ICGrid: Enabling Intensive Care Medical Research on the EGEE Grid

Harald Gjermundrød [a], Marios D. Dikaiakos [a], Demetrios Zeinalipour-Yazti [a]
George Panayi [b], Theodoros Kyprianou [b]

[a] *Department of Computer Science, University of Cyprus, 1678 Nicosia, Cyprus*
[b] *Intensive Care Unit, Nicosia General Hospital, Nicosia, Cyprus*

Abstract. Healthcare Information Systems are nowadays among the world's largest, fastest-growing and most information intensive industries. Additionally, Intensive Care Units are widely considered as the most technologically advanced environments within a hospital. In such environments, physicians are confronted with the challenge of storing and analyzing terabytes of data that stream in real-time from inpatients to centralized clinical information systems. In this paper we present the system architecture and early experiences from *ICGrid (Intensive Care Grid)*, a novel system framework that enables the seamless integration, correlation and retrieval of clinically interesting episodes across Intensive Care Units, by utilizing the EGEE infrastructure. ICGrid is based on a hybrid architecture that combines i) a heterogeneous set of monitors that sense the inpatients and ii) Grid technology that enables the storage, processing and information sharing task between Intensive Care Units. ICGrid makes it feasible for doctors to utilize the EGEE Infrastructure through simple, intuitive user interfaces, while the infrastructure itself handles the complex task of information replication, fault tolerance and sharing.

Keywords. Intensive Care Medicine, Grid Computing, EGEE

1. Introduction

An Intensive Care Unit (ICU) is an acute care environment deploying multi-disciplinary team skills to resuscitate patients with potentially reversible multi-system dysfunction/failure and in danger of imminent death. The physiological condition of ICU patients is marked by rapidly evolving and frequently life-threatening derangements as well as 'silent' yet important alterations in homeostasis. Therefore, effective and reliable monitoring using multiple, patient-attached sensors is of ultimate importance in order to ensure early diagnosis, timely and informed therapeutic decisions, effective institution of treatment and follow up.

Modern ICU technology supports the continuous monitoring of patients through patient-attached sensors. Data retrieved from these sensors are presented to ICU staff as a continuous real-time flow of screen-displayed numerical values and waveforms. The capability to collect, store, process, and share such data along with the knowledge and the experience of ICU medical staff can bring tremendous benefits to all aspects of Intensive Care Medicine (practice, research, education), in accordance to the emerging trends and challenges of *e-Science* [13,14]. Nevertheless, current ICU facilities provide limited if any capabilities for longer-term storage, off-line analysis, and easy sharing of data. Such capabilities clearly require access to advanced computational, storage, and software resources. Furthermore, the sharing of data between ICUs requires internetworking that takes into account the security and privacy considerations that go with sensitive medical equipment and data.

In recent years, Grids have emerged as wide-scale distributed infrastructures that support the sharing of geographically distributed, heterogeneous computing and storage resources [10]. Grids represent a promising approach for providing Intensive Care Units with easy access to high-end computational and storage resources. Furthermore, Grid middleware provides services that can be used to support and manage the secure, controlled and coordinated sharing of data deployed on Grid resources [2,3]. These capabilities have been demonstrated successfully in computational- and data-grid applications spanning across a wide spectrum of application areas [4,11,12,18].

In this paper, we present the design and implementation of *ICGrid*, a Grid application that enables the retrieval of data from patient-attached medical sensors found in modern Intensive Care Units, the filtering and annotation of these data by ICU medical staff, and the storage and replication of annotated data-sets on the EGEE infrastructure [1]. Also, ICGrid supports the controlled sharing of annotated ICU data-sets between collaborating hospitals; to this end, it provides functionalities for the distributed searching of stored metadata annotations and the retrieval of data-sets through the Grid. The ICGrid infrastructure (application and data) can be utilized in scenarios such as medical education, early diagnosis and for defining early warning systems that identify when human life is in jeopardy. The remainder of this paper is organized as follows. In Section 2 we present the context of modern Intensive Care Units and the prospects and challenges that arise for medical research from the sharing of ICU data-sets; furthermore, we describe the capabilities offered by the EGEE infrastructure. Section 3 introduces the architecture of the ICGrid system and the key issues that were addressed in the ICGrid design. Section 4 discusses the main implementation and deployment challenges faced. We conclude in Section 5, with a summary of our conclusions and future work.

