

GReIC

Services for Heavy User Communities

EGI Technical Forum 2011

S. Fiore and G. Aloisio

SPACI and University of Salento

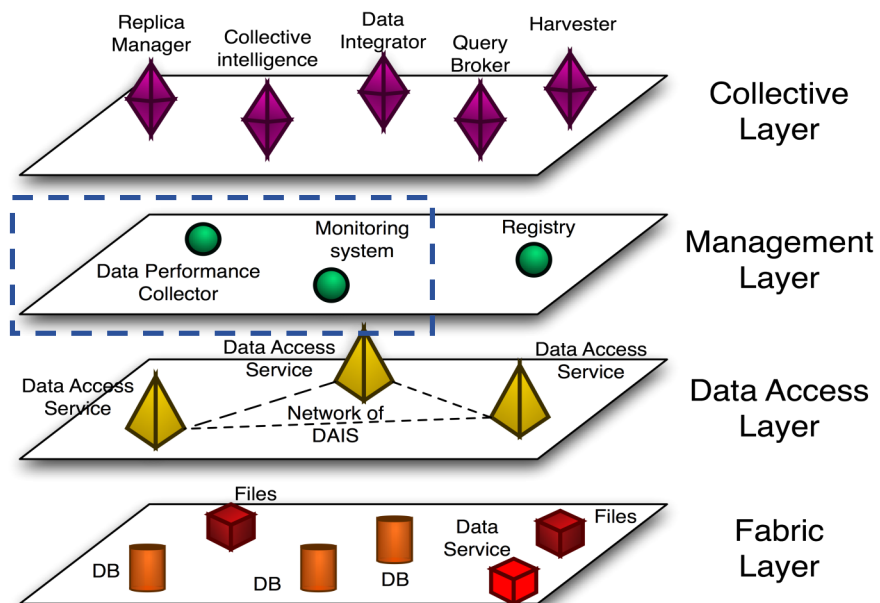
The GReIC Project: main goal and service

- Grid Relational Catalog is a project which aims at designing and developing a set of efficient, secure and transparent Data Grid Services (Starting date, January 2001).
- **GReIC Service** aims at providing a large set of functionalities to access to both relational and non relational Databases in a grid environment.



DoW (I) - “Management Layer”

An important task will be related to the monitoring and control functionalities connected with the underlying P2P infrastructure of the GRelC system. Such a management framework will be managed through the GRelC Portal by means of a new set of web pages exploiting the dashboard approach (charts, reports, table, diagrams able to provide global and local views about the status of the system). Users will be able to configure, manage and query their own GRelC services, exploiting a wide set of management functionalities embedded into the GRelC Portal. A key point will be to make easier (few steps in a web-based wizard) the *gridification process* (bringing into the grid) of a database resource.



GRELC
AN EASY WAY TO MANAGE GRID DATABASES

Home Components Downloads Documentation News Events Publications Staff Deployment Portal

Query
Synchronous Query
Server - gandalf
GRelC DAs
Database - sakila
Query - select * from actor

actor_id	first_name	last_name	last_update
1	PENELOPE	GUINNESS	2006-02-15 04:34:33
2	NICK	WAHLBERG	2006-02-15 04:34:33
3	ED	CHASE	2006-02-15 04:34:33
4	JENNIFER	DAVIS	2006-02-15 04:34:33
5	JOHNNY	LOLLOBRIGIDA	2006-02-15 04:34:33
6	BETTE	NICHOLSON	2006-02-15 04:34:33
7	GRACE	MOSTEL	2006-02-15 04:34:33
8	MATTHEW	BERHANSSON	2006-02-15 04:34:33
9	JOE	SWANK	2006-02-15 04:34:33
10	CHRISTIAN	GABLE	2006-02-15 04:34:33
11	ZERO	CAGE	2006-02-15 04:34:33
12	KARL	BERRY	2006-02-15 04:34:33
13	UMA	WOOD	2006-02-15 04:34:33
14	VIVIJEN	BERGEN	2006-02-15 04:34:33
15	CUBA	OLIVIER	2006-02-15 04:34:33
16	FRED	COSTNER	2006-02-15 04:34:33
17	HELEN	VOIGHT	2006-02-15 04:34:33
18	DAN	TORN	2006-02-15 04:34:33
19	BOB	FAWCETT	2006-02-15 04:34:33
20	LUCILLE	TRACY	2006-02-15 04:34:33

