# Grid-based Environmental Risk Analysis System

**Thilo Ernst, Andreas Hoheisel, and Steffen Unger**

*Institute for Computer Architecture and Software Technology (Fraunhofer FIRST), Berlin, Germany (steffen.unger@first.fraunhofer.de)*

**Abstract:** A computer-supported system for analyzing and managing pollutant-related environmental risks is being developed in the ERAMAS (Environmental Risk Analysis and Management System) project. The objective is to use sophisticated simulation programs to forecast and evaluate the dispersion of carcinogenic and chemically toxic substances in the atmosphere, the soil and the groundwater and to calculate the risk they pose to human beings. The system is being implemented using the technology of the Fraunhofer Resource Grid (FhRG). This technology facilitates coupling of the various project partners' local hardware and software components via the Internet, thus making the system extremely flexible. Depending on the problem in hand, it can select the most suitable models and system information and thus give optimal consideration to the respective constraints, whether in terms of the concrete area affected or the type of substance released. The design of ERAMAS as a distributed system also guarantees that it is dynamically adapted to the current state of the art because the different simulation models are monitored by their respective developers. ERAMAS provides lightweight web-based user interfaces (also suitable for low-bandwidth clients such as mobile devices) implemented using the VirtualLab platform developed in collaboration with DLR (German Aerospace Center).

***Keywords:*** Simulation Models; Environmental Risk analysis; Grid Computing; e-science Platform

## 1. INTRODUCTION

ERAMAS is an acronym for the **Environmental Risk Analysis and Management System** being developed by Fraunhofer FIRST in collaboration with *Ingenieurbüro Beger für Umweltanalyse und Forschung* and the *Dresdner Grundwasser Consulting GmbH* in Germany [ERAMAS, 2004]. ERAMAS is a simulation-based analysis framework for risks caused by chemically toxic or carcinogenic substances released during accidents in industrial installations, the transport of dangerous goods or by terrorist attacks. It is designed to be applicable both for real-time emergency management and for risk mitigation activities such as simulation-aided studies concerning the design of approval procedures or emergency plans.

Figure 1 shows an overview of the simulation models involved in ERAMAS regarding the different transportation paths in the atmosphere and the soil. In the environmental simulation domain, this is a rather typical scenario – nevertheless, with the software technology widely used in this domain today, integrating and coupling even a small number of heterogeneous simulation models and making them available to technically unskilled users amounts to a very complex and time-consuming effort.

In ERAMAS, this task is made much easier by a Grid-based component model which allows the integration of legacy applications without making changes to their internals, and by a Petri-net-based workflow model coordinating the collaboration of the software components involved in the simulation (Figure 3).

## 2. IMPLEMENTED SIMULATION MODELS

The analysis and forecast components employed in ERAMAS are based on approved simulation packages. The following components are available:

- A *diagnostic windfield model (dwm)* that calculates three-dimensional realistic wind fields considering topography from sparse input data, e.g., measurements from a weather station such as wind speed and direction, temperature, stability class, etc.
- Several *source modules* for the simulation of the pollutant emission through a chimney, jet release, line source, or evaporation from a puddle. To consider uncertainties in the input data, we use a Monte Carlo Simulation for these source modules.