2. Background: The ICGrid Motivation and Context

In this section we familiarize the reader with the required background and the respective terminology for the two main areas addressed by the ICGrid framework: i) Intensive Care Medicine and ii) the EGEE Grid. We highlight the open challenges with regards to information technology in Intensive Care, summarize the state-of-the-art and then outline the benefits of our proposed system framework.

Clinical Environment: An Intensive Care Unit (ICU) is the only environment in clinical medicine where all patients are monitored closely and in detail for extended periods of time, using different types of *Medical Monitoring Devices (MMD)*. An MMD may be defined as a collection of sensors that acquire the patients' physiological parameters and transform them into comprehensible numbers, figures, waveforms, images or sounds. The acquisition is performed by attaching specially designed invasive or non invasive interfaces, such as cables, transducers, catheters etc., to the inpatients body. This in effect enables the extraction of parameters from the vitals parts of inpatients. The MMD can range from *ventilators* to *drug administration pumps*, *blood gas analyzers* and specialized equipment for *medical imaging*. MMDs are nowadays microprocessor-based and have their own screens and data export capabilities.

The acquired measurements can then be utilized in order to: i) take timely and informed therapeutic decisions, ii) ensure early diagnosis and finally iii) have effective institution of treatment and follow up. Note that the patient monitoring is always provided at the bedside, where the results of the acquisition are displayed on the screens in the form of alphanumeric data and waveforms. In order to assist administration, the operation of this setup can be complemented by centrally administered monitors, which connect through LAN-based technologies to a number of distributed monitors, and which consolidate all the information in a concise representation. This way the patients' acquisition data are saved in trends memories for a specific time period.

Technological Shortages: Taking clinical decisions for the ICU patients based on monitoring can be a very demanding and complex task requiring thorough analysis of the clinical data provided: *even the most skilled physicians are often overwhelmed by huge volumes of data, a case that may lead to errors, or may cause some form of life threatening situation* [7]. Consider for instance a physician that monitors a set of medical monitors in order to identify some abnormal condition in the projected state of an inpatient. Although such monitors enable the physician to *react* to alerts by triggering some medical procedure, they cannot provide means for *proactively* exploiting the real-time signals in order to uncover interesting patterns, predict trends and correlate the state of the inpatient to other similar cases. Providing systems that actively learn from previously stored data and suggest diagnosis and prognosis is a problem that, to our knowledge, has been overlooked in previous Intensive Care Medicine research.

Traditionally, medical research is guided by either the concept of patients' similarities (clinical syndromes, groups of patients) or dissimilarities (genetic predisposition and case studies). Clinical practice also involves the application of commonly (globally) accepted diagnostic/therapeutic rules (*evidence-based medicine* [9]) as well as *case-tailored approaches* which can vary from country to country, from hospital to hospital, or even from doctor to doctor within the same hospital. These different approaches in treating similar incidents produce knowledge which, most of the times, remains a personal/local expertise, not documented in detail and not tested against other similar data. Global sharing of this cumulative national/international experience would be an important contribution to clinical Medicine in the sense that one would be able to examine and follow up implementation of and adherence to guidelines as well as to get the benefit of sharing outstanding experience from physicians.