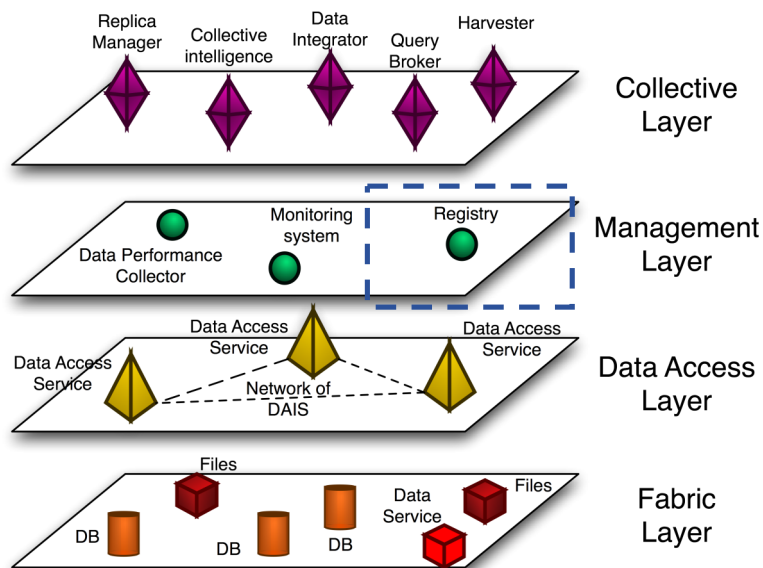
Previous 1 2 3 4 5 6 7 8 9 10 Next

Your DN is CN=grelc04.unile.it,L=HPCC University of Lecce,OU=Host,O=INFN,C=IT - Your proxy is valid until 27/09/2007 01:57:21

DoW (II) - “Management Layer”

A key task will be the creation of the EGI Database of Databases, a registry service accessible through a specific GRelC Portal web page that will contain all of the information about the grid-databases available in the GRelC System. Users will be able to:

- *query the registry* (exploiting a keyword-based approach) asking for specific databases, filtering by VO, keywords, domain, etc. This will help people working in a specific domain to quickly identify available and related resources, identify key people working on specific subjects, easily contact them to establish collaborations, etc.
- *join a specific grid-database*, submitting via the web a request to the grid-database administrator to know more about the supported VOs, etc.;
- *add comments* on the available data and the related data sources, being part of a community exploiting a collaborative and Web2.0 oriented approach. All of this data will be available for future users, creating a knowledge base centred around community-oriented topics.



The screenshot shows the Global Registry web interface. It includes a Settings section with filters for Project Name, Host Name, V.O., Publish Date, and Tag. Below the settings is a table listing database entries.

Database Name	Description	Publish Date	V.O.	Project Name	Host Name	Score
DB1	Bioinformatics database	08/29/2009 16:35	Virtual Organization 1, Virtual Organization 2	GRelC	GOOGLE	★★★★★
DB2	Eath Science database	08/20/2010 16:34	Virtual Organization 1, Virtual Organization 3	GRelC	GOOGLE	★★★★★
DB3		09/04/2010 17:06	Virtual Organization 3	Priva	ESCHER	★★★★★

DashboardDB and EGI (I)

- A new system (**DashboardDB**) more targeted on the GRelC service has been designed during Y1
 - It represents a unified environment (web based) joining social, management and monitorings aspects
 - Key aspect: focus on “grid-databases”

- Non functional requirements:
 - Pervasivity, user-friendliness and transparency
 - A web based solution is a good candidate
 - Security
 - ...taking into account the security implementation must not be a barrier for new users
 - Look and Feel
 - Technolgical impacts on the adopted software libraries