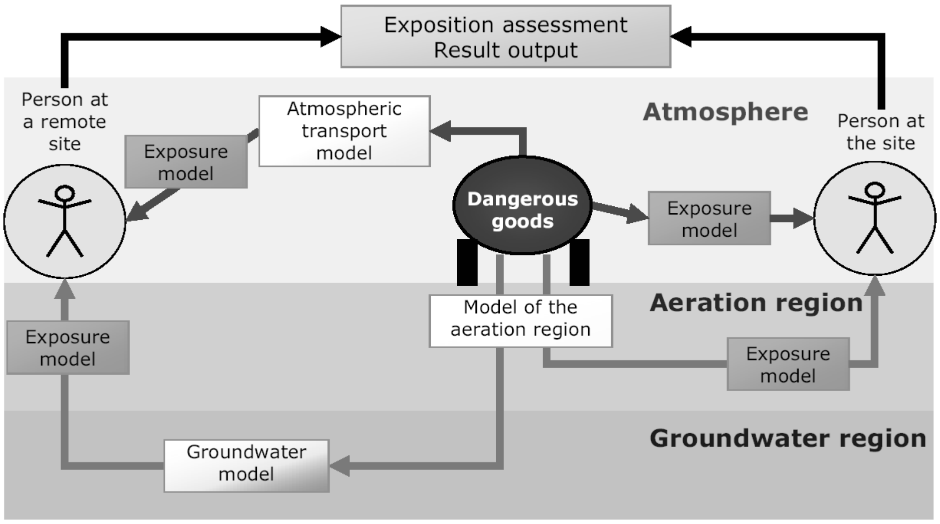


**Figure 1.** ERAMAS integrates a variety of physical simulation models in order to calculate the transport of carcinogenic and chemically toxic substances in the atmosphere, the ground, and the exposure to humans.

- A *Lagrange model* for atmospheric pollutant transport simulation. From the emissions and the three-dimensional wind field it calculates the spatial and temporal distribution of the pollutants. As our Grid framework is not optimized for tightly coupled applications, this model uses MPI for parallel execution on a Linux cluster.
- A one-dimensional simulation model for the calculation of the pollutant transport in the aeration region (*HYDRUS*).
- A three-dimensional simulation model for the calculation of the pollutant transport in the water saturated ground (*MODFLOW*).
- Several models for the simulation of the pollutant exposure to the human being for different kinds of ingestion and inhalation paths.
- Modules for the risk analysis as a function of the land usage and the resulting distribution of pollutants on the ground.
- An open-source geographical information system (*GRASS*)
- A MySQL database server

More background on ERAMAS is available in German in [ERAMAS, 2004, and Unger et al., 2003].

## 3. ERAMAS ARCHITECTURE

ERAMAS is being developed using the technology of the *Fraunhofer Resource Grid (FhRG)* that simplifies the coupling of the heterogeneously distributed software, hardware, and data resources (see Figure 2). The Fraunhofer Resource Grid is a Grid initiative of five Fraunhofer institutes funded by the German federal ministry of education and research with the main objective to develop and to implement a stable and robust Grid infrastructure within the Fraunhofer-Gesellschaft, to integrate available resources, and to provide internal and external users with an easy-to-use interface for controlling distributed applications and services on the Grid [FhRG, 2004]. The component environment supports loosely coupled software components where each software component represents an executable file (command-line application) that reads input files and writes output files. (This smallest unit of execution is also referred to as an *atomic job*). Communication between components mainly takes place via file transfer (GridFTP). Arbitrary legacy code can be integrated easily using shell scripts to encapsulate programs or fixed combinations of them.

Up to now, the FhRG architecture does not directly support tight coupling schemes like CORBA, MPI, or HLA. However, tightly coupled applications can be included as a whole, forming one atomic job each. The FhRG Petri-net-based workflow model and the Grid Job Handler implementing it, is described more detailed in Section 5.2 and [Hoheisel, 2004].

ERAMAS uses the VirtualLab platform [Ernst et al., 2003] for providing a convenient portal-based

