# The GEMSS Grid: An Evolving HPC Environment for Medical Applications

## Authors

D.M. Jones, J. W. Fenner, G. Berti, F. Kruggel, R. A. Mehrem, W. Backfrieder, R. Moore, A. Geltmeier

## Affiliations

*D. M. Jones, J. W. Fenner, R. A. Mehrem:*
Medical Physics & Clinical Engineering, The University of Sheffield, Sheffield, UK

*G. Berti:*
C&C Research Laboratories, NEC Europe Ltd., Sankt Augustin, Germany

*F. Kruggel:*
Max-Planck-Institut für neuropsychologische Forschung, Leipzig, Germany

*W. Backfrieder:*
Institut für Biomedizinische Technik und Physik, Universität Wien, Wien, Austria

*R. Moore:*
IDAC Ireland Ltd., Cork, Ireland

*A. Geltmeier:*
ASD Advanced Simulation and Design GmbH, Rostock, Germany

## Corresponding author

*name:* D. M. Jones
*telephone:* +44 1142712234
*fax:* +44 1142713403
*email:* d.jones@sheffield.ac.uk
*address:* Medical Physics & Clinical Engineering, Room I100 Royal Hallamshire Hospital, Glossop Road, Sheffield S10 2JF, United Kingdom

# Abstract

*Objectives:* Grid Enabled Medical Simulation Services (GEMSS) aims to provide a health-computing Grid platform suited to the provision of accessible compute-intensive health applications. The viability and performance of the Grid resource is to be tested using a series of six applications, intended to highlight strengths and weaknesses of the Grid infrastructure.

*Methods:* The six applications are diverse and cover surgical planning, medical image processing and reconstruction, bio-fluids simulation and radiosurgery treatment planning. These are being hosted on High Performance Computing facilities in Austria and Germany with middleware built on top of existing Grid and Web technologies. This approach maintains compliance with current standards and ensures future extensibility and interoperability. The GEMSS middleware includes features that support quality of service aspects, business and security. The latter is particularly important within the legal context of European-wide health computing, and secure operation designed to safeguard patient data and confidentiality must be demonstrated. A spectrum of end users has started to evaluate the viability of the Grid from the corporate, health, research and small business sectors and exercise the computing resource in a realistic manner.

*Results:* The applications have been adapted to accommodate the requirements of the Grid infrastructure and demonstrated successful execution. Despite the constraints placed upon compute provision by the health sector, it appears that a secure yet accessible and flexible Grid resource is a viable proposition.

*Conclusions:* GEMSS demonstrates the possibility of much wider access to HPC facilities for users of healthcare applications without compromising security, accessibility and flexibility.

## Keywords

Computer Simulation [MESH unique ID: D003198]
Image Processing, Computer-Assisted [MESH unique ID: D007091]
Computer Communication Networks [MESH unique ID: D003195]
Computer Security [MESH unique ID: D016494]

# 1 Introduction

Grid-enabled Medical Simulation Services (GEMSS) is a Grid computing viability study [1] concerned with the provision of advanced health simulation and image processing applications running as services on a Grid platform. The project will create an environment in which computationally demanding tools relevant to the health sector can be made accessible to a wide spectrum of users. Accessibility is critical, since high-powered computational tools are of little use if they are not accessible to those who would use them. The GEMSS tools are aimed at users within many different sectors of the healthcare system, including biomechanics laboratories, hospital surgery and radiology units and designers of medical devices in industry.

GEMSS aims to provide a transparently accessible health computing resource suited to solving problems of large magnitude. The viability of this approach will be evaluated through six diverse medical applications, including maxillo-facial surgery planning, neuro-surgery support, medical image reconstruction, radiosurgery planning and fluid simulation of the airways and cardiovascular system. The developing GEMSS Grid infrastructure will include robust error handling, a mechanism for paying the grid provider in accordance with use of the resource, and recognition of data security and legal issues.

# 2 A GEMSS Application

A GEMSS compliant application must possess certain characteristics, notably:

- The application must operate in a client/server configuration.
- Data must be separated from the application software so that only data is transferred to/from the Grid and the application problem runs remotely as a batch job.
- The application must be compliant with the GEMSS security model.
- The application must support the GEMSS integral business model.
- The application must be compliant with the GEMSS Quality of Service model.

In addition to these attributes, the Grid-enabled applications are intended to provide an accessible computing resource, which includes unobtrusive connection of client to Grid server and intuitive graphical user interfaces.