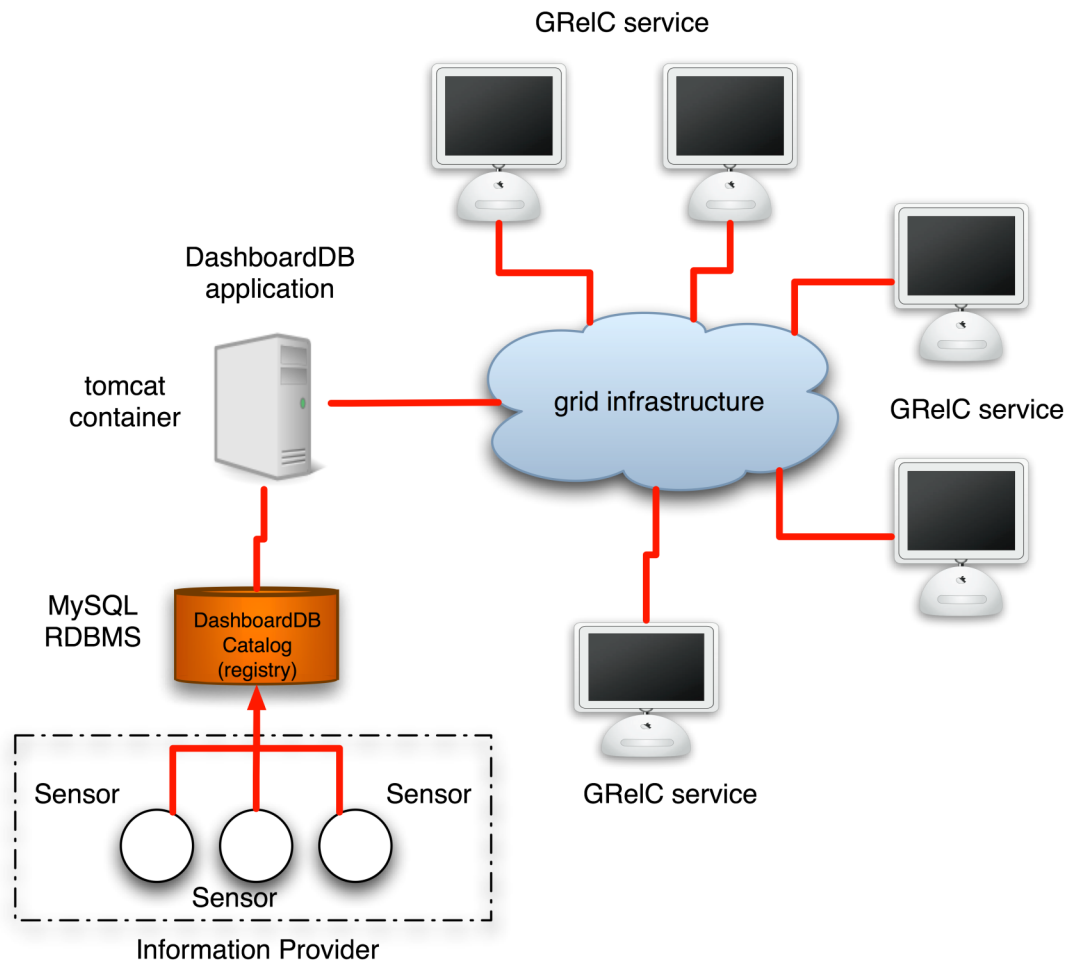
DashboardDB and EGI (II)

- Functional requirements:
 - Monitoring of GRelC service instances
 - Provision of specialized views related to the “network” of GRelC services
 - Database/VO association
 - Database distribution
 - etc.
 - Creation of a community oriented registry of grid-database resources
 - Discussion groups
 - Tagging capabilities
 - etc.