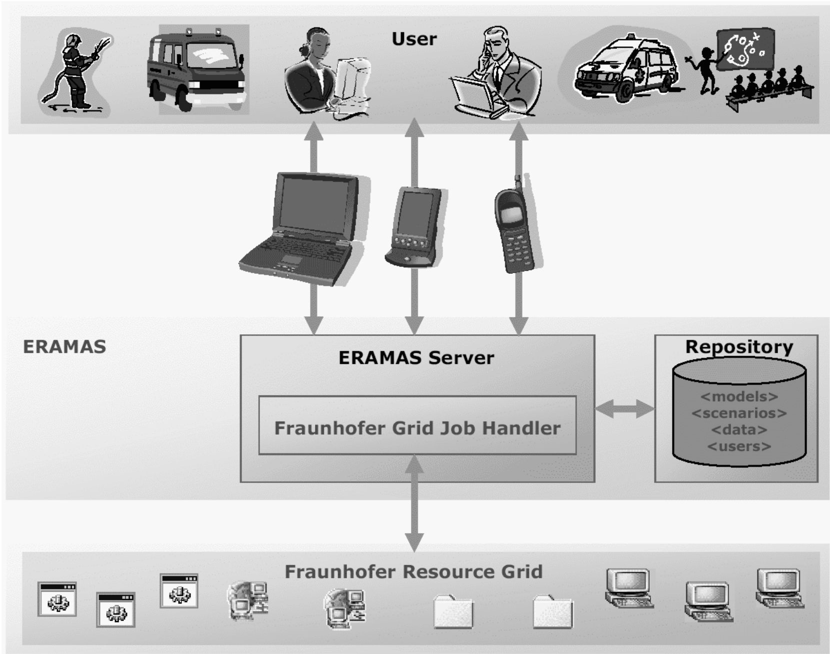


**Figure 2.** The ERAMAS architecture. The user connects to the ERAMAS server by HTTP. The ERAMAS Server uses the Fraunhofer Grid Job Handler in order to enact the ERAMAS workflow.

web user interface. The genericity of this user interface mechanism extends the ease of integration provided by the underlying FhRG Grid component model up to the UI level.

Currently, ERAMAS has the status of a demonstration prototype. We plan to use it as a production system in the near future.

## 4. WHY DO IT ON THE GRID?

ERAMAS is a system with considerable resource demands which arise not only from the inner complexity of its components, but also from complex workflows and usage scenarios in which a substantial number of component instances need to be executed – for instance in parameter studies. Such a system cannot be expected to run on a single workstation; the need to exploit parallel and distributed computing techniques is obvious. However, the primary advantage of using Grid technology in ERAMAS is not mainly the performance gain from being able to use additional resources, but rather the organizational advantages in building and maintaining a distributed, highly heterogeneous simulation system. We use the Grid to organize the workflow of coupled simulations and to provide uniform access to a wide variety of hardware and software (including data) resources.

The component abstractions offered by the Fraunhofer Resource Grid make coupling of a wide range of models (and data sources) very easy – detailed knowledge of the internals of the components is no longer needed. In the long term, we are convinced that the Grid will provide the perfect exchange and transfer medium for simulation models and scientific software (and data) in general – once the Grid technology is sufficiently mature, standardized and openly available, it can be expected to become widely deployed, at least throughout the research and technology sector. For first-time or casual users, invoking a program directly on the Grid (most often through a web portal) will be the quickest and most convenient way to use it. Provided that it will get easy to publish software on the Grid as well, this can enable a much smoother and unobstructed flow of knowledge manifested in scientific software into its potential application domains in industry, engineering and government than we see it today. The consequences of this transition must be considered not only on the technical level. Lowering the barrier-to-entry both on the "producer" and on the "consumer" side, makes lots of application scenarios realistic that today are simply not economically viable: For instance, a small, specialized company will be able to work with a large selection of programs that would have been prohibitively expensive to purchase "rented" as application services on the Grid; a research institute can realistically offer a specialized software package as a Grid service that would otherwise never get used outside the department it was developed in. Of course this era of broad e-Science [Ernst and Wauer, 2003]