## *2.1 The Grid Client*

At its simplest, the Grid client is an internet-enabled personal computer on which software has been installed to allow communication with Grid servers through the GEMSS middleware. Security aspects are foremost, and a validated security infrastructure is required before GEMSS access will be permitted. The economic aspects of the GEMSS service also mean that the user must demonstrate the authority to pay for use of the Grid resources. Both aspects are relevant to the Quality of

Service (QoS) component that clarifies the contractual responsibilities of client and service provider in running the computational job.

## 2.2 Job Execution and the Grid Server

The GEMSS design supports a three step process to job execution. A business step is used to open accounts and fix payment details. A quality of service negotiation step follows, which defines the terms and conditions of the compute job. Finally, once a contract is in place, the job can be submitted and executed. Fig. 1 illustrates this process.

## 2.3 Grid Services

Services are mechanisms that support the uploading of input data, job execution, simple process monitoring and the download of results. The GEMSS infrastructure is being designed to use SOAP messaging to WSDL defined web services, hosted using an Apache web server and Tomcat AXIS. HTTPS will be initially used to securely transmit these SOAP messages, and the WS-Security standard applied where possible. For large file transfers SFTP will be investigated, as well as SOAP attachments. The client side applications handle the creation of the service input data and visualization of the service output data.

## 2.4 Security Model

Such is the importance of patient confidentiality and data protection in the medical services sector that a robust security model is a necessity. Mechanisms must be in place to prevent data ‘snooping’ and corruption. For developing application code the most important requirements are:

- The use of input data files with a well-defined structure. This permits rigorous screening of incoming files to check that they conform to an acceptable format and do not contain malicious code.
- Preventing unchecked shell commands in the input data. Free access to a shell provides opportunities to subvert the system.
- Job execution should be confined to the execution directory. Jobs should not write to an area outside their run directory - this contains damage from subverted processes.

The full GEMSS infrastructure is designed to provide many other security measures, including a public key infrastructure, X.509 compliance, RSA encryption, explicit authorization for sequential job handling actions and intrusion detection.

# 3. Applications

The six GEMSS applications are deliberately diverse in order that the flexibility of the Grid can be effectively evaluated. For similar reasons, some of the applications utilise custom written code whilst others make use of commercial code. Details of the applications are provided below.

## *3.1 Maxillo-facial surgery simulation*

This application is concerned with pre-operative planning for maxillofacial surgery. The use of a rigid external distraction system for midfacial distraction osteogenesis is a new method to correct for the underdevelopment of the midface, surpassing traditional orthognathic surgical approaches for these patients. The treatment consists of a midfacial osteotomy (*bone cutting*) followed by a halo-based distraction (*pulling*) step. Typically, the surgeon's experience is the only means of estimating the outcome of the treatment. This application aims to improve upon this by modelling the distraction process [2] (with geometry based on CT data) to enable prediction of its outcome and allow the user to 'try out' several treatments *in silico* before selecting the most promising one.

The simulation process begins with filtering of CT data to remove streak artefacts followed by segmentation of different tissue components. 3D visualisation then allows the surgeon to specify cuts and distraction parameters. From these parameters, a geometric model of the surgical operation is generated, following which the facial distraction process can be simulated by an FEM (Finite Element Model). Finally, the results of this simulation are visualized and used to judge the quality of the surgery outcome.

This is a prime example of a compute intensive application suitable for Grid computing. Using real patient CT data, the largest file sizes for the CT scan set and the FEM model are typically 100-200MB. The segmentation, mesh generation and bone cutting pre-processing steps typically require 1-2GB of memory to run. The FEM step is the most computationally demanding one, and may require up to 15GB of memory to run. In the *thick client* scenario it is only this step that is executed on the Grid. The flexibility of the Grid middleware permits alternative scenarios, with the extreme case being a *thin client* in which the processing for all steps, including the interactive ones, is performed on the server. This considerably reduces the computational demands placed on the client system. A geometrically linear solution step took approximately 7min when using 24 AMD Athlon MP 1900+ processors running at 1.6 GHz. An accurate nonlinear simulation takes considerably longer (several hours).

By allowing access to high performance computing through the Grid, the simulation time can be reduced to a level acceptable for clinical implementation, with the potential to improve the outcome of this surgical procedure. Some development of the pre-processing tools, such as the interface used to specify the location of bone cuts, is required in order to make this application suitable for general use by surgeons.