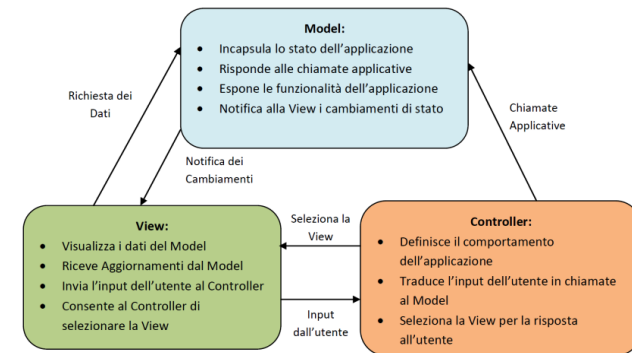
- Important Features
 - Permalink support
 - Support for multiple views
 - *Based on countries, goal, etc.*

DashboardDB: Architecture

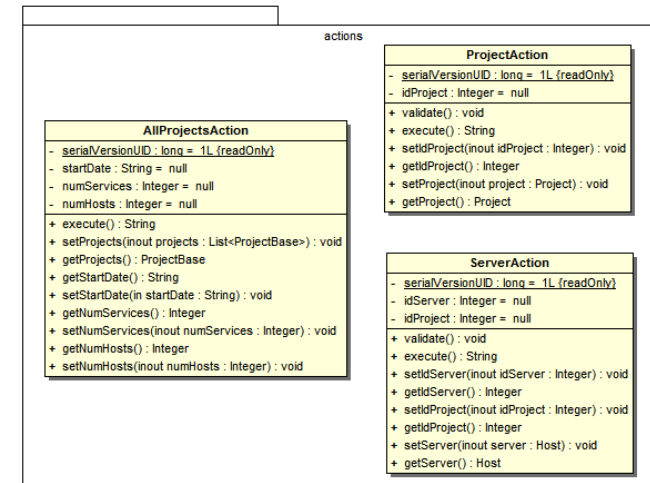
System Architecture



Model View Controller Pattern



Main actions



The DashboardDB Registry: main view

GReIC Registry



Registry Information

Grid DB Number: 2
GReIC Services Number: 2
Permalink: [code](#)

Filters

Filter By:
Project Name: -
V.O.: -
Publish Date:
Host Name: -
Tag:
Conditions: All Any

Grid

Discussion Group	Database Name	Description	Publish Date	V.O.	Projects Name	Host Name	Score
Leave	DB 1	Bionformatics database related to the UNIPROT databank. Relational and federated approach running on top of Postgresql 9.0.1	09/28/2010 19:45		Climate-G.	Gandalf	☆☆☆☆☆
Join	DB 2	Eath science database	09/28/2010 19:45				

General information

Filters

Grid Database information

Permalink



The DashboardDB Registry: grid-DB view

Grid-DB Details

The screenshot displays the 'DB 1 Details' page in the DashboardDB Registry. The page is divided into several sections:

- Information:** Contains a 'Rate' section with a star rating and a 'Description' field. The description reads: "Bionformatics database related to the UNIPROT databank. Relational and federated approach running on top of Postgresql 9.0.1". Below the description are 'Save Changes' and 'Clear' buttons.
- Join Grid-DB:** A section titled 'Join DB' with a 'Send Request' button. The text says: "Send a request to the administrators, if you want to get access to this Grid-DB."
- Tag Cloud:** A section titled 'Tag Cloud' with an 'Add Tag' input field and an 'Add' button. Below the input field, the tags 'bioinformatics', 'genome', and 'protein' are displayed.
- Comments:** A section titled 'Comments' containing a table of user comments and their dates.

Green arrows highlight specific features: one points to the 'DB 1 Details' tab, another to the 'Rate' section, a third to the 'Description' field, a fourth to the 'Join Grid-DB' section, a fifth to the 'Tag Cloud' section, and a sixth to the 'Messages list' (comments table).