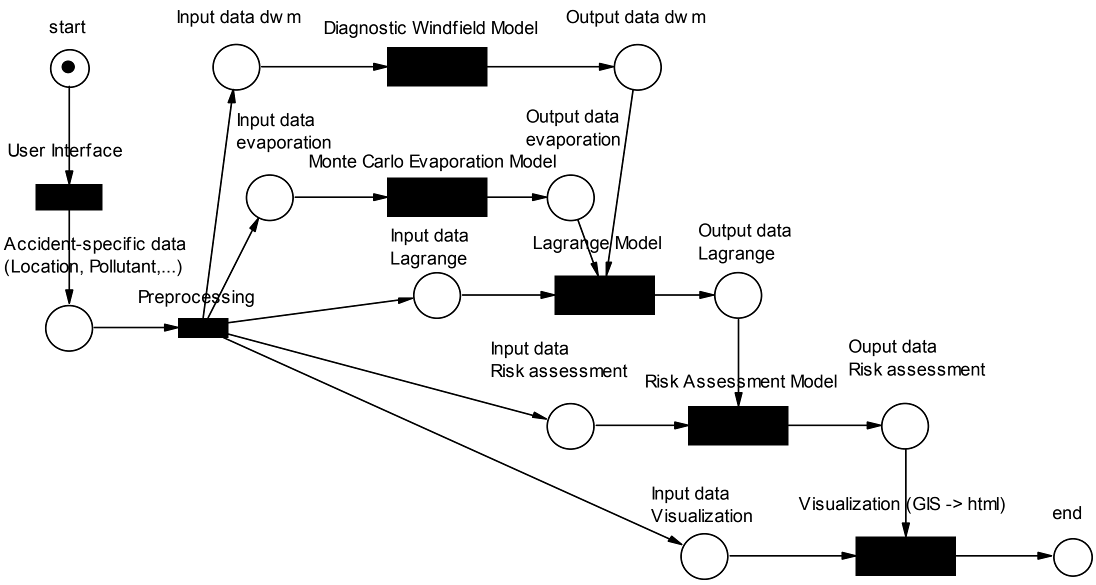


**Figure 3.** Petri net describing the Grid job for simulating atmospheric pollutant transport in the atmosphere.

will not arrive all by itself – ERAMAS represents a step in this direction. In other words, ERAMAS can be viewed as introducing Grid- and web-based e-science methods into the environmental simulation and risk management community and to develop and deploy a dedicated platform for the purpose.

With respect to the risk management-related project goals it is also mandatory to take care of quality-of-service aspects. However, there are many open questions in this area today, so the ERAMAS prototype does not provide such a guarantee right now. On the positive side, the Grid technologies it relies on are in our view an ideal base for addressing these issues.

## 5. REALIZATION WITHIN THE FRAUNHOFER RESOURCE GRID

**Software Component Framework.** Most of the software components that are integrated within ERAMAS are provided as legacy code. In order to make these components available to the Grid architecture they have to comply with the software component model [Hoheisel, 2002]. This component model is rather simple – it just specifies mandatory command line parameters and defines the input and output files that are processed by the component. Furthermore, the standard output, standard error and standard input of the component can be used by the Grid architecture. Arbitrary executables can easily be wrapped using shell scripts in order to enforce compliance with the software component model.

Following this, the software component is packed together with its static data and the encapsulating shell script into a tar.gz archive that is stored somewhere on the Grid, so the software component can be deployed automatically on a suitable hardware resource on demand (a planned extension will also allow replicated component storage in order to improve reliability and availability). Each software component is described separately by an XML-based resource description (GResourceDL) in order to support task-mapping mechanisms for problem-solving environments and resource discovery with regard to the dependencies between Grid resources. The FhRG component framework supports loosely coupled communication between the software components. This communication mainly takes place by transferring the corresponding input and output files via GridFTP. The main advantage of this approach is that it does not require any modification or recompilation of the software components, so existing legacy code can be included easily in the framework ("black-box integration"). The coordination and synchronization of the execution of multiple application tasks is done by the so-called *Grid Job Handler* that acts as a higher-level service on top of existing Grid middleware.