## 3.2 *Quasi real-time Neurosurgery support by non-linear image registration*

This application corrects for the brain shift phenomenon encountered with image-guided surgical planning. The occurrence of surgically induced deformations invalidates positional information about functionally relevant areas acquired from functional MRI (fMRI) data. This problem is addressed by non-linear registration of pre-operative fMR images to intra-operative MRI, or to intra-operative 3D ultrasound data [3].

The data processing chain has been tested using one pre-op anatomical MRI (aMRI) dataset, one pre-op fMRI dataset, and four intra-op aMRI datasets. At the moment, execution time of the image processing chain is approximately 25 minutes on a 2-processor AMD AthlonMP 1.5 GHz machine. To be useful in surgical intervention, the whole neurosurgery registration imaging chain (including image acquisition) needs to be performed in less than 10 minutes. Because near-linear speedup is expected using the parallel implementation of the software, this should be eminently achievable by using multi-processor Grid resources.

To date, a dedicated local cluster with multiple processors has been used, but such clusters are expensive, require effective maintenance, and are often under-worked. A Grid resource offers a much more cost-effective means of achieving a rapid solution, but clearly needs to offer ready availability and high reliability for this application.

## 3.3 *Near real-time Cranial Radiosurgery Simulation*

This application is designed to run Monte Carlo simulations of Gamma Knife® therapy used in Stereotactic Radiosurgery. The Gamma Knife® (supplier: Elekta Ltd, Crawley, UK) uses 201 $^{60}$Co sources to deliver a high radiation dose to brain lesions. At present, the GammaPlan® treatment planning system is used to calculate the radiation dose distribution. Although this provides rapid results, a superior description of the radiation distribution can be obtained using complex, compute-intense Monte Carlo simulations. However this comes at the expense of significantly increased compute times. The improved physics of the Monte Carlo method will be particularly apparent at boundaries between materials with different linear attenuation coefficients and may indicate significant deviation from the GammaPlan® solution in areas that lie close to sensitive structures such as the eyes. The treatment involves the fitting of a stereotactic frame and treatment should occur shortly thereafter. This means that rapid computations of dose distributions are needed. For this application, previously developed 'RAPT' software [4] has been Grid-enabled and acts as a wrapper that adapts the EGS Monte Carlo computer code [5] to the radiosurgery application. It is well suited to massively parallel computation because there is no inter-process communication during calculation, leading to an almost linear speedup with the number of processors. Nevertheless, Monte Carlo simulation to the desired level of clinical accuracy is extremely computationally demanding. In a test case in which the dose distribution within a water phantom was calculated, Monte Carlo calculation using 20 million photons took approximately 1.2 hours on a PC with a single 2.5GHz Pentium processor. For effective treatment planning, results should be obtained within a matter of minutes. Current Grid tests have reduced this to under 15 minutes using 8 processors.

### 3.4 Inhaled Drug Delivery Simulation

In this application, the process of inhaled drug delivery is modelled. Many factors affect this process, such as inhaler design, formulation of the medication, airway geometry, and pharmacokinetics of the drug absorption process. The initial application of this software is in the evaluation of new designs of inhaler or formulations of the medicament. In the future, it may also be possible to perform patient-specific therapy, in which the dosing regime is tailored to the airway geometry of a particular patient [6]. By far the most computationally demanding stage of the simulation is the fluid dynamical analysis of the airflow used to determine the distribution of drug deposition. Because the input flow-rate varies over the course of a breath, a transient analysis involving solution at several hundred time-steps is required. CFX (supplier: ANSYS Inc, Canonsburg, PA, USA) is the computational fluid dynamics solver used to calculate the pattern of airflow within the inhaler and major airways (Fig. 2), and this is coupled to a compartmental model that represents the compliance of the peripheral airways. Unlike application 3.2, instantaneous availability of the Grid resource is not required. A typical requirement for solution time would be several days. This application would therefore benefit from inexpensive computing power, perhaps available at periods of low demand, which would be negotiated through the Quality of Service component of the GEMSS infrastructure. At present, this application requires expert CFX knowledge for successful execution of the fluid dynamics analysis. In order to improve accessibility to non-expert users, a more user- friendly interface is being developed using EASA (supplier: AEA Technology plc, Harwell, UK). The EASA software acts as a Grid portal that can be used to provide an intuitive interface for the target application. It reduces the expertise required by the user, through the use of pre-prepared parameterised models that represent the most common scenarios found in inhaled drug delivery.