User & Comment	Date
Are you planning to federate other DBs into this one?	You 09/21/2011 09:08
Great! Now it's fine! Thank you very much.	You 09/21/2011 09:02
I have just updated it! Please check and thanks for your comment.!	fiore 09/21/2011 09:02
This DB is not updated. Please check it and let me know...	You 09/21/2011 09:01

DashboardDB: the Registry

The screenshot displays a web application interface for a database registry. At the top, there are tabs for 'GReIC Registry' and 'DB 1 Details'. The main content area is titled 'User & Comment' and contains a list of messages. A table on the right side of the messages lists the user and the date/time of each post. Annotations with green arrows point to various parts of the interface: 'Users posting messages ("Who")' points to the user names in the table; 'Messages ("What")' points to the text of the messages; 'Date/time ("When")' points to the date/time column in the table; 'Join/Leave a discussion group' points to the 'Leave' button; and 'Add new comments' points to the comment input field. A pagination bar shows 'Page 1 of 1' and 'Displaying topics 1 - 4 of 4'. A 'Leave' button is accompanied by a note: 'Please click on the Leave button if you want to leave this DB discussion group (you'll no longer receive by email notifications about new comments posted here)'.

User & Comment	Date
Are you planning to federate other DBs into this one?	You 09/21/2011 09:08
Great! Now it's fine! Thank you very much.	You 09/21/2011 09:02
I have just updated it! Please check and thanks for your comment.!	fiore 09/21/2011 09:02
This DB is not updated. Please check it and let me know...	You 09/21/2011 09:01

Page 1 of 1 | Displaying topics 1 - 4 of 4

Comment:

Add Clear Leave

Please click on the Leave button if you want to leave this DB discussion group (you'll no longer receive by email notifications about new comments posted here)

DashboardDB: Security aspects


Security Management:

- User Registration
- User Authentication
- User Profile Management
- User Authorization
- Guest users (access to public projects)

User Profile

First Name*:	<input type="text" value="dashgadmin"/>
Last Name*:	<input type="text" value="dashgadmin"/>
Country:	<input type="text" value="Italy"/> ▼
User DN:	<input type="text"/>
Username:	<input type="text" value="dashgadmin"/>
E-mail:	<input type="text" value="admin@admin.com"/>
Password:	<input type="text"/>
Password (confirm):	<input type="text"/>

User Registration

First Name*:	<input type="text"/>
Last Name*:	<input type="text"/>
Country:	<input type="text"/> ▼
User DN:	<input type="text"/>
E-mail*:	<input type="text"/>
Username*:	<input type="text"/>
Password*:	<input type="text"/>
Password (confirm)*:	<input type="text"/>
Captcha*:	

Username Password

DashboardDB: Permalinks and Mashup

By including into a target web page a simple line of code like:

...

```
<iframe  
src="http://host:8080/dashboardDB/.../ProjectRegistry...?request_lo  
cale=en&idProject=...&frame=.../ProjectRegistry...%3Frequest_locale%3De  
n%26idProject%3D5" height="600" width="100%"></iframe>
```

...

you can embed the DashboardDB registry into your web application in a straightforward manner like a YouTube video.

Authorization can be turned on/off into the target web page

Reusability can strongly be addressed by exploiting permalink capabilities (key issue for software sustainability)

DashboardDB: “embedding” the registry

Wednesday, 21st September 2011 12:19 PM +0200



GRELC

AN EASY WAY TO MANAGE GRID DATABASES

[Home](#)
[Components](#)
[Downloads](#)
[Documentation](#)
[News](#)
[Events](#)
[ig_GRELC](#)
[Portal \(BETA\)](#)

Headlines

- 18-08-2010** **GRELC DAIS version 3.1.3 is now available for gLite 3.1**
- 16-11-2009** **GRELC Portal v2.0 online today**
- 16-11-2009** **GRELC DAIS version 3.1.2 is now available for gLite 3.1**
- 12-06-2009 GRELC & Climate-G
- 14-05-2009 CFP HDDM09 Workshop
- 08-04-2009 DaGreS09 Workshop Schedule
- 24-12-2008 CFP DaGreS09 Workshop
- 20-12-2008 GRELC C Client API New release
- 13-10-2008 Climate Modelling and EGEE
- 29-09-2008 HPDataGrid08 Workshop Schedule

Events / Invited Lectures

- 16-09-2010 GRELC Talk @ EGI-TF, Amsterdam
- 16-09-2010 GRELC @ EGI-TF, Amsterdam
- 03-05-2010 GRELC @ EGU, Vienna
- 08-12-2009 GRELC @ AHM - Oxford, UK
- 09-11-2009 GRELC Tutorial @ ICITST09
- 09-11-2009 GRELC Talk @ ICITST09

Mission

The **Grid Relational Catalog Project (GRELC)** aims at providing a set of advanced data grid services to transparently, efficiently and securely manage Databases on the Grid.

