

# AQUAGRID: an extensible platform for collaborative problem solving in groundwater protection

G. Lecca · C. Lai · F. Murgia · R. Biddau · L. Fanfani · P. Maggi

Received: 12 September 2008 / Accepted: 21 January 2009 / Published online: 18 February 2009  
© Springer-Verlag 2009

**Abstract** AQUAGRID is the subsurface hydrology computational service of the Sardinian GRIDA3 infrastructure, designed to deliver complex environmental applications via a user-friendly Web portal. The service aims to provide to water professionals integrated modeling tools to solve water resources management problems and aid decision making for contaminated soil and groundwater. In this paper, the AQUAGRID application concept and enabling technologies are illustrated. At the heart of the service are the computational models to simulate complex and large groundwater flow and contaminant transport problems and geochemical speciation. AQUAGRID is built on top of compute-Grid technologies by means of the EnginFrame Grid framework. Distributed data management is provided by the Storage Resource Broker data-Grid middleware. The resulting environment allows end-users to perform groundwater simulations and to visualize and interact with their results,

using graphs, 3D images and annotated maps. The problem solving capability of the platform is demonstrated using the results of two case studies deployed.

**Keywords** Data grid · High-end computational clusters · Hydrogeochemical modeling · Problem solving platform

## Introduction

Monitoring complex environmental systems is extremely challenging because it requires capturing the processes governing the impacted system, elucidating human and ecologic risks, limiting the data collection costs and satisfying the interests of multiple stakeholders. In addition, environmental data collections, originating from heterogeneous sources like sensors, satellites, high performance computer simulations, high throughput devices and so on, are scattered in diverse locations/institutions with varying access policies. As a result, multidisciplinary problem solving platforms, providing tools for data sharing and re-use, integrated modeling, and e-collaboration and community building, are worldwide regarded as a key enabler for effective and efficient decision making in science and innovation (Beagrie 2007).

Environmental applications due to their intensive data processing and highly distributed organization, are uniquely positioned for the uptake and exploitation of Grid and Web 2.0 technologies. The Grid represents a solution allowing different organizations to federate together to deliver high value-added services, remotely accessed by users via user-friendly Web portals. These tasks require integrated provision of resources, high performance networking and computing, sharing of software efforts, and lowering of barriers between expert groups participating in the service development. Up to now, a limiting factor of a wider

---

Work presented at the EGU2008

G. Lecca (✉) · C. Lai · F. Murgia  
Center for Advanced Studies,  
Research and Development in Sardinia (CRS4),  
località Piscina Manna,  
09010 Pula, Cagliari, Italy  
e-mail: giuditta@crs4.it  
URL: <http://www.crs4.it>

R. Biddau · L. Fanfani  
Department of Earth Science, University of Cagliari,  
Via Trentino 51,  
09127 Cagliari, Italy

P. Maggi  
NICE srl,  
Via Marchesi del Roero 1,  
14020 Cortanze, Asti, Italy

acceptance of Grid solutions in science and engineering has been the complexity of the Grid technology as perceived by end-users. To enhance user experience and productivity, new computing models and paradigms characterized by easy-to-use high level online user interfaces, resource virtualization, service oriented computing and semantic Web technologies are increasingly being adopted. In a near future, semantic and ontology technologies will add to these knowledge infrastructures flexible and extensible schemas for data ingestion, access and processing, reducing the dependence on human intervention (EC FP7 DEGREE Project Consortium 2008).

In this paper, one of the five applications of the Sardinian GRIDA3 infrastructure (*Shared Resources Manager for Environmental Data Analysis and Applications-GRIDA3* project, <http://grida3.crs4.it>) is illustrated along with its enabling technologies. The AQUAGRID service is a powerful online tool to solve water resources management problems, oriented towards water professionals and decision makers, hiding to the end-users all unnecessary complexity of data analysis and simulation over distributed resources. The GRIDA3 platform, as an application service provider (Goulas et al. 2005), embraces a number of environmental application services, exposing them via standard Web technologies. Besides groundwater management, the other environmental challenges addressed by the project are the remediation of industrial polluted sites, the non-rational use of soil and water resources, and the hydrogeological risk related to meteorological extreme events and climate variability. The platform aims to develop a distributed data infrastructure along with a Grid-based decision support system, integrating GIS technologies, monitoring services and mathematical models for simulation, optimization, analysis, and risk assessment in spatial planning and environmental protection. The partners of the GRIDA3 initiative are CRS4, Nice srl, and the University of Cagliari.

### State of the art

Different domains of Earth and Environmental Sciences have been proven to be a prolific ground for the use and uptake of emerging technologies. One of the most important steps of this exploitation process is the joining-up of Web with Grid technologies to create what are called either Web-based Grid Portals or Grid-based Web portals. The Grid adds another fundamental layer to the Internet, enabling direct access to and use of underlying resources, such as computers, storage, scientific instruments and experiments, sensors, applications, data, and middleware services. Based on widely accepted Grid and Web services standards, resources communicate with each other and deliver results as services back to the user (Gentzsch 2007).

Typically the Grid-based Web portals orchestrate several dedicated services of one or more distinct Grid infrastructures (e.g. at European, national or organization scale). They provide a general, proven framework model which facilitates the addition and easy porting on the Grid of new science applications. They may also provide ready access to large Grid-based data collections and support the easy integration of new data collections for use in the Grid-based applications data processing.

Among various examples of existing Grid portals that are easily accessible from the net, we cite the LEAD (*Linked Environments for Atmospheric Discovery*), ESG (*Earth System Grid*) and GEON (*GEOScience Network*) portals. LEAD (Droegemeier et al. 2005) makes meteorological data, forecast models, and analysis and visualization tools available to end-users at one convenient access point, supported by high-performance computing systems. LEAD (<https://portal.leadproject.org/gridsphere/gridsphere?cid=portal-home>) end-users become active participants who can acquire and process their own data. ESG (<http://www.earthsystemgrid.org/home/home.htm>) is a research project sponsored by the U.S. DOE Office of Science whose goal is to provide discovery and secure access to very large datasets for earth sciences research (Pouchard et al. 2003). ESG through a combination of Grid and emerging community technologies, distributed federations of super-computers and large-scale data and analysis servers is developing a seamless and powerful environment for climate research. GEON (<http://www.geongrid.org/>) is a multi-disciplinary research initiative conducted jointly by earth and computer scientists to understand the complex dynamics of Earth systems by means of interlinking and sharing multidisciplinary data sets and tools. GEON is designed to create an infrastructure and a cultural environment to enable collaboration and scientific discovery (Keller et al. 2005).

In the field of water resources management, the SegHidro portal (<http://portalseghidro.dca.ufcg.edu.br>) is designed for researchers and decision makers to execute and couple computational models in a workflow to support optimal decisions about water reservoir management, agricultural planning and flood control. The portal provides a framework which allows seamless integration of the models, meaning that each phase of the flow may be executed by a different expert and that the resulting data is shared among other portal users (Voorsluys et al. 2007).