**Workflow Orchestration and Enactment.** Our approach is centered on a Petri-net-based workflow model that allows the graphical definition of arbitrary workflows with only few basic graph

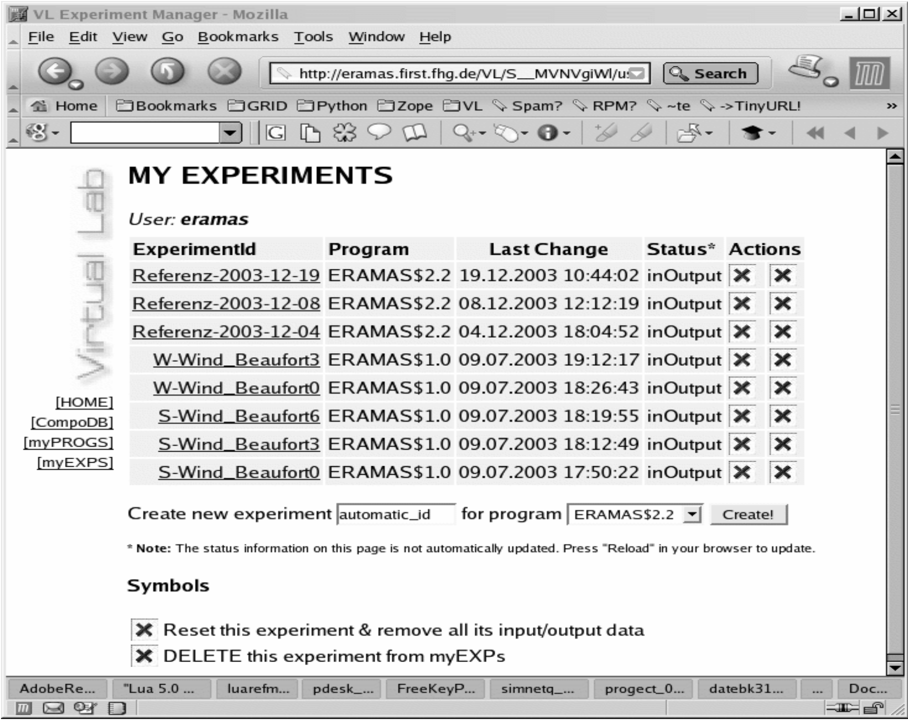


**Figure 4.** Within VirtualLab, each simulation of a specific scenario is managed as a virtual experiment. The user can access his data from past simulation runs and create new experiments using a standard web browser.

elements – just by connecting data and software components. The output of the workflow orchestration process is an XML-based *Grid Job Definition Language (GJobDL)* document, which defines the Grid job. The GJobDL description of a Grid job contains the resource descriptions of the basic resources that are required to define the Grid job and the model of the Grid job workflow and the dataflow using the concept of Petri nets [Petri, 1962]. The GJobDL document can be stored as a file or be transmitted directly to the Grid Job Handler Web Service in order to enact the workflow. Within ERAMAS we use predefined Grid job templates, so the user does not have to take care of these coupling mechanisms. An example Petri net that represents the workflow of a sub part of the ERAMAS simulation is displayed in Figure 3. For further details about the Petri net approach of the Fraunhofer Resource Grid refer to [Hoheisel and Der, 2003].

The Grid Job Handler is responsible for the execution of each Grid job on a set of suitable hardware resources. Therefore, the Grid Job Handler parses the Grid job description, resolves the dependencies between the Grid resources, and searches for sets of hardware resources that fulfill the requirements of each software component. A meta scheduler is used to select the best-suited hardware resource of each set of matching hardware resources. The Grid Job Handler maps the resulting atomic jobs to the Globus Resource Specification Language (RSL) [Globus, 2000] and submits them via GRAM to the corresponding Grid nodes. The Grid Job Handler itself is deployed as a Web Service with possibilities to create, run and monitor Grid jobs remotely.

## 6. WEB ACCESS TO ERAMAS VIA VIRTUALLAB

The simulation components integrated in ERAMAS are pure command-line applications, i.e., they have no graphical user interface. As specialized simulators usually originate from the research sector, this can be considered normal, but it severely conflicts with the goals of and application scenarios envisioned for ERAMAS, which call for strong support of the less technically skilled user, e.g. users from on-site emergency response teams. We bridged this gap by relying on the VirtualLab web platform architected by one of the authors at the German Aerospace Center (DLR), and now further developed in an ongoing collaboration DLR-Fraunhofer FIRST. VirtualLab [Ernst et al., 2003] can be characterized as a pre-Grid Science Portal which allows making scientific programs available for direct on-line execution as web applications in a set of domain-specific virtual laboratories.