Finally, although a number of dedicated and commercially available information systems have been proposed for use in ICUs [8], which support real-time data acquisition, data validation and storage, analysis of data, reporting and charting of the findings, none of these systems is appropriate in our application context because of the following limitations:

- ICUs lack the required high performance and dedicated storage resources for long-term collection of medical data, and the computational power to perform advanced processing and data-mining on these data.
- Most ICUs lack the qualified information technology personnel that can undertake the complex task of running in-house computations.
- The sharing and collaborative processing of medical data collected by different ICUs raises important privacy, anonymity, information integrity challenges that cannot be addressed by existing commercial ICU information systems.

An estimate of the amount of data that would be generated daily is given in the following scenario. Suppose that each tuple is 50 bytes (it is stored as text) and that there are 100 hospitals with 10 beds each, where each bed has 20 sensors. Assuming that each bed is used for 2 hours per day, the data collected amounts to 6.7055 GB per day. After compression, it would be reduced to 675 MB of data per day, but this number only represents the data from the sensors. Additional information includes metadata, figures, etc.

Grids represent a promising venue for addressing the challenges described above, since all of the aforementioned limitations are common advantages of Computational and Data Grids, with success stories in biomedicine, computational chemistry and high energy physics. It is worth noting that Grids have recently been adopted for the storage and sharing of human biological data [20].

2.2. The EGEE Grid

Currently, Grid computing infrastructures assemble an extensive collection of resources and expertise in production-quality operational support. For instance, the *EGEE* Grid (Enabling Grids for E-sciencE) [1], which is the largest Grid infrastructure in operation, at the time of this writing assembles over 200 sites around the world with more than 30,000 CPU's and about 5PB of storage. Access to EGEE resources is managed by the gLite middleware [2]. gLite comprises a variety of data management services, such as the *Storage Element*, which provides an interface to storage resources available at local Grid sites, the *Logical File Catalog*, which holds information about the location of files and replicas held at different Storage Elements, the *File Transfer Service*, which is responsible for replicating files across different Storage Elements, the *Metadata Grid Application (AMGA)*, which manages metadata, the *Encryption Data Service*, which manages the encryption of data stored in Storage Elements, etc.

Users that want to access a Grid infrastructure and make use of its services need to join a *Virtual Organization* (VO) supported by the infrastructure and its resource providers. A Virtual Organization is a dynamic collection of individuals and/or institutions that share resources in a controlled and mutually agreed fashion [10]. More specifically, EGEE users registered within a particular Virtual Organization obtain security credentials for single Grid sign-on that enable them to obtain controlled access to resources belonging to that particular VO, despite the fact that such resources span different EGEE sites across different countries. End-user resource access is obtained via a *User Interface (UI)* machine, which runs the services required to submit jobs and files to EGEE, to monitor job and resource status, to retrieve data from Storage Elements and Logical File Catalogs, etc.

The resources and services available on EGEE and other Grid infrastructures of similar scale are clearly adequate for storing and managing ICU-related data. In our work, we develop the software tools that are required to establish and operate a Data Grid infrastructure dedicated to Intensive Care Medicine, on top of EGEE.

3. ICGrid Architecture

The envisioned Data Grid infrastructure for Intensive Care Medicine will comprise:

- An *ICU Virtual Organization*, bringing together Intensive Care Units, hospitals, medical schools, medical research institutions, and resource providers willing to work collaboratively in order to promote research and education in Intensive Care Medicine through the sharing of data and knowledge produced inside the ICUs.
- *Grid Sites*, providing adequate storage and metadata facilities to securely store and maintain the huge data-sets collected from the medical sensors found in modern ICUs.
- The *ICGrid framework*, which: i) enables the easy retrieval of data-sets from the medical sensors; ii) supports the local buffering, anonymization and annotation of data by authorized personnel inside the ICU, using computing resources isolated from the Internet; iii) supports the uploading and replication of annotated data-sets to a Storage Element and metadata catalog of a nearby Grid site supporting the ICU-VO, and iv) provides members of the ICU-VO with access to searching capabilities and specialized data-mining applications.

In the remainder of this section, we present the architecture of the ICGrid framework and discuss some key design decisions.