### GRIDA3 platform and enabling technologies

The GRIDA3 platform is made of a cluster of high-end computers located at CRS4 and federated data servers located at the departments of the University of Cagliari

involved in the project. The GRIDA3 portal (Fig. 1, AQUAGRID service) is built on top of compute-Grid technologies by means of the EnginFrame Grid framework (<http://www.enginframe.com>). Distributed data management is provided by the Storage Resource Broker data-Grid middleware (<http://www.sdsc.edu/srb>). The Web user interface is made up merging several technologies. The main structure is arranged by EnginFrame widgets, such as text areas, input fields, buttons, etc. (Fig. 2). These widgets populate a library of plugins that can be shared and re-used among different applications. Data with geographical meaning is graphically rendered using Google Maps (Figs. 3 and 5). In general, user interaction and facilities are automatically handled by native library. Auxiliary geographical information not immediately available to the user, can be reached with the GeoNames library (<http://www.geonames.org/>), if available. For instance, in the Web form to ingest into the system end-user data to be analyzed, once located a point on the map is possible to obtain related geographical information such as province name, postal code and so on. Case studies input data are managed by means of the PostgreSQL database (<http://www.postgresql.org/>). An *ad-hoc* relational schema has been created to

manage the geochemical data collections provided by the Department of Earth Science of the University of Cagliari (DISTER). PostgreSQL database has been extended with PostGIS technology (<http://postgis.refractory.net/>). The PostGIS enabled geochemical database allows to perform some additional functionalities such as the creation of spatial objects, spatial indexing, standard input/output representations, spatial operators and functions (e.g. on-the-fly coordinate transformation from different coordinate systems adopted by end-users to the WGS84 system adopted by Google Maps). Partner data are stored in the project database while end-user data are stored in a companion database for end-user data lineage.

Other open source modules are currently being integrated into the AQUAGRID application service: the UMN Map Server (<http://mapserver.gis.umn.edu/>), an environment for constructing spatially-enabled Web applications; the Geographic Resources Analysis Support System GRASS (<http://grass.itc.it/>, mirror site), a Geographic Information System (GIS) to perform geospatial data management and analysis, image processing, graphics/maps production, spatial modeling, and visualization; and mscross (<http://sourceforge.net/projects/mscross>) a Web-GIS client, developed at CRS4 as a

**Fig. 1** The AQUAGRID subsurface hydrology application portal. *Left side* menu shows available computational services (CODESA-3D and PHREEQC). Each service provides an *Application Management*

tool running demonstrative simulations based on dataset collections provided by partners and a *New locations* tool allowing end-users to build *ex-novo* applications

The screenshot shows the AQUAGRID application service interface. The browser title is 'AQUAGRID - GRIDA3 - Mozilla Firefox'. The page header includes 'AQUAGRID' and 'Grida3'. The navigation bar contains 'Application Home', 'My Data', 'My Jobs', and 'Cluster Load'. The main content area displays the 'Contents of spooler Codesa3D - KORBA' with a table of files:

File	Date	Size
inputData	Nov 5, 10:47	4096
Logs	Feb 10, 11:03	4096
Results	Feb 10, 11:03	4096

Below the file table, there is a table of job details:

ID	Project	Time	Status	Host	Name
148894[100]	g3giuditta	Feb 10 10:51	DONE	Codesa3D	

The job details section shows 'CPU: 25.5s' and 'Details' for 'Resource usage'. The submission and execution details are as follows:

- Submission:**
  - Job manager: lsf
  - Time: Feb 10, 10:51
  - Host: grd3port
  - Directory: /spoolers/g3giuditta/tmp14524.ef
- Execution:**
  - Time: Feb 10, 10:51
  - Host: stria16
  - Directory: /spoolers/g3giuditta/tmp14524.ef
- Termination:**
  - Time: Feb 10, 10:51
  - Total cpu usage: 25.5
  - Exit code: 0

The footer of the page includes logos for NICE and EnginFrame, and the text 'Powered by EnginFrame'.

**Fig. 2** AQUAGRID application service: Inspection of the run status and user directories (input, output and log files), after job submission. Cluster load button in the horizontal bar on top of the page allows to inspect also cluster load

JavaScript interface to UMN Mapserver, to dynamically create Web-GIS applications with support for WFS and WMS protocols, panning and zooming.

### EnginFrame Grid Portal

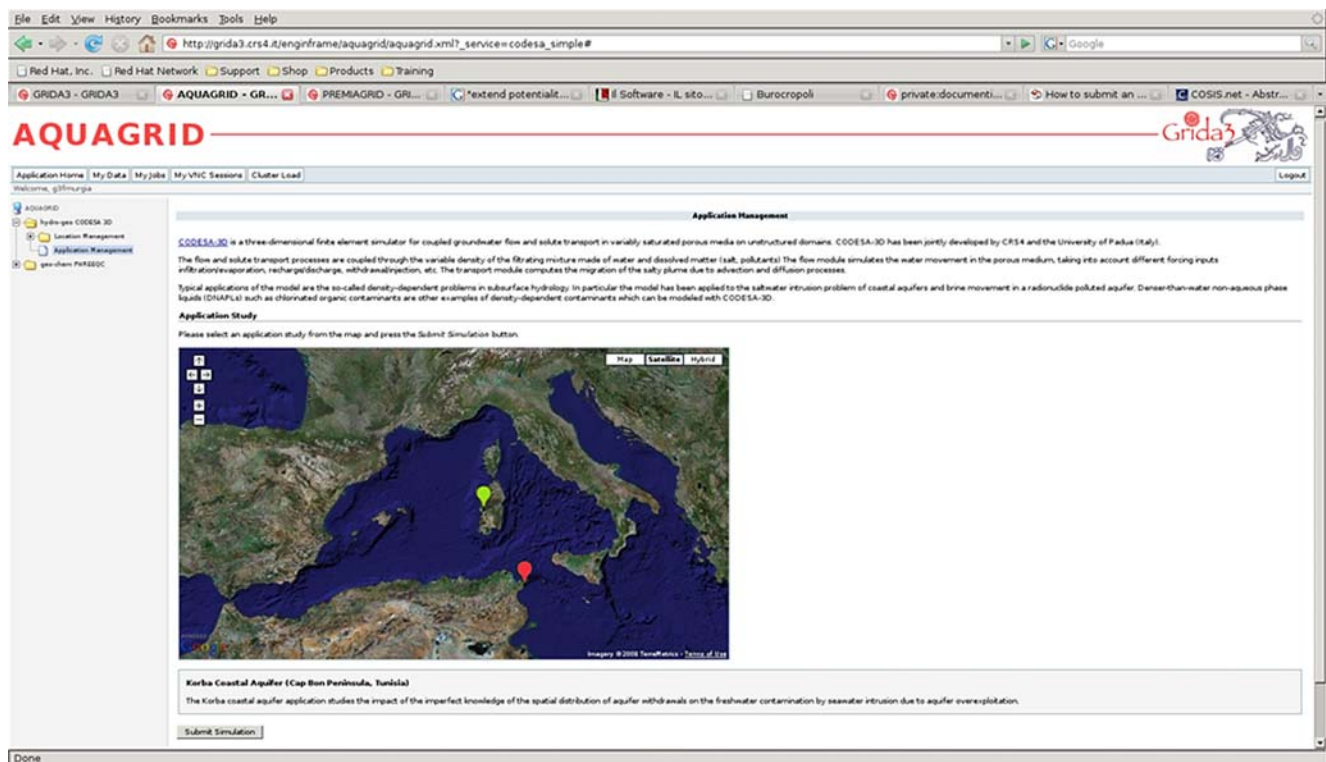
EnginFrame is a commercially supported Grid Portal that has proven track record of successful production deployments within corporate networks and research Grids like EGEE (<http://www.eu-egee.org/>) and DEISA (<http://www.deisa.org/>). EnginFrame role is to enable efficient Internet and Intranet access to Grid-enabled infrastructures: computing resources, data, software licenses, batch and interactive applications, by any client using a standard Web browser. Its open and evolutionary framework is based on Java, XML, XSLT and Web Services and facilitates the deployment of user-friendly, application- and data-oriented portals. Users and administrators can easily submit and control Grid-enabled applications, as well as monitor workload, data, and licenses from within the same user dashboard, hiding the heterogeneity and complexity of native interfaces of the underlying Grid infrastructure. It supports several open and vendor-neutral standards and seamlessly integrates with JSR 168 compliant enterprise portals, distributed file systems, data-grid middlewares and graphical user interface (GUI) virtualization tools.