Targeted at facilitating technology transfer and scientific exchange, the system is consistently focused on reducing barriers-to-entry for users (but also component authors) so they no longer need to act as system administrators or developers when the goal is just to use some piece of scientific software or offer it for on-line execution, respectively.

VirtualLab contains a subsystem for dynamically generating flexible and easy-to-use web user interfaces for command-line applications from abstract descriptions of their input datasets. Together with its generic web portal features (protected user areas persistently storing simulation runs, integrated documentation management, web-based administration), VirtualLab is thus able to provide a powerful web access layer for ERAMAS.

Figure 4 shows the VirtualLab-based web user interface. The user provides input data for the input files that are transferred from the VirtualLab web server to a Grid gateway node. Also, a Grid job description (including workflow description) is generated and submitted to the Grid Job Handler Web Service, which takes care of its execution. Progress can be monitored from the user's browser through a Java applet. After completion of the Grid job, the user can access the resulting data using his web browser.

## 7. CONCLUSIONS AND FURTHER WORK

ERAMAS is a work in progress, but already now demonstrates the vast potential Grid and web-based e-science methods will have when routinely available to the broader scientific computing community. For the near future, it is planned to migrate the FhRG and ERAMAS to Globus Toolkit V3. As our component model does not require the internals of components to be modified, we expect this to proceed rather smoothly. In the mid term, we will further develop ERAMAS and FhRG and will more closely integrate them with central parts of VirtualLab, with the goal of building a Grid and web-based platform for broad e-science. The application domain of environmental simulation and risk management will remain a main focus of our prototyping and case study work.

## 8. ACKNOWLEDGEMENTS

The project is funded by the *Arbeitsgemeinschaft industrieller Forschungseinrichtungen "Otto von Guericke" (AiF)* in the program "Innovationskompetenz mittelständischer Unternehmen" (PRO INNO).

## 9. REFERENCES

ERAMAS homepage, http://www.eramas.de/, 2004.

Ernst, T., Rother, T., Schreier, F., Wauer, J., and Balzer, W., DLR's Virtual Lab – Scientific Software Just A Mouse Click Away. *IEEE CiSE,* 5 (1), 2003.

Ernst, T., Wauer, J., Grid Content Evolution and Management – Enabling e-Science for the Rest of Us. In: *Proceedings of the Cracow Grid Workshop '03*, Cracow, Poland, 2003.

Fraunhofer Resource Grid homepage, http://www.fhrg.fhg.de/, 2004.

Hoheisel, A., Ein Komponentenmodell für Softwarekomponenten des Fraunhofer Resource Grid. Internal report, Fraunhofer FIRST, http://www.andreas-hoheisel.de/docs/FhRGSoftwareComponent.pdf, 2002.

Hoheisel, A., Der, U., An XML-based Framework for Loosely Coupled Applications on Grid Environments. In: Sloot, P.M.A. et al. (eds.), ICCS 2003. *Lecture Notes in Computer Science*, 2657, 245–254, 2003.

Hoheisel, A., User Tools and Languages for Graph-based Grid Workflows. Proc. GGF10-Grid Workflow Workshop (to appear), 2004.

Petri, C. A., Kommunikation mit Automaten. Ph.D. dissertation, Bonn, 1962.

The Globus Project, The Globus Resource Specification Language RSL v.1.0, http://www-fp.globus.org/gram/rsl_spec1.html, 2000.

Unger, S., Hoheisel, A., Beger, E., Beims, U., ERAMAS – Analyse- und Managementsystem von schadstoffbedingten Umweltrisiken, *Technische Überwachung,* 44 (4), Springer-VDI-Verlag Düsseldorf, 46–49, 2003.