3.1. Architecture Overview

Figure 1, illustrates the architecture of the envisioned ICGrid infrastructure, which comprises a number of different sites that are geographically distributed,

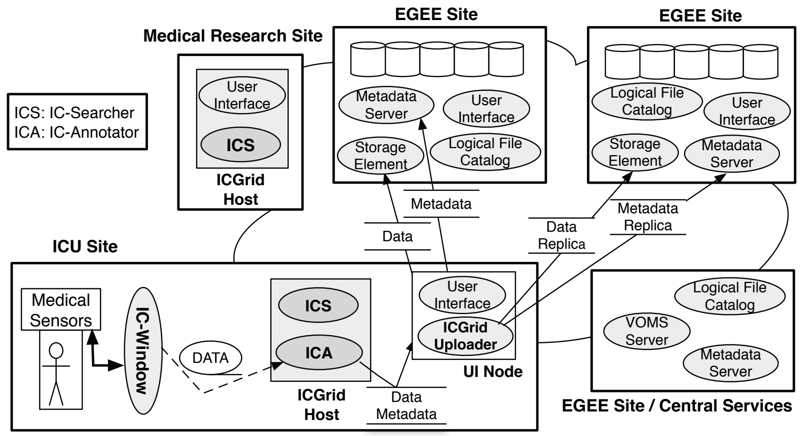

Figure 1. ICGrid System Architecture. White rectangles represent different sites of the infrastructure (each site represents resources of one administrative domain/institution), shaded rectangles represent computer nodes, and shaded ovals depict required Grid services and tools of the ICGrid framework.

belong to different organizations, and are connected through the Internet. In particular: i) *ICU Sites*, which are equipped with the required tools for the retrieval of data from Medical Monitoring Devices, the annotation thereof by expert physicians, and the transfer of collected information to the ICU Grid. A typical ICU Site will be hosted inside the Intensive Care Unit of a hospital that participates to the ICU-VO. ii) *ICGrid Storage Sites*, which support the ICU-VO by contributing their computing facilities (computers, disk farms, tapes) for the storage and processing of data and knowledge published by the ICU sites. iii) *Medical Research Sites*, which participate to the ICU-VO and have access to ICGrid tools for searching and processing data stored throughout the ICGrid Storage Sites. The ICGrid infrastructure is supplemented by one or more *EGEE Sites*, which host the Grid middleware services required for the coordinated operation of the distributed ICGrid resources (VO management, naming services, etc).

The *ICGrid framework* is based on a hybrid architecture that combines a heterogeneous set of monitors that sense the inpatients and three Grid-enabled software tools that support the storage, processing and information sharing tasks: the IC-Window (ICW), the IC-Annotator (ICA) and the IC-Searcher (ICS). The first was developed using Visual Basic while the others were developed in Java and make use of the following services of the gLite middleware of EGEE [2]: i) The *User Interface*, which is the entry point into the Grid infrastructure; ii) The *Storage Element* [19,6], which stores the collected data in data-archive files, replicated to provide fault tolerance, access load sharing, and faster access for the distributed users; iii) The *Logical File Catalog* [16], which is a service that keeps track of the archive files and replicas; iv) The *AMGA* [17] service, which is a metadata storage and retrieval system, complemented with replication, fault-tolerance and access load sharing functionality; and v) the *VOMS* [5] service, which generates the X509 proxy certificates required to authenticate the members of an EGEE Virtual Organization.

The diagram of Figure 1 depicts the acquisition and annotation of parameters of an inpatient at an ICU Site (bottom left) and the transfer of data replicas to two ICGrid Storage Sites. The transfer comprises of the actual sensor data, denoted as *Data*, and the information which is provided by physicians during the annotation phase, denoted as *Metadata*. Note that the *Data* is a selected subset of the acquired signals, those with the highest clinical interest. In particular, the physician that observes the inpatient annotates these signals with metadata in an offline phase. What is considered an interesting incident depends on the subjective opinion of the physician on duty. Consequently, we utilize the notion of a *Clinically Interesting Episode (CIE)* to refer to the captured sensor data along with the metadata that is added by the physician to annotate all the events of interest.