Its authentication system provides delegation to a rich set of external authorization mechanisms (OS/NIS/PAM, LDAP, Microsoft Active Directory, MyProxy, Globus, etc) and provides internal, session level single-sign-on across multiple services. The authorization system allows to control in a fine-grained fashion which users can access a given resource, granting or denying operations according to a set of predefined policies. Leveraging the authorization system, it is also possible to tailor the Grid portal according to the specific users roles or access rights. EnginFrame also provides comprehensive data management, virtualization and correlation via the *spooler* concept. A spooler in the EnginFrame terminology is a dedicated data container to host files provided by users (e.g. input files uploaded using a Web browser) or generated by the services (e.g. output or temporary files). Users can interact with the data contained in the spoolers associated with their jobs, for example downloading output files (Fig. 2) or, when possible, previewing them inside the Web browser. Services associated with these functionalities are re-used among the different applications services of the GRIDA3 infrastructure.

EnginFrame features a three-tiers architecture (Fig. 7):

- The *Resource Tier* abstracts computing resources and data management. It consists of one or more Agents deployed on the back-end infrastructure. Agents manage





**Fig. 3** AQUAGRID-CODESA-3D hydrogeological application service: Case-study selection (via a mouse click on the map) and seawater intrusion simulation submission (via the “submit” button). The color of case-study locator on the map changes according to the user actions

computing resources on user’s behalf and interact with the underlying operating system, job scheduler, or Grid infrastructure to execute EnginFrame services (e.g. starting jobs, moving data, retrieving cluster load, etc.);

- The *Server Tier* exposes services to users. It consists of a Server (a Java Web Application deployed inside a Java Servlet Container) that interacts with the EnginFrame Agents and manages the interaction with the users;
- The *Client Tier* provides an easy-to-use interface based on established Web standards like XHTML and JavaScript. This tier is independent from the specific software and hardware environment used by the end user and usually consists of the user’s Web browser.

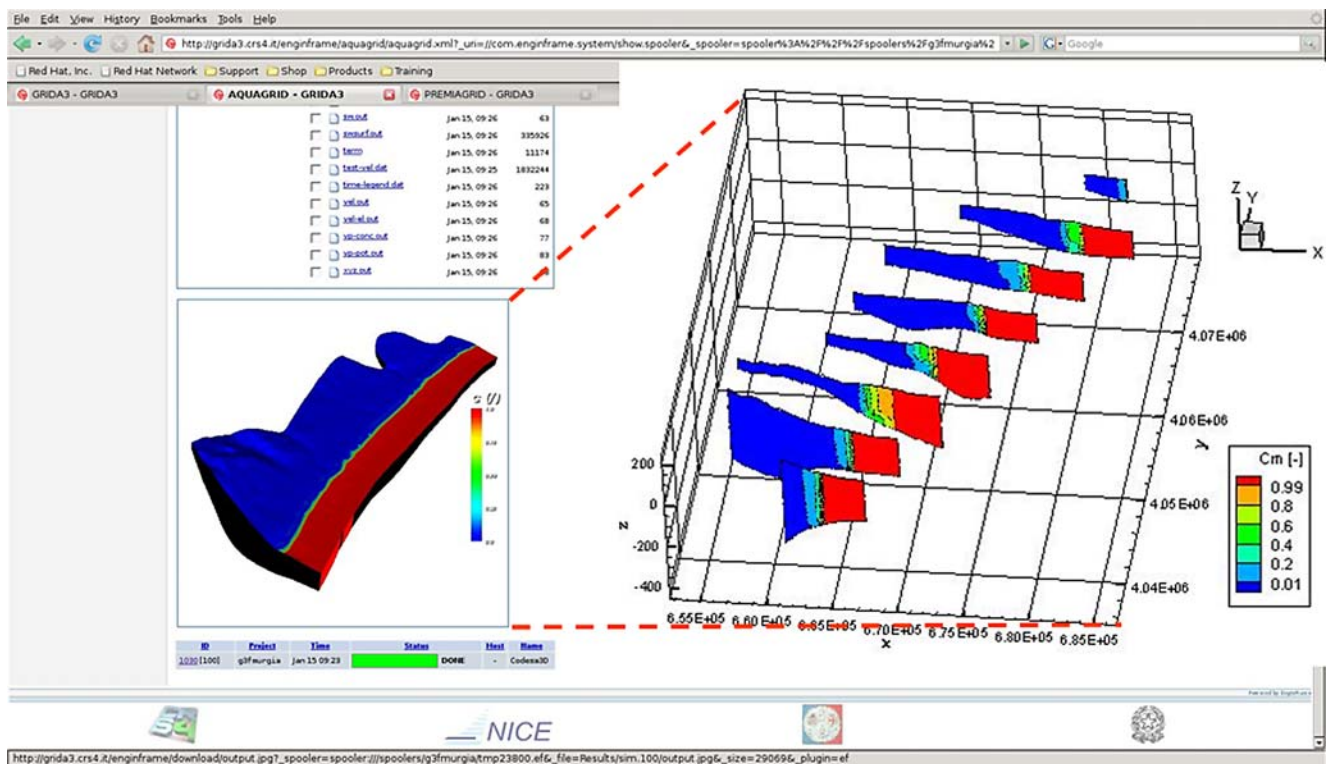
Most of the information managed by EnginFrame are described by dynamically generated XML documents. The source of such information is typically the service execution environment: an XML abstraction layer (Fig. 7) is aimed to submit service actions and translate raw results coming from the computational environment into XML structures. The XML abstraction layer is designed to decouple EnginFrame from the actual Grid working environment, hiding the specific Grid technology solution. All these characteristics make possible to easily extend EnginFrame functionalities by developing ad-hoc plugins

for specific computational and data Grid middlewares or legacy applications.