The GRELC middleware is currently used within several grid research projects to support bioinformatics experiments on distributed and huge data banks as well as the metadata management related to Earth Observation System applications, etc.

At the moment the GRELC Data Access and Integration Service (GRELC DAIS) allows users to access and interact with different DBMSs both relational (PostgreSQL, MySQL, Oracle, DB2, SQLite,



RSS feed



University of Salento, Lecce



SPACI Consortium



CMCC



EGEE RESPECT Program



Climate-G

Welcome dashgadmin dashgadmin [Sign out](#)

GRELC Registry

Registry Information

Grid DB Number: 2
 GRELC Services Number: 2
 Permalink: [code](#)

Filters

Filter By:

Project Name: Host Name:
 V.O.: Tag:
 Publish Date: Conditions: All Any

Grid

Discussion Group	Database Name	Description	Publish Date	V.O.	Projects Name	Host Name	Score
Leave	DB.1	Bionformatics database	09/28/2010 19:41		Climate-G, GRELC	Gandalf	★★★★★
Join	DB.2	Eath science database	09/28/2010 19:41		Climate-G, GRELC	Scalmeta	★★★★☆

Ongoing activities and new ones planned for Y2

Ongoing activities and new ones planned for Y2:

- **Porting** of the GRelC software on:
 - **gLite 3.2** (SL5.x) very soon (some problems with 64bits SSL libraries on SL5, prevented the team to release the software at the end of Y1)
 - ... and on **EMI** soon after that
- HUC support activities:
 - **LS**: A GRelC service has been deployed in our site to support LS database management (user support activity). In particular a use case regarding the **UNIPROT data bank** has been implemented
 - **ES**: the **Climate-G Portal** will integrate the **DashboardDB monitoring** facility.
 - **Tutorial** and **training** events (next event scheduled in December at the PDCS2011 conference, Dallas, Texas)
 - Participation in “**user community oriented**” activities (i.e. ES, LS), initiatives and conferences (AGU2011, EGU2011 and EGU2012, etc.)
 - **Project website** and **GILDA tutorials**

HUC Life Sciences Support: the UNIPROT use case

- In Q4-Q5 a **new use case for LS** has been jointly defined with bioinformatics people at the University of Salento. The main goal of this use case was to make the Uniprot database available to the LS community through a GRelC service interface.
- A relational-based schema of the Uniprot database has been designed and implemented.
- An ETL (Extraction-Transformation-Loading) tool to move the data from the Uniprot/Swiss-Prot flat file into a relational DB has been implemented and tested jointly with the bioinformatics group.
- The database schema includes **30 relational tables (13GBs of data)**.
- The relational version of the Uniprot DB has been deployed on the machines provided by SPACI to support these use cases.
- The database allows submitting queries like:
 - **Query 1:** *Given a protein, select the OG (OrGanelle) that indicates if the gene coding for a protein originates from mitochondria, a plastid, a nucleomorph or a plasmid.*
 - **Query 2:** *Given a protein, select the specie, its classification and taxonomy.*

Contact point for this activity: maria.mirto@unisalento.it

UniProtKB/Swiss-Prot Release 2011_05 of 03-May-2011 30 Tables

Table Name	Num_entry
OriginDB	129
Gene	84260
Organism	13008
ordlocname	381179
orfname	73583
organel	845
topic_comment	40
db_organism_identifier	1
organism_class	8347
synonyms	52157
primary_identifier	3937759
keyword_name	1052
sequence_type	1
status_entry	1
Molecule	526969

Table Name	Num_entry
organism_taxonomy	12463
organism_classification	122209
originated_by	537622
gene_synonyms	56117
accession	694964
accession_number	711407
gene_codified_by	458981
orf_codified_by	76407
ord_codified_by	381473
molecule_organel	20223
comment	2206025
feature	3358177
referenced_into_db	8711973
keyword	3250350
reference	931428

HUC Life Sciences Support: the UNIPROT use case

Advantages

- Reducing the redundancy present into the flat file;
- Reducing the inconsistency of data that could have different values in the flat file;
- More performing searches querying the relational database, by using the GReIC service;
- Complex queries by using a standard language such as SQL.