Metadata consists of information about the institution, physician, sensors, patient, intervals of the signals, along with some annotation of the signals. In order to protect patient privacy, information about the inpatient is anonymized. The only information that is collected for the inpatient is height, weight, age, and sex. The metadata is encoded as an XML-document defined by an application-tailored schema. On the other hand, the collected data consists of a set of tab-separated text files, one for each sensor. Each line in these files contains a timestamp, the recorded physical parameter and the state of parameter at the given time-stamp (e.g., if the parameter indicates some alert). All the files that belong to an episode are archived together to form the *Data* archive.

These two files (Data and Metadata) must be transferred to a storage unit that can be accessed by all the authorized and authenticated parties that will be entities of the ICGrid-VO. Thus, the services must satisfy certain security properties, something that is already inherent in the Grid. A Grid user authenticates by presenting an X509 proxy certificate [15] and as a result, the security properties of authentication, authorization, integrity, and non-repudiation are provided. Access to the files that are stored on the Grid is controlled by the readily-available Access Control Lists (ACL).

3.2. Design Decisions

Encryption: Our framework is designed around the concept of open access and sharing of medical knowledge. In order to accommodate the inherent privacy constraints we apply local anonymization procedures at each ICU-site, such that any non-disclosed information is filtered-out, prior to uploading the data to the Grid. This makes encryption redundant to a very large degree. Nevertheless, should such a functionality become a necessity, it could easily be accommodated by the Encrypted Data Storge component that is available by the gLite middleware.
Data Storage: Our current scheme stores data and metadata in plain text and XML format respectively. There is certainly a tradeoff between maximum flexibility and optimality of space utilization. We opted for the former alternative since it facilitates the incorporation of tools that perform searches and do signal processing on the acquired signals. Nevertheless, if storage becomes a concern, we can easily utilize an off-the-shelf relational database or a proprietary binary representation.

Metadata Management: The physical separation of the metadata from the collected data and the utilization of the AMGA Metadata service, provided by the middleware logic, introduces a number of advantages. Firstly, it improves scalability, as the metadata can be distributed across several sites without disrupting the operation of the application logic. Secondly, this distribution certainly improves fault tolerance and optimized access to the indexed information, leading to a simplified software development cycle.

4. ICGrid Implementation

In this section we present the most significant implementation details of the ICGrid framework and procedures for installing and operating the system.

One of the ICGrid framework design criteria was to make the use of the Grid infrastructure completely transparent to the end-user. Therefore we developed a collection of user friendly interfaces which allow physicians to get access to the EGEE resources without any prior knowledge of Grid technologies. In particular, we built three GUI-based tools: *Intensive Care Window (IC-Window)*, *Intensive Care Annotator (IC-Annotator)* and *Intensive Care Searcher (IC-Searcher)*, which enable the task of *acquisition*, *annotation* and *search* respectively.

The IC-Window tool (see Figure 2 (left)) interacts with the a number of patient monitoring devices. We have implemented a full-fledged interface to access the Phillips IntelliVue MP70 medical monitor, one of the most technologically advanced medical monitors on the market. We are also working on providing connectivity to other UDP/IP and RS232-based devices that will supplement our framework with more acquisition capabilities. Our objective is to provide open implementations to the various proprietary and closed protocol of medical monitors, thus contributing at the same time to more general field of medical Informatics.

The IC-Annotator tool (see Figure 2 (right)) assists the end-user with annotating the collected sensor data and uploading this information to the grid. The end-user follows a sequence of steps to annotate interesting incidents that occurred during an episode with the aid of an easy and intuitive GUI. First, the location of the directory where the collected sensor data is stored is specified. Second, the end-user enters the metadata, which is information about the institution, the physician, and minimal patient information. Finally, interesting incidents that occurred during the episode as well as optional HL7 information is entered.