For instance, to create the AQUAGRID subsurface hydrology service, a service definition file (SDF), developed in XML and JavaScript, has been developed. Common services can be shared and re-used among the applications as application plugins, increasing the overall efficiency and productivity of the GRIDA3 framework. As an example, the public domain software Paraview (<http://www.paraview.org>) and Grace (<http://plasma-gate.weizmann.ac.il/Grace/>) have been integrated into AQUAGRID service to extend the EnginFrame capabilities to preview within the browser simulation results as 2D and 3D images (Figs. 4 and 6).

### Data virtualisation

The SDSC Storage Resource Broker (SRB) technology has been used to design the data-grid infrastructure. SRB client-server middleware provides a uniform interface for connecting to heterogeneous data resources over a network. It allows the organization of data from heterogeneous systems into easily accessible logical collections. In combination with the Metadata Catalog, SRB supports location transparency by accessing data objects through



**Fig. 4** AQUAGRID-CODESA-3D hydrogeological application service: 3D visualization of the aquifer domain within the Web browser. The zoom panel on the *right side* shows aquifer transects with contour lines of salt concentration in groundwater

queries on their attributes rather than on their physical locations. In Beltrame et al. (2006) has been demonstrated how the EnginFrame compute-Grid portal can be coupled with SRB data-Grid middleware to develop services in a heterogeneous and distributed environment. Within the project framework, an experimental data-Grid infrastructure have been established within the Datacrossing decision support system (<http://datacrossing.crs4.it>) for a two zone federation among CRS4 and the University of Cagliari (Manca et al. 2006). Additional partners can be easily associated to the GRIDA3 federation.

### Web 2.0 and mash-up

The slogan “Web2.0” refers to the second generation of Web-based communities and services, providing to Web developers the capability to realize graphical interfaces with almost the same level of interactivity of desktop ones. Web 2.0 applications have the intrinsic advantages of delivering software as a continually-updated service that gets better the more people use it, consuming and remixing data from multiple sources, while providing their own data and services in a form that allows remixing by others, creating network effects through an *architecture of participation* (Oreilly 2007). Web 2.0 takes advantages by a number of

technology services such as social networking, aggregation services, data mash-ups, etc.

AQUAGRID application integrates Web 2.0 solutions in the form of mash-ups. In web development, a mash-up is a web application that combines data from more than one source into a single integrated tool. An example is the use of cartographic data from Google Maps to add location information to hydrogeochemical data, thereby creating a new and distinct web service that was not originally provided by either source (Fig. 8). In addition, browsers have a complete support to Cascading Style Sheets (CSS), and JavaScript that make concrete the Dynamic HyperText Markup Language (DHTML). The Document Object Model (DOM) tree of current displayed page is completely accessible and can be modified by the command processing performed directly by the browser without server requests for new pages (e.g. texts and colors can change according to the keyboard and mouse inputs, see Fig. 3). The server interaction takes advantage from the Asynchronous JavaScript and XML (AJAX) paradigm. An utility integrated into all the most common browsers lets to create JavaScript function sending requests to services exposed by the same server where the Web page has been downloaded. The response of an AJAX request ranges from a simple URI to a block of XML/HTML, from JSON data to a JavaScript code.

**Fig. 5** AQUAGRID-PHREEQC geochemical application service: Case-study selection (via “field studies” menu) and speciation simulation submission (via “submit” button) for samples of the selected field campaign. *Gray* locators on the map correspond to the

water samples of the selected campaign. Basic metadata (name, date, etc.) are shown in the framed box just below the map when clicking on a locator (*highlighted in red*)

### AQUAGRID application service

AQUAGRID (Lecca et al. 2007) is the subsurface hydrology service of the GRID3 infrastructure (Fig. 1). It integrates simulation tools based on complex non linear models, optimization engines to rank optimal solutions, geo-database for data storage, 2D/3D visualization tools, spatial analysis tools and Google Map mash-ups for data rendering and visualization. The system is open and based both on in-house developed (CODESA-3D, Lecca et al. 2007) and open source (e.g. PEST, Doherty 2002; HYDRO-GEN, Bellin and Rubin 1996; PHREEQC, Parkhurst and Appelo 1999) software already in place. Other open source modules (Web-GIS mapping tools and R statistical analysis library <http://www.r-project.org/>) are being progressively implemented into the system when required.