### 3.5 Compartmental modelling approach for the Cardiovascular System

The expertise gained in developing the previous application (3.4) has been used to model the cardiovascular system in a similar manner. The CFX software is again used to model the pattern of fluid flow, but in this case the fluid is blood. A full 3D model of the whole of the circulatory system (which includes many millions of capillaries) would be prohibitively large, and therefore the 3D model is truncated at a computationally tractable point. A compartment model is then coupled to the 3D model and represents the behaviour of the peripheral circulatory system [7]. The compartment system is implemented using a state-space method [8], in which any linear system can be represented using four matrices. By changing the input matrices, the compartment can be instantly re-configured from, say, a simple resistance and compliance in parallel to a more complex branching network. Transient fluid dynamics analysis must be performed because of the time-varying nature of blood flow, and this is therefore the most computationally demanding part of the simulation process. Using this simulation tool, the benefit of various surgical procedures, such as arterial stenting, grafts and valves, can be assessed in terms of the improvement to pattern of blood flow.

### *3.6 SPECT Image reconstruction service*

Visualisation of the distribution of radiopharmaceuticals by Single Photon Emission Computed Tomography (SPECT) provides valuable complementary information to the representation of anatomy from high-resolution imaging modalities such as X-ray, CT and magnetic resonance imaging. A variety of SPECT reconstruction algorithms exist, an example being filtered back-projection (FBP). This finds extensive use in current clinical practice but it is only applicable to single slices. Modern fully 3D iterative reconstruction algorithms [9] provide enhanced image reconstruction for the whole image volume, by considering principal 3D effects of data acquisition, but this comes at the expense of high computational effort (Fig. 3). If N is the length of the image voxel cube, then the number of calculations is of order $N^6$ for full 3D reconstruction compared to order $N^4$ with slice reconstruction. The Grid is an ideal platform for fully 3D reconstruction, and its accessibility offers an opportunity for other imaging centres to avail themselves of this reconstruction technology. It is envisaged that in a fully developed Grid, this might be just one of many image processing resources, available through a common interface, which could be used to aid diagnosis.

## 4 Discussion

It is important that the GEMSS Grid can accommodate a diverse range of applications since this is likely to be representative of relevant applications that are being developed in the wider world. Some of the applications are time critical whilst others are not. Some of the applications rely on third-party commercial software, whilst others are being developed from first principles. Some applications require the transmission of very large input and/or results files across the internet - therefore, a fast, reliable connection may be required. These aspects will be evaluated using a series of test scenarios relevant to each application.

In respect of cost-effective operation, it is advantageous to have a variety of applications that are complementary in terms of their resource needs. Some users will want access to large computational resources at very low cost, but can accept delays and even intermittent availability – applications 3.4 and 3.5 fit into this category. Others need guaranteed access on demand (application 3.2 is a good example here), and in the long term may be prepared to pay a premium for this facility. A resource provider that can address both groups can afford to install enough resources to satisfy their premium-rate, high-priority users, and sell surplus time at low priority to cheap-rate users. This complementarity is core to the Grid, as a service provider that addresses only high-priority users will find it difficult to meet their availability requirements without installing excess resources, which will not be more cost-effective than in-house resources. Through the process of Grid-enabling, computer resources can be used more efficiently, and the availability of clinical applications can be widened beyond the small group of specialist centres with HPC facilities.

# 5 Conclusions

The notion of a Grid platform for health computing is attractive, but its practical implementation imposes constraints in areas such as security and quality of service. In spite of this, there is evidence that the GEMSS architecture can demonstrate a secure, accessible, flexible, reliable, time-critical interface to HPC facilities and ultimately have the potential for significant impact in the health arena.

## Acknowledgements

This project is funded by the European Commission in the Information Society Technologies (IST) Programme under cross-program theme 'Grid Testbeds' - Framework V Project No. IST-2001-37153. The authors are pleased to acknowledge the contribution made by all project partners to ensure the implementation of the GEMSS Grid. We would also like to thank Dr. Hierl, Leipzig, for his support on the maxillo-facial surgery application.

