aei.mpg.de:24004 | 168:02 hours ago Mar 26, 2008 10:02:38 AM | 0:59 hours ago Apr 2, 2008 10:05:33 AM |

Figure 2: Screenshot of a Cactus portlet page listing the most recent Cactus simulations

A list of Cactus simulations comprises, on an overview level, the following information retrieved via specific metadata queries:

- a descriptive title for the simulation (as specified by the user at startup time)
- the login ID of the owner who ran the simulation
- the location and name of the Cactus parameter file used by this simulation
- the name of the execution host where the simulation ran, along with the connection URL of the Cactus built-in monitoring/steering interface (if available)
- the dates of the simulation startup and its last update with the AstroGrid-D metadata information service

Since the total number of simulations can be rather long, users may refine the list of simulations shown in the Cactus portlet by setting filter options to match certain metadata criteria, for instance

to display only simulations with a certain simulation title or of a certain user, or limit the overall list to the most recent N entries.

The simulations table on the Cactus portlet's overview page also presents hidden metadata in some columns as hyperlinks, directing the user to other portlet webpages which are then dynamically created from results of a more refined query in the metadata for a particular simulation:

1. a job information page with a complete list of all metadata taken at simulation startup as well as during runtime, eg. periodic announcements of the current simulation status in periodic intervals (see figure 3 as an example)
2. the full contents of the parameter file used by this simulation (see figure 4 as an example)
3. the URL of the monitoring/steering thorn if activated for this simulation

The screenshot shows a web interface for 'Cactus Metadata'. The main content area is titled 'Simulations' and contains a 'Back to list of simulations' button. Below this is a section titled 'Job Information Published at Simulation Startup' with the following details:

- Simulation: *BNS Inspiral*
- Started at: Wed Mar 26 11:04:17 CET 2008
- by User: baiotti
- Running on: node0058.admin
- # processors: 36
- Parameter file: [Meudon-moving-6lev-move3-45km_all-time-refinement_dtfac.35_larger-finest.par](#)
- Working directory: /data21/baiotti/Meudon-moving-6lev-move3-45km_allTref_NoMoveEnd_dtfac.35_finer-grids-in-origin_larger-finest-grid
- Executable: cactus_belladonna_git
fresh version 4.0.b16, compiled on Wed Mar 26 09:39:47 CET 2008
- Config ID: config-belladonna_git-belladonna.aei.mpg.de-datal-baiotti-Cactus
- Build ID: build-belladonna_git-belladonna.aei.mpg.de-baiotti-2008.03.26-08.30.59-9753
- Simulation ID: run-Meudon-moving-6lev-move3-45km_all-time-refinement_dtfac-node0058.admin-baiotti-2008.03.26-10.04.14-9773
- Run ID: run-Meudon-moving-6lev-move3-45km_all-time-refinement_dtfac-node0058.admin-baiotti-2008.03.26-10.04.14-9773

Below the job information is a section titled 'Simulation Events Published During Runtime (in descending order)'. It features a table with the following columns: Time, CCTK iteration, CCTK time, Event, Key, and Value.

Figure 3: Complete presentation of available metadata for a specific Cactus simulation (as selected by the user)

Finally it is possible to select two or more simulations on the overview page and click on an action button to compare the parameter files used by these simulations. Again, this action triggers the execution of a dynamically generated metadata query which internally retrieves the contents of the parameter files; they are parsed accordingly, and the comparison results are presented eventually on a webpage in a table, sorting individual parameters in table rows, colour-coded for identical (green) or different (red) values (blue if the parameter isn't set in one of the simulations compared).

Internal Handling of Metadata in the Cactus Simulations portlet

All Cactus simulation metadata is stored in RDF format and managed by an AstroGrid-D information service based on the OpenRDF framework Sesame[11, 12]. The Cactus portlet uses the Jena Semantic Web Framework API[13] to access this metadata information service. According to the user's interaction with the Cactus portlet, specific SPARQL queries are generated dynamically and send to the information service; the results are received in XML form and processed further before finally presenting them to the end user in the form of a HTML page.

```

Contents of parameter file Meudon-moving-6lev-  
finest.par

Cactus::cctk_run_title = "BNS inspiral"
Cactus::cctk_full_warnings = yes
Cactus::terminate = never
Cactus::cctk_final_time = 800.0

ActiveThorns = "Fortran"
ActiveThorns = "LocalInterp AEILocalInterp"
ActiveThorns = "Slab"
ActiveThorns = "IOUtil"
ActiveThorns = "InitBase Carpet CarpetLib CarpetInterp CarpetReduce CarpetSlab"
ActiveThorns = "NaNChecker"
ActiveThorns = "Boundary CartGrid3D CoordBase SymBase "
ActiveThorns = "CarpetRegrid2 SphericalSurface"
ActiveThorns = "ADMBase ADMCoupling ADMMacros CoordGauge SpaceMask StaticConformal"
ActiveThorns = "Time"
ActiveThorns = "MoL"
ActiveThorns = "BSSN_MoL"
ActiveThorns = "ADMAnalysis"
ActiveThorns = "AHPFinderDirect"
ActiveThorns = "WaveExtract"
ActiveThorns = "SphericalHarmonics SummationByParts PsiKadellia"
ActiveThorns = "ADMConstraints"
ActiveThorns = "CarpetIOScalar CarpetIOASCII CarpetIOBasic CarpetIOHDF5"
ActiveThorns = "Dissipation"
ActiveThorns = "TimerReport Formaline"
    
```