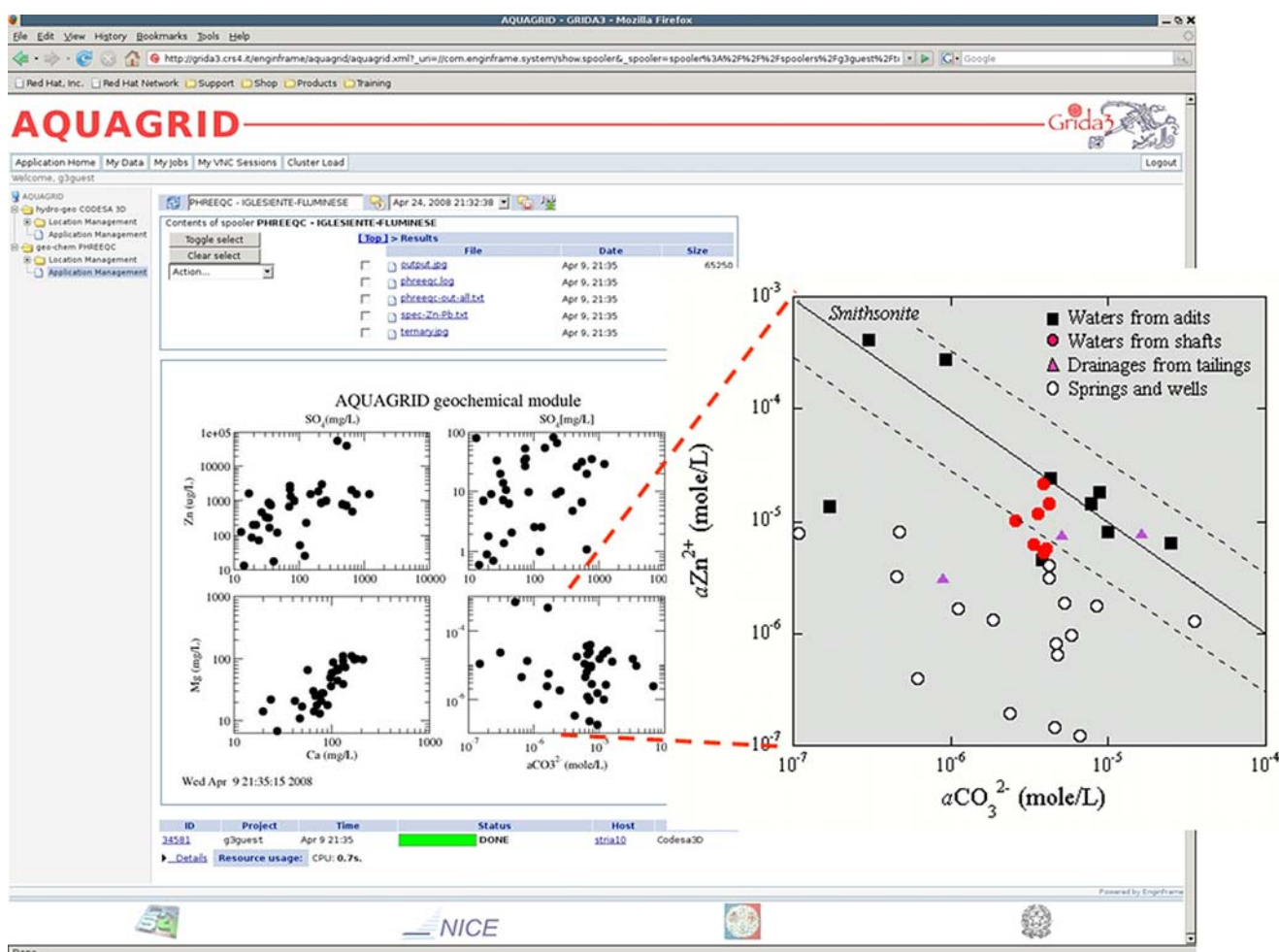
The AQUAGRID service is designed mainly as a front-end user interface for large scale dynamic processing of

hydrogeochemical datasets. The user interface currently provides authorized users with two different Web applications: 1) coupled groundwater flow and solute transport simulation (CODESA3D) and 2) geochemical data analysis and interpretation (PHREEQC). Major fields of application are: planning of cost-effective remediation scenarios, design of optimal aquifer development schemes and monitoring networks, and soil and water risk analysis assessment. Potential end-users of the application are water professionals and decision makers of private and public bodies, without any specific expertise of the underlying physical models and technological background.

### User experience

The use of Web oriented user interfaces (UI) foster the activity of anyway and anytime access to the information.





**Fig. 6** AQUAGRID-PHREEQC geochemical application service: rendering of output files. The application allows to preview graphs (and text files) within the Web browser. The zoom panel on the *right side* shows mine water classification for the selected graph

Since AQUAGRID aims at reaching the largest number of users, not only very skilled ones such as experts and technicians, the main aim of such UI is to minimize the learning curve effort moving from not at all to very skilled users. To design the UI developer has to consider:

- the sequences of transactions the user may perform;
- which transactions each category of users may perform;
- the type of data inserted;
- the type of data returned.

Accordingly, the AQUAGRID user categories have been identified as:

- *guest*: access without password; general browsing only on public areas;
- *basic*: access with password; they have a general view of the provided services; they can use the provided datasets; they don't interact directly with the Grid layer;
- *expert*: access with password; they have a personal workspace; they can provide their own and new data-

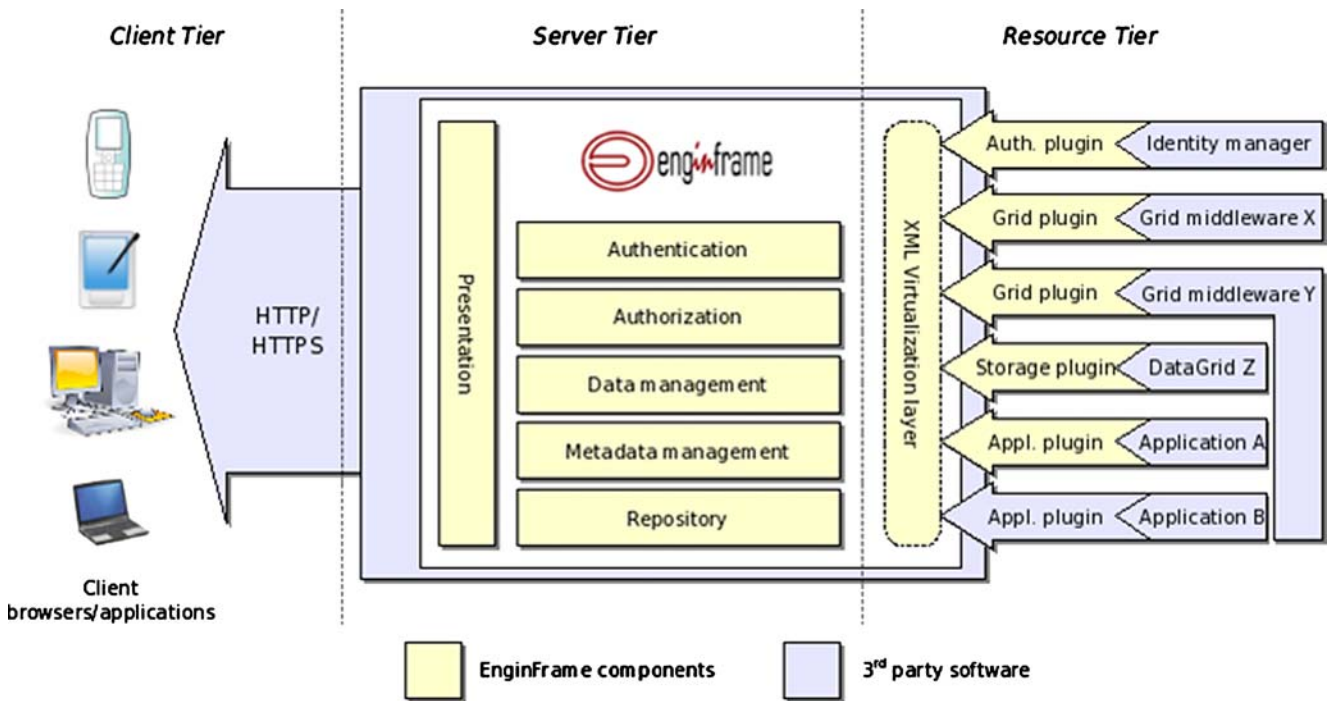
sets, they can submit and monitor jobs and download results;

- *administrator*: superuser; they have the global control of the application.