The IC-Searcher tool, is a user interface that supports search over the acquired and stored data. Although IC-Searcher is still at an early stage, it already supports connectivity to the AMGA metadata service for the acquisition of data using free text searches. In the future we will support the automatic identification of similar episodes using high performance timeseries similarity methods executed using grid resources: *Given a real signal from an inpatient, we want to find other signals with a similar temporal movement.* Additionally, we will also support other data mining techniques, such as predicting the future value of a signal and clustering similar inpatient states.

The IC-Annotator and IC-Searcher tools are developed using the Java programming language, which provides platform independence, given that the EGEE

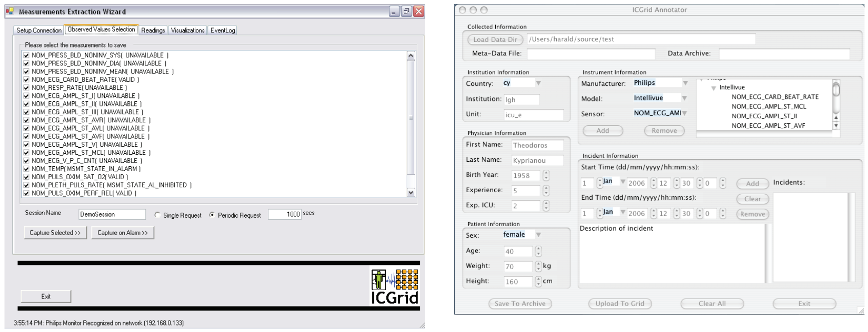

Figure 2.

a) IC-Window: enables physicians to acquire the clinical state of inpatients.
b) IC-Annotator: enables physicians to annotate the respective clinical state.

Grid infrastructure currently only supports a very limited number of Linux operating system distributions. On the contrary, the IC-Searcher is developed with the Visual Basic .NET environment under Windows, as this client is not required to interact with the Grid infrastructure directly. Note that the IC-Window can operate in a stand-alone mode, independently from the rest ICGrid framework, thus creating an invaluable tool for clinical monitoring using real-time graphs and a variety of other data management tools.

Installing the ICGrid framework at an ICU site is a straightforward procedure which is not much different from registering a new EGEE Grid user. In particular, users are issued a private/public key pair that is created and installed on the machine where the IC-Annotator and IC-Windows will operate. This provides a secure channel between the EGEE infrastructure and the ICU-site. This is succeeded by a certificate request from the local Grid certification authority and the issuance of an account on the local Grid entry point. Once the installation procedure is completed, the end-user can interact with the system only with the ICGrid tools. The fact that ICGrid utilizes available services of the EGEE infrastructure once again highlights the importance of Grids as generic infrastructures that can be customized to fit a variety of application scenarios hiding in that way most of the inherent software development and administration complexities.

5. Conclusions and Future Work

In this paper we present ICGrid, a framework that paves the way for Intensive Care Medical Research on the EGEE Grid. ICGrid is based on a hybrid architecture that combines a heterogeneous set of monitors that sense the inpatients and Grid technology that enables the storage, processing and information sharing task between Intensive Care Units. Our preliminary results have so far been extremely encouraging. In the future we plan to extend the framework to use the computational resources provided by the Grid to perform deep-searches in the collected data. Additionally, we want to offer physicians the capability to perform exploratory data analysis in order to maximize insight into the distributed ICGrid data repository, uncover important events, outliers and anomalies.

Acknowledgements

This work was supported in part by the European Union under projects EGEE (#IST-2003-508833) and CoreGRID (#IST-2002-004265).