Figure 4: Example Cactus Simulations portlet webpage illustrating the comparison of parameter files for two Cactus simulations

User-defined SPARQL Queries

As stated above, dynamic metadata queries are typically generated internally by the Cactus portlet based on the interaction of the user with the portlet (eg. by following a given URL or by clicking an action button in a web formular). The details and the complexity of the metadata management (RDF representation, RDF schema used to describe Cactus simulation metadata, formulation of metadata queries in the SPARQL language) are hidden to the user as much as possible.

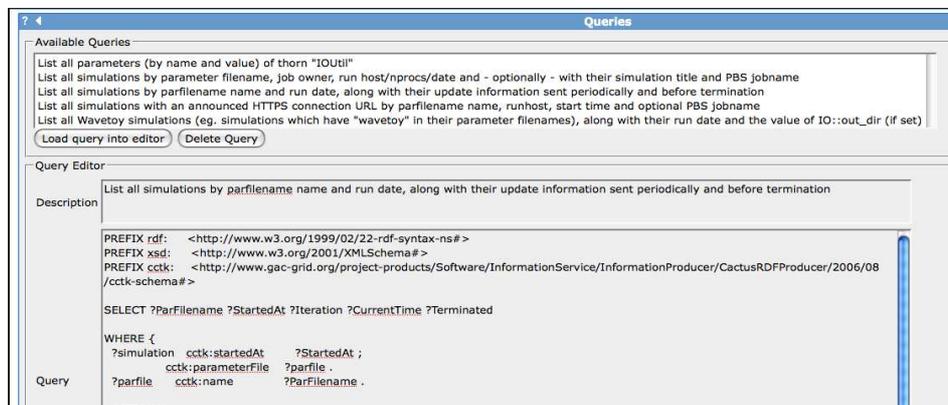


Figure 5: SPARQL Query editor provided by the Cactus Simulations portlet

However, for development purposes the Cactus portlet was extended by additional functionality which allows a user to define arbitrary SPARQL queries directly in a query editor (figure 5) and save them for later reuse; the results of such queries are displayed in a simple table-like structure, with no further context information specific to the actual application for which the metadata was collected.

2.3 GEO600 Portlet

In close collaboration with the D-MON project[14], the GEO600 portlet was developed to present a job statistics across all Grid resources for data analysis jobs of the GEO600 use case[4]. The main intension here was to demonstrate the production-mode execution of gravitational wave data analysis on the Grid (both on the German D-Grid as well as on the mostly U.S. based Open Science Grid). Like the ClusterFinder portlet, the GEO600 portlet is integrated in the official AstroGrid-D user portal (<http://cashmere.aip.de:8080/gridsphere/gridsphere/guest/geo600/r/>).

GEO600 job statistics information is automatically generated by the GEO600 Grid application and stored as RDF metadata in an AstroGrid-D metadata information service. The GEO600 portlet retrieves the job statistics information via SPARQL metadata queries to this AstroGrid-D service, and presents the results to the portal user in various ways, either in graphical or textual format. As an example, figure 6 shows a screenshot of job statistics information presenting the query results about the number of submitted and currently executed GEO600 jobs, along with the total CPU time spent by them, as a timeline for the month of January 2008. The GEO600 portlet also provides submenus to select a different period of time, or to switch from displaying accumulated values across all resources to showing results for an individual Grid resource only.