Since its design the UI development has followed several criteria based on habits of the users, such as perception of information and commands, use of information and commands, security, errors tolerance, niceness. The analysis of usability of UI is driven by test methodologies centered on the final users. The evaluation of the UI is based on a methodology driven by four conditions:

- panel establishment: a representative and heterogeneous group of users;
- scenarios: use cases defining use contexts, aims, interaction with the application;
- evolutionary design: the panel evaluates the application relating to more complex scenarios;
- monitoring: the application shouldn't be static but should change with user needs.





**Fig. 7** Enginframe architectural view. The application services (e.g. AQUAGRID application) are 3rd party software integrated into the system by means of a XML service definition file

**CODESA-3D simulator**

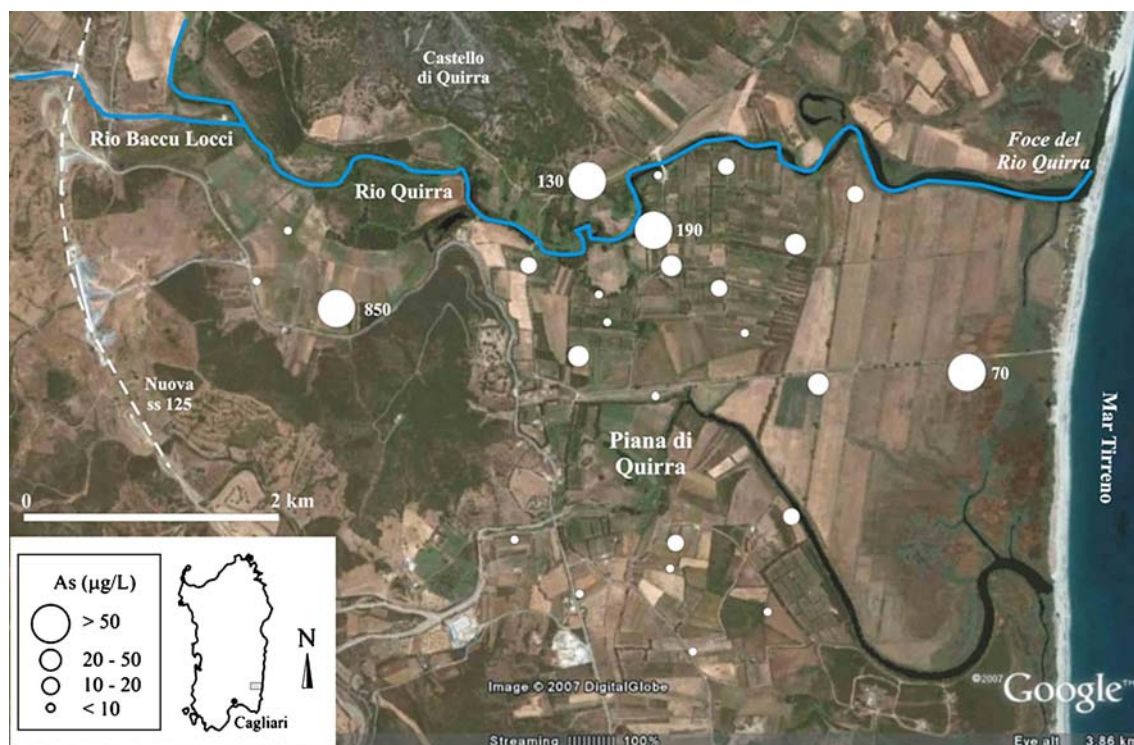
The AQUAGRID hydrogeological service is based on the CODESA-3D computer code (Lecca 2004). The numerical model simulates complex and large density-dependent groundwater flow and contaminant transport problems in aquifer systems. Model forecasts aim to support better policy and decision making for freshwater resource management in coastal zones. Often in groundwater modeling, available field data are not sufficient to adequately characterize either subsurface physical properties and model forcing inputs (e.g. spatial and temporal distribution of rainfall infiltration or aquifer withdrawals). Thus, stochastic simulations, based on Monte Carlo analysis, are used to assess and quantify the impact of input uncertainty on resulting predictions (Lecca and Cau 2006). To circumvent the large computing time required to complete several 3D density-dependent simulations and to post-process model outputs, distributed computing is exploited.

The application (Fig. 3, service *Application Management*) allows registered end-users to run on the infrastructure a set of demonstrative 3D seawater intrusion simulations for two coastal aquifer pilot sites: Oristano-Italy (Lecca and Cau 2004) and Korba-Tunisia (Kerrou et al. 2007). Case-study selection is performed choosing a site locator on the Google Map mash-up. Figures 2–4 illustrate a series of actions performed by end-users when deploying a typical CODESA-3D application. Users are allowed to

perform simulation by clicking the “submit” button (Fig. 3), to online inspect run status (Fig. 2), cluster load, model input and output text files, to download them on the remote computer, and to visualize in the browser a 3D image of the simulated groundwater contamination field (Fig. 4). Further developments (service *New Application*) will allow the user to partly modify input settings of the demonstrative case studies to generate in a dedicated workspace a number of “personalized” application scenarios. In general, end-user will be allowed to adjust physical model parameters, forcing terms and boundary conditions, with the 3D computational mesh kept fixed. Besides the main purpose of the generation of end-user scenarios to forecast system behaviour under natural and anthropogenic stresses, the service will be relevant also for educational purposes.

**PHREEQC simulator**

The AQUAGRID hydrogeochemical service is based on the PHREEQC open source computer program (Parkhurst and Appelo 1999). The program, developed by the United States Geological Survey (USGS), simulates chemical reactions and transport processes in natural or polluted water. PHREEQC is based on equilibrium chemistry of aqueous solutions interacting with minerals, gases, solid solutions, exchangers, and sorption surfaces. It can be used as a speciation program to calculate saturation indices and



**Fig. 8** AQUAGRID-PHREEQC geochemical application service: arsenic concentration in groundwater visualized on a Google Map

distribution of aqueous species. Its extensible chemical database allows application of the reaction, one dimensional transport, and inverse modeling capabilities to almost any chemical reaction that is recognized to influence rain, soil, ground and surface water quality.

The PHREEQC application (Fig. 5, service *Application Management*) allows to run, on a high end Linux cluster of PCs, a number of geochemical simulations based on data collected in 17 field campaigns conducted by the DISTER for a total of 20,000 chemical analyses and 940 water samples. The user performs the selection of one the data collections to be analysed using a Web menu. The selection dynamically loads from the geo-database the coordinates of the corresponding groundwater samples on a Google Map mash-up (same Figure). Each sample, when highlighted by a mouse click, allows to inspect its basic sample metadata (name, sampling date, etc.) in the framed box located just below the map. Finally, the “submit” button launches the simulation of speciation for the whole set of samples. Since the PHREEQC output text file is large and cannot be easily inspected on the browser, the user may optionally choose in the “selected output” menu synthetic results (e.g. saturation index, specific chemical elements) to be summarized in a single text file. After the job completion, user retrieves and downloads input and output results as text files, graphs and maps (Figs. 6 and 8). Further developments will allow exposing results also as Web-GIS thematic maps overlaying basic geographic layers.