References

[1] Enabling Grids for E-SciencE project. http://www.eu-egee.org/.
[2] gLite Middleware. http://glite.web.cern.ch/glite/ (accessed June 2006).
[3] Globus project. http://www.globus.org.
[4] WISDOM: Initiative for grid-enabled drug discovery against neglected and emergent diseases. http://wisdom.eu-egee.fr/ (last accessed June 2006).
[5] R. Alfieri, R. Cecchini, V. Ciaschini, L. dellAgnello and?A. Frohner, A. Gianoli, K. Lorentey, and F. Spataro. VOMS, an Authorization System for Virtual Organizations. In *First European Across Grids Conference*, February 2003.
[6] J-P Baud, J. Casey, S. Lemaitre, C. Nicholson, D. Smith, and G. Stewart. LCG Data Management: From EDG to EGEE. In *UK eScience All Hands Meeting*, 2005.
[7] B. Hayes-Roth et al. Guardian: A prototype intelligent agent for intensive care monitoring. *Artificial Intelligence in Medicine*, 4:165–185, 1992.
[8] B.M. Dawant et al. Knowledge-based systems for intelligent patient monitoring and management in critical care environments. In Joseph D. Bronzino, editor, *Biomedical Engineering Handbook*. CRC Press Ltd, 2000.
[9] DL Sackett et al. *Evidence-Based Medicine: How to Practice and Teach EBM*. Churchill Livingstone, 2nd edition, 2000.
[10] I. Foster and C. Kesselman. In I. Foster and C. Kesselman, editors, *The Grid: Blueprint for a New Computing Infrastructure*, chapter 4: Concepts and Architecture, pages 37–64. Elsevier, 2004.
[11] Ian Foster and Carl Kesselman. Scaling System-Level Science: Scientific Exploration and IT Implications. *Computer*, 39(11):31–39, November 2006.
[12] Fabrizio Gagliardi, Bob Jones, Mario Reale, and Stephen Burke. European DataGrid project: Experiences of deploying a large scale testbed for E-science applications. *Lecture Notes in Computer Science*, 2459:480–499, 2002.
[13] Jim Gray. eScience: the Next Decade Will be Interesting! Talk at ETH, Zurich, Switzerland, May 22 2006.
[14] Jim Gray, David T. Liu, Maria A. Nieto-Santisteban, Alexander S. Szalay, Gerd Heber, and David DeWitt. Scientific Data Management in the Coming Decade. *ACM SIGMOD Record*, 34(4):35–41, January 2005.
[15] R. Housley, W. Polk, W. Ford, and D. Solo. Internet X.509 Public Key Infrastructure, RFC 3280. http://www.ietf.org/rfc/rfc3280.txt, April 2002.
[16] P. Kunszt, P. Badino, A. Frohner, G. McCance, K. Nienartowicz, R. Rocha, and D. Rodrigues. Data storage, access and catalogs in gLite. In *Local to Global Data Interoperability - Challenges and Technologies*, pages 166 – 170, June 2005.
[17] N. Santos and B. Koblitz. Distributed Metadata with the AMGA Metadata Catalog. In *Workshop on Next-Generation Distributed Data Management HPDC-15*, June 2006.
[18] Peter M.A. Sloot, Alfredo Tiramo-Ramos, Ilkay Altintas, Marian Bubak, and Charles Boucher. From Molecule to Man: Decision Support in Individualized e-Health. *Computer*, 39(11):40–47, November 2006.
[19] Graeme A. Stewart, David Cameron, Greig A Cowan, and Gavin McCance. Storage and Data Management in EGEE. In *5th Australasian Symposium on Grid Computing and e-Research (AusGrid 2007)*, January 2007.
[20] L. Sun, P. Hu, C. Goh, B. Hamadicharef, E. Ifeachor, et al. Bioprofiling over Grid for eHealthcare. In V. Hernandez, I. Blaquer, and T. Solomonides, editors, *Challenges And Opportunities of Healthgrids: Proceedings of Healthgrid 2006 (Studies in Health Technology and Informatics)*. IOS Press, 2006.