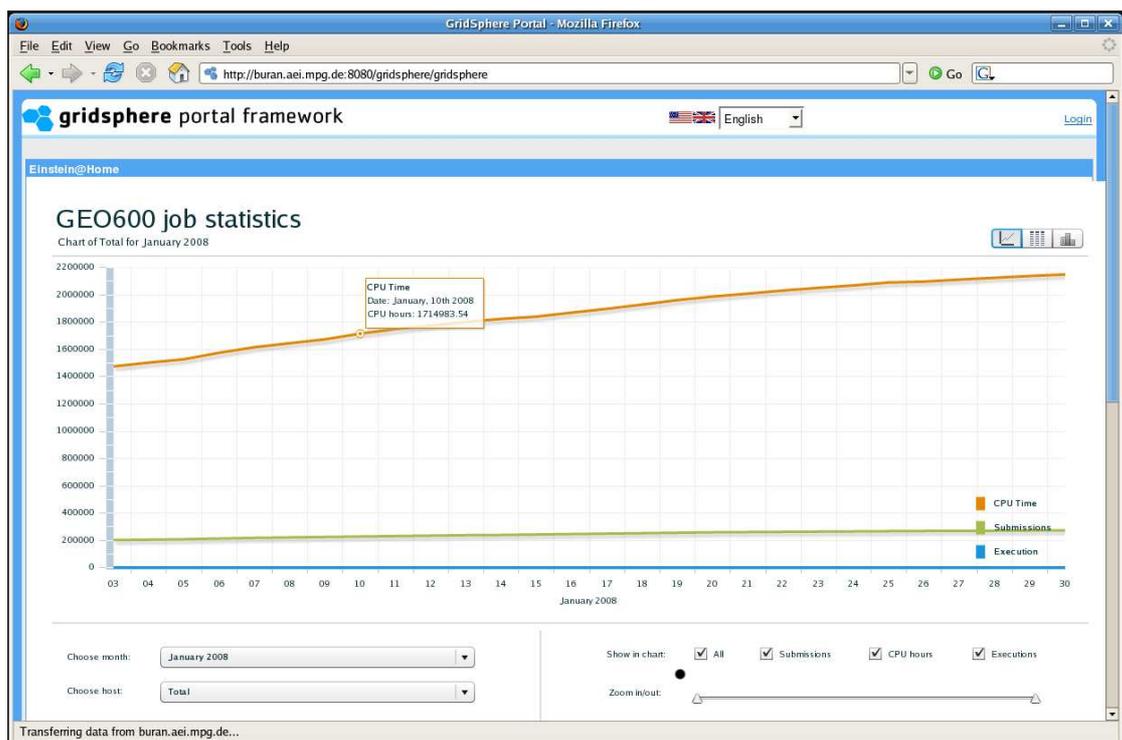


Figure 6: Screenshot of the job statistics page as shown in the GEO600 portlet

Similar to the Cactus simulations portlet, a SPARQL query is dynamically generated on each user interaction based on the submenu selection parameters as currently chosen. Query results are retrieved as an XML document which is further processed within the GEO600 portlet, and finally a Flash player graphics (either a timeline diagram or a table with the textual information) is generated from it and presented to the user as an embedded HTML object in a portal webpage.

3 Summary

This deliverable document describes how metadata generated by AstroGrid-D application use cases and services are being managed, processed, and presented to the end user in a portal user interface such as the AstroGrid-D user portal. Due to the different types of metadata and the users' interactions with them, it was concluded that corresponding user interfaces must be designed with the actual application in mind so that they are intuitive and user-friendly also to the ordinary scientist who cannot be expected to have a deep knowledge about the underlying Grid and web technologies.

Three portlets have been implemented for specific AstroGrid-D use cases (ClusterFinder, Cactus, GEO600). Their individual user interfaces and internal handling of metadata are documented in some detail. All portlets are integrated in a web-based portal (based on the GridSphere portal framework) which serves as a single point of contact for the end user, providing a common look-and-feel graphical user interface even for rather different use cases.

Although the type of metadata and the internal handling may differ from portlet to portlet, the common use of the standardized RDF metadata format, along with an appropriate API to deal with RDF metadata (the Jena framework), was found to greatly simplify portlet code development and reuse. This approach is advised to be followed also for the design and implementation of future AstroGrid-D application portlets.

F: References / Bibliography

References

- [1] Distributed File Management – Data- and Replica-management in AstroGrid-D. AstroGrid-D Deliverable Document D3.2
http://www.gac-grid.org/project-documents/deliverables/wp3/D3_2.pdf
- [2] AstroGrid-D Job Monitoring
<http://www.gac-grid.org/project-products/Software/job-monitoring.html>
- [3] AstroGrid-D Use Case *Cactus*
<http://www.gac-grid.org/project-products/Applications/Cactus.html>
- [4] AstroGrid-D Use Case *GEO600 Data Analysis*
<http://www.gac-grid.org/project-products/Applications/Geo600.html>
- [5] AstroGrid-D Metadata Information Service *Stellaris*
<http://www.gac-grid.org/project-products/Software/InformationService.html>
- [6] AstroGrid-D Use Case *CluserFinder*
<http://www.gac-grid.org/project-products/Applications/ClusterFinder.html>
- [7] Cactus Users Portal
<https://portal.cactuscode.org>
- [8] Numerical Relativity Portal
<https://portal.aei.mpg.de>
- [9] GridSphere Portal Framework
<http://www.gridsphere.org>
- [10] Prototype Implementation of grid-enabled Monitoring Methods – Documentation and Test Report AstroGrid-D Deliverable Document D6.4
http://www.gac-grid.org/project-documents/deliverables/wp6/WG6_D6_4.pdf
- [11] OpenRDF.org Homepage
<http://openrdf.org/>
- [12] Advanced Prototype Implementation of Metadata Information Providers – Documentation and Test Report; AstroGrid-D Deliverable Document D2.6
http://www.gac-grid.org/project-documents/deliverables/wp2/D2_6_Information_Providers.p
- [13] *Jena* – A Semantic Web Framework for Java. Homepage
<http://jena.sourceforge.net/>
- [14] D-MON – Horizontal Integration of Resource and Service Monitoring in D-Grid.
Project homepage <http://www.d-grid.de/index.php?id=401>