The same application (service *New Location*) allows to assemble and run an *ex-novo* geochemical simulation in a dedicated end-user space, by supplying a set of aqueous solution samples via a user-friendly Web form. The form ingests into the application database chemical analyses of water samples. Users may locate the sample directly on the Google map obtaining the corresponding geographical coordinates and viceversa. Additional geographical information is provided by the GeoNames database.

Data concerning field studies are automatically collected from the database and visualized on the map, according to user selection (Fig. 5). The data needed by the computation are extracted and properly formatted for the specific input model. Geochemical data from a new location supplied by end-users are similarly processed.

### Grid-enabled application results

A number of demonstrative case studies have been already deployed into the AQUAGRID portal. To illustrate main portal functionalities through the case study results, the seawater intrusion simulation of the Oristano coastal aquifer (central western Sardinia, Italy) and the geochemical speciation simulation of the groundwater flowing in the Iglesias-Fluminense mine district (southern-western Sardinia, Italy) are described in the following. The complete description of the available case-studies is provided at the

GRIDA3 project wiki pages (<http://grida3.crs4.it/grida3/doku.php?id=activity:or4>) linked to the computational portal (<http://grida3.crs4.it/enginframe/aquagrid/index.xml>).

### The Oristano coastal aquifer (Sardinia, Italy)

The objective of the case study is to improve the understanding of the groundwater degradation mechanism in the complex hydrodynamic environment (diverted upstream surface waters, lagoons and channels) of the Oristano coastal basin (central western Sardinia, Italy) and to support the decision between alternative aquifer exploitation scenarios. Coastal areas experience dramatic demographic and socio-economic development, especially in the Mediterranean area, with growing and conflicting demands on water resources from different economic activities (e.g. tourism and agriculture). This anthropogenic pressure leads to often irreversible degradation of natural resources that severely limit the basis for further development. Two subsequent modeling phases are carried out on the GRIDA3 infrastructure: 1) model parameter calibration against field measurements in an iterative fashion and 2) hydrogeological parameters uncertainty analysis via the Monte Carlo technique. The modeling approach relies entirely on a suite of in-house developed and open-source computational modules.

*Automatic calibration* Performance evaluation of a groundwater model is a complex task due to the large heterogeneity of hydrogeological properties, the scarcity of appropriate field data and the variability of boundary conditions. The procedure is based on the coupling of the CODESA-3D model with a nonlinear optimization model able to find the “best fit” between adjustable model parameter and control data obtained from the field (Lecca 2004; Lecca and Cau 2006). The PEST model-independent tool (Doherty 2002) has been deployed to calibrate hydraulic parameters of the groundwater system against head field data in an iterative fashion controlled by a user-defined exit threshold.

*Monte Carlo analysis* A stochastic modeling technique is used to assess the risk of aquifer salinization, under model parameter uncertainty. The Oristano aquifer consists of two main productive units interbedded by a thin, possibly vanishing, clayey aquitard. Hydrogeological data and previous modeling studies revealed hydraulic interconnection between aquifers. Available data does not enable, however, an adequate characterization of spatial discontinuities in the confining layer, leading us to identify aquitard conductivity as the target parameter to be varied during stochastic simulations. Using a Monte Carlo tech-

nique, a range of aquifer system configurations is explicitly simulated for three different schemes of pumping (phreatic aquifer only, deep aquifer only, and both aquifers simultaneously). Results show that predicting the effect of aquitard spatial discontinuities is essential to quantify the potential risk of groundwater contamination due to different exploitation schemes and to identify aquifer protection zones showing some level of vulnerability to seawater intrusion (Lecca and Cau 2004). Figure 4 shows a typical output visualization by means of an image of the simulated salt concentration map plotted on the 3D aquifer geometry. Aquifer transects are also shown to display concentration profiles in the domain interior.

### The Iglesias-Fluminese district (Sardinia, Italy)

The Iglesias-Fluminese mining district, located in SW Sardinia (Italy), was the prominent area for the exploitation of lead and zinc in Italy. The Pb-Zn deposits were intensely mined in the nineteenth and twentieth centuries with a peak (more than 40 mines in operation) in the 1950-decade. The cessation of mining activity (1995) left large quantities of waste dumps and flotation tailings estimated at about 48 million m<sup>3</sup> for the whole area. Mining activities have a relevant influence on the quantity and quality of water resources in the surrounding environment. Dewatering during exploitation and rebound after the mine closure degrades the quality of groundwater, which is no more suitable for a valuable use.

The impact of past mining activity on the quality of groundwater is assessed, by means of the AQUAGRID application service, through a comparison between the chemistry of mine-influenced waters (adits and mine shafts, drainages from wastes and tailings abandoned on the ground) with the chemistry of springs and wells in the same area at sites relatively far from any mine legacy. In the GRIDA3 project wiki pages, a complete description of each groundwater sample along with the geology and mineralogy of the area is provided. The speciation of metals governs their mobility, bioavailability and toxicity. Knowing metal speciation in contaminated soils, sediments and water is needed to improve the ability of decision makers to assess the risks of metal contamination and develop effective and economic remediation strategies.

From the compute portal, is possible to submit the simulation of speciation for all the catalogued water samples (Fig. 5). Geochemical data processing and interpretation yields high dissolved SO<sub>4</sub>, Zn, Cd and Pb contents in groundwater interacting with mine works (Cidu et al. 2008). In the Iglesias-Fluminese mining district, the thermodynamic equilibria, calculated for the most common Zn and Pb minerals, show that most of the mine samples



are at the equilibrium with smithsonite ( $\text{ZnCO}_3$ ) while few samples approach that with cerussite ( $\text{PbCO}_3$ ) and many are not far from it. All the samples are largely undersaturated with respect to anglesite ( $\text{PbSO}_4$ ). In this carbonate environment, therefore, the Zn content in the mine waters is controlled by the equilibrium with smithsonite, while Pb dissolved content depends on its availability at the different sites. The undersaturation of all waters with respect to cerussite is explained by Pb speciation in solution; the most abundant Pb species in waters circulating in the Cambrian carbonate formations is  $[\text{PbCO}_3]^0$ ; this aqueous species is stable under the observed pH conditions and persists at long distance from Pb sources, constituting a major environmental threat. Figure 6 summarizes some output results as binary graphs visualized in the Web browser, after the job completion. Input and output text files are also available to users for inspection and download.

### Summary and future work

We have presented a Grid-enabled subsurface hydrology application designed to deliver complex environmental simulations via a user-friendly Web portal. The application AQUAGRID is built on top of compute-Grid technologies by means of the EnginFrame portal. Data management, distribution and visualization mechanisms are based on the tools provided by the SRB data-Grid middleware and Web-GIS and relational database technologies. Main application results concern a number of ready-to use demonstrative case studies developed by project partners to support decision making in water resources management, and *ex-novo* computational services available to end-users to build their own applications.