## References

1. The GEMSS project: Grid-enabled medical simulation services. EU IST project number IST-2001-37153; 2002-2005: www.gemss.de
2. Koch RM, Roth SHM, Gross MH, Zimmermann AP, Sailer HF. A framework for facial surgery simulation. In: Proceedings of the 18th spring conference on Computer graphics; ACM Press; 2002. p. 33–42.
3. Tittgemeyer M, Wollny G, Kruggel F. Visualising deformation fields computed by non-linear image registration. Computing and Visualization in Science 2002;5(1):45-51.
4. Gill A, Surridge M, Scielzo G, Felici R, Modesti M, Sardu G. RAPT: A Parallel Radiotherapy Treatment Planning Code. In: Liddell H, Colbrook A, Hertzberger B, Sloot P, editors. High Performance Computing and Networking Europe, Lecture Notes in Computer Science: Springer; 1996. p. 183-193.
5. Electron Gamma Shower (EGS) web page: http://www.slac.stanford.edu/egs/
6. Ley S, Mayer D, Brook BS, van Beek EJR, Heusell CP, Hose DR, Rinck D, Kauczor HU. Radiological imaging as the basis for a simulation software to advance individualised inhalation therapies. Eur Radiol 2001;11(Suppl):216-217.
7. Li JK-J. The Arterial Circulation: Physical Principles and Clinical Applications. Totowa, NJ: Humana Press; 2000.
8. Adby PR. Applied circuit theory: Matrix and computer methods. New York: John Wiley; 1980.
9. Backfrieder W, Forster M, John P, Engelbrecht G, Benkner S. Fully 3D Iterative SPECT Reconstruction in a High Performance Computing Environment. In: Proceedings of the International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences; Las Vegas; 2003. p. 216-221.

## Figure Captions

*Figure 1 caption:*
Three-step process to running a Grid job

*Figure 2 caption:*
Computed flow streamlines in the upper part of the lungs

*Figure 3 caption:*
(a) Filtered Back Projection (FBP) (b) iterative 3D reconstruction algorithm (ML-EM) (c) profiles showing improved contrast for 3D reconstruction algorithm

# Figures

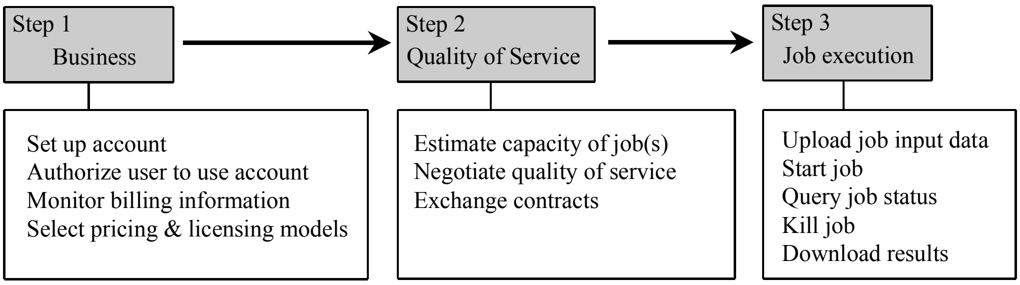

**Figure. 1**

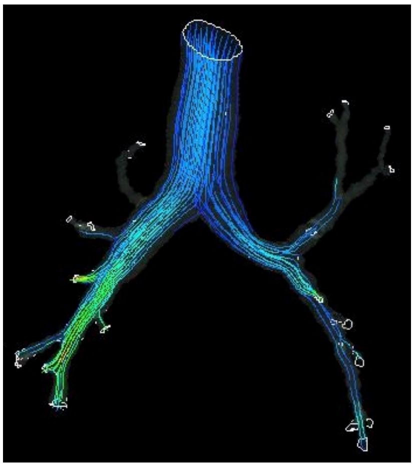

**Figure. 2**

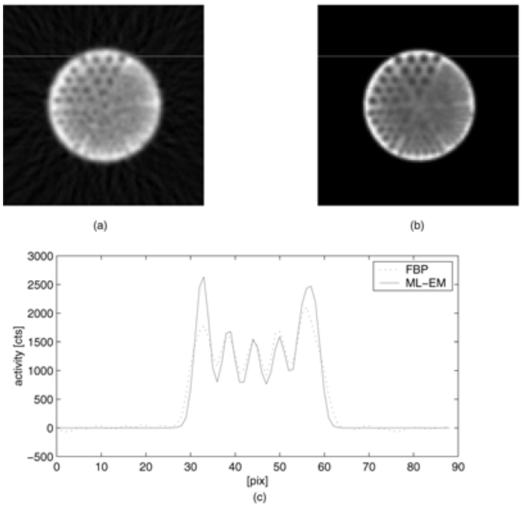

**Figure. 3**