Next steps

- Taking into account the user requirements, in the near months it is expected to increase the number of biological data banks accessible via the GReIC interface
- The UNIPROT data bank will be published on the DashboardDB registry

User support: the GRelC WebSite



Main sections:

- Download (rpms available)
- News
- Publications
- Events
- Deployment
- Documentation
- Components
-

GRelC Website

URL: <http://grelec.unile.it/>

Mailing List

mail: grelec-user@sara.unile.it

User support: tutorials on GILDA

GReIC DAS User Tutorial on GILDA Grid CT Wiki Website

Info about:

- Log in to the grid
- Query Submission



A screenshot of a Mozilla Firefox browser window displaying the GReIC Data Access Service 2.2.0 page. The browser's address bar shows the URL: https://grid.ct.infn.it/twiki/bin/view/GILDA/GReICDataAccessService. The page content includes a navigation menu on the left with options like "Log In or Register", "GILDA Web", "Create New Topic", "Index", "Search", "Changes", "Notifications", "Statistics", and "Preferences". Below the menu is a "Webs" section with links to EUMEDGRID, GILDA, ICEAGE, Main, PIZS2, Sandbox, TRIGRID, and TWiki. The main content area features the "GReIC Data Access Service 2.2.0" title, a "Supervisor" Prof. Giovanni Aloisio, "Project P.I." Sandro Fiore, Ph.D., "Team members" Massimo Cafaro, Alessandro Negro, Salvatore Vadacca, and a "Website" link to http://grelc.unile.it. There are also sections for "Introduction" and "GReIC Project" with descriptive text.

For any information about GILDA t-Infrastructure please
contact roberto.barbera@ct.infn.it & grid-prod@ct.infn.it

GReIC DAS Tutorial link:

<https://grid.ct.infn.it/twiki/bin/view/GILDA/GReICProject>

**Special thanks to the GILDA
Staff for their support**

Some useful information

Fon any information

Project P.I.: S. Fiore (sandro.fiore@unisalento.it)

GReIC WebSite: <http://greic.unile.it>

GILDA support: <https://grid.ct.infn.it/twiki/bin/view/GILDA/GReICProject>

Mailing lists: greic-user@sara.unisalento.it



Some useful references

[1] S. Fiore, et al., The Climate-G Portal: The context, key features and a multi-dimensional analysis,, Future Generation Computer System, Vol 28, pp.1-8 (2012), [doi:10.1016/j.future.2011.05.015](https://doi.org/10.1016/j.future.2011.05.015).

[2] S. Fiore, G. Aloisio, Special section: Data management for eScience. Future Generation Computer System 27(3): 290-291 (2011)

[3] S. Fiore, et al., The Data Access Layer in the GReIC System Architecture, Future Generation Computer System, 27(3): 334-340 (2011), <http://dx.doi.org/10.1016/j.future.2010.07.006>

[4] S. Fiore, et al., The GReIC Project: from 2001 to 2011, ten years working on Grid-DBMSs, in Grid and Cloud Database Management, Springer. Edited by S. Fiore and G. Aloisio.

[5] S. Fiore, G. Aloisio, P. Fox, M. Petitdidier, H. Schwichtenberg, S. Denvil, J. D. Blower, A. Cofino, The Climate-G testbed: towards large scale distributed data management for climate change, Proceedings of the International Conference on Computational Science ICCS 2011, June 1 - June 3, 2011, Nanyang Technological University, Singapore, Procedia Computer Science, Elsevier, pp. 567-576.

[6] S. Fiore and G. Aloisio, “[Grid and Cloud Database Management](#)”, 2011. Springer, ISBN 978-3-642-20044-1