Among the lessons learned from this ongoing experience are: the paramount importance of the UI design as a means to provide customized access and control to the resources and to hide all unnecessary complexity of underlying technologies, and the fundamental need to build a dedicated framework to share data, computational models, simulation results and other project resources. Such open platforms can effectively enable the comparison and fully validation of different approaches to solve water contamination problems allowing the integration of different interrelated aspects of the solution (methods and techniques for site characterization, data analysis, monitoring, modeling, optimization and identification of the most efficient and cost-effective remediation technologies) and enacting cross-organizational teamwork and partnership.

Future developments will include macro-application services chaining the computational services of the GRIDA3 infrastructure to deliver solutions customized to the end-user operational requirements (e.g. Web-GIS maps

showing groundwater zones classification according to national and European environmental protection laws and regulations) and the launch of heterogeneous user groups focused on the application to exchange ideas to improve the service by means of user feedbacks.

**Acknowledgements** The work has been partially funded by the GRIDA3 (Italian Ministry of Research and Education MIUR Prog. N. 1433/2006, <http://grida3.crs4.it>), DEGREE (EC FP6 IST-2005-034619, <http://www.eu-degree.org>) and CyberSAR (MIUR PON N. 1575/2004, <http://www.cybersar.com>) projects. The financial support of Sardinian Regional Authorities is also gratefully acknowledged.

### References

- Beagrie N (2007) E-infrastructure strategy for research: final report from the OSI preservation and curation working group. UK Office of Science & Innovation (OSI), retrieved on Sept 1st, 2008 at <http://www.nesc.ac.uk/documents/OSI/preservation.pdf>
- Bellin A, Rubin Y (1996) HYDRO\_GEN: a spatially distributed random field generator for correlated properties. *Stoch Hydrol Hydraul* 10(4):253–278
- Beltrame F, Maggi P, Melato M, Molinari E, Sisto R, Tortorolo L (2006) SRB Data Grid and Compute Grid integration via the EnginFrame Grid Portal, in Technical Report SDSC TR-2006-1, February 2006, pp 15–20
- Cidu R, Biddau R, Fanfani L (2008) Impact of past mining activity on the quality of groundwater in SW Sardinia (Italy). *J Geochem Explor*. doi:10.1016/j.gexplo.2008.02.003
- Doherty J (2002) PEST model-independent parameter estimation, watermark numerical computing
- Droegemeier K, Brewster K, Xue M, Weber D, Gannon D, Plale B, Reed D, Ramakrishnan L, Alameda J, Wilhelmsen R, Baltzer T, Domenico B, Murray D, Wilson A, Clark R, Yalda S, Graves S, Ramachandran R, Rushing J, Joseph E (2005) Service-oriented environments for dynamically interacting with mesoscale weather. *Computing in Science and Engineering*, IEEE Computer Society Press and American Institute of Physics 7(6):12–29
- FP7 DEGREE Project Consortium (2008) Earth science white paper on grids (ES Grid Roadmap Deliverable D6.1.2), retrieved on Sept 1st, 2008 at [http://www.eu-degree.eu/DEGREE/internal-section/wp6/DEGREE-D6.1.2\\_v2.8.pdf/view](http://www.eu-degree.eu/DEGREE/internal-section/wp6/DEGREE-D6.1.2_v2.8.pdf/view)
- Gentzsch W (2007) National grid initiatives: Lessons learned and recommendations, RENC1, Duke & D-Grid, January 21, accessed on Sept 1st, 2008 at [http://www.renci.org/publications/reports/Grid\\_Initiatives\\_Jan\\_22\\_2007\\_final.doc](http://www.renci.org/publications/reports/Grid_Initiatives_Jan_22_2007_final.doc)
- Goulas G, Alefragis P, Housos E (2005) SchedSP: a Grid-based application service provider of scheduling solutions: Research articles, *Softw Pract Exper* 35:1195–1216, Wiley
- Keller G, Seber D, Sinha A, Baru C (2005) The GEOsciences Network (GEON): One step towards, *Geophysical Research Abstracts*, 7, 05726
- Kerrou J, Lecca G, Murgia F, Renard P (2007) Grid-enabled simulation of the impact of exploitation uncertainty on the Seawater intrusion of the Korba Aquifer (Tunisia). Proceedings of IST-Africa Conference, Maputo (Mozambique), May 2007 Maputo (IST-Africa 2007 best paper award)
- Lecca G (2004) Modeling issues: the CODESA-3D code. In: Benavente J et al (ed) *Monitoring, modeling and management of coastal aquifers*. Water Research Institute, University of Granada, Spain. (ISBN 84-699-7903-5)



- Lecca G, Cau P (2004) Stochastic modeling for seawater intrusion risk assessment in exploited coastal aquifers: the Oristano (Sardinia, Italy) case study. In: Miller CT et al (ed) Proceedings of the XV international conference on computational methods in water resources (CMWR XV). Elsevier Science Ltd, Chapel Hill, NC. June 13–17, 2004
- Lecca G, Cau P (2006) Automatic calibration of a 3D groundwater model applied to the Muravera-Flumendosa coastal aquifer (SE Sardinia, Italy). In: Binning PJ, Engesgaard PK, Dahle HK, Pinder GF, Gray WG (eds) Proceedings of the XVI international conference on computational methods in water resources. Copenhagen, Denmark, June, 2006
- Lecca G, Murgia F, Maggi P, Piras A (2007) AQUAGRID: the subsurface hydrology Grid service of the Sardinian regional Grid infrastructure. In: Gavela R, Marechal B, Barbera R, Ciuffo LN, Mayo R (eds) Proceedings of the Third EELA Conference, pp 357–365
- Manca S, Cau P, Bonomi E, Mazzella A (2006) The datacrossing DSS: a data-GRID based decision support system for groundwater management. Proceedings the 2nd IEEE International Conference on e-Science and Grid Computing, Dec. 4–6, 2006, Amsterdam, Netherlands
- Oreilly T (2007) What is Web 2.0: Design patterns and business models for the next generation of software. Communications & Strategies, No. 1, p 17, First Quarter 2007
- Parkhurst DL, Appelo CAJ (1999) User's guide to PHREEQC (version 2)-A computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations: U.S. Geological Survey Water-Resources Investigations Report 99-4259, pp 312
- Pouchard L, Cinquini L, Strand G (2003) The earth system grid discovery and semantic web technologies, proceedings of the workshop on semantic web technologies for searching and retrieving scientific data. Sanibel Island, Florida, October 20, 2003. In: Ashish N (ed), NASA Ames Research Center and Carole Goble, University of Manchester
- Voorsluys W, Araújo E, Cirne W, Galvão CO, Souza EP, Cavalcanti EP (2007) Fostering collaboration to better manage water resources. Concurrency and Computation: Practice & Experience, Workshop on Grid Computing Portals (GCE 2005) 19(12):1609–1620