

The Impact of Grid on Health Care Digital Repositories

Flavia Donno
CERN
CH-1211, Geneva 23, Switzerland
flavia.donno@cern.ch

Elisabetta Ronchieri
INFN-CNAF
Viale Berti Pichat 6/2, I-40126 Bologna, Italy
elisabetta.ronchieri@cnaf.infn.it

Abstract

Grid computing has attracted worldwide attention in a variety of applications like Health Care. In this paper we identified the Grid services that could facilitate the integration and interoperation of Health Care data and frameworks world-wide. While many of the current Health Care Grid projects address issues such as data location and description on the Grid and the security aspects, the problems connected to data storage, integrity, preservation and distribution have been neglected. We describe the currently available Grid storage services and protocols that can come in handy when dealing with those problems. We further describe a Grid infrastructure to build a cooperative Health Care environment based on currently available Grid services and a service able to validate it.

1. Introduction

The concept of “Grid computing” was coined almost fifteen years ago by scientists in Argonne National Laboratory and the University of Chicago to denote a distributed computing infrastructure for advanced science and engineering [1]. At that time, Grid computing experienced a slow development and Grid applications were present only in academia funded mainly by U.S.A. funding agencies. The name “Grid” comes from the metaphor of “Electrical Grids” and the idea to get access to a resource (e.g., electricity) by using a plug [1]. However, there is no unified definition of the term “Grid” and the definitions often address different application areas.

Currently, Grid computing has raised interest worldwide in academia, industry, and government with fast development cycles. In particular, a variety of applications ranging from aerospace, astronomy, bioinformatics, chemistry, environment, finance, physics and Health Care have been identified.

Grid computing aims at the definition of a global infrastructure able to share geographically distributed resources (such as data, storage, computers, software, tools, applications, instruments, and networks) securely, at anytime, and anywhere. Great efforts, huge funding and resources have been made available

through national, regional and international initiatives aiming at providing Grid infrastructures, Grid core technologies, Grid middleware and Grid applications. The Grid software layers are shown in Figure 1. The integration of Grid and Health Care has created a new area called Health Grid [2] which is having a great impact on almost every aspect of Health Care, from diagnosis, treatment, primary/acute care to social services.

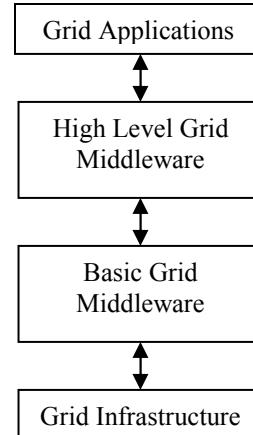


Figure 1: Grid Layers.

Grid technologies [3, 4] are being introduced in many Health Care areas including medical applications [5]. The Grid builds an infrastructure that provides resources to medical personnel who access them following the necessary policies in Health Care. These resources may include computational resources, storage, equipment (e.g., scanners), or human resources (specialists). The access to a medical resource could be given to the clinicians when entering a case into the Grid system. Health Care specific policies (such as privacy or authorization) are enforced according to a service agreement [6].

This paper is organized as follows. Section 2 describes some Health Care case studies, whilst Section 3 focuses on Health Care requirements. Section 4 reports on related works. Section 5 details the Grid storage technology, followed by a description

of existing storage protocols in Section 6. Section 7 details how a Grid infrastructure can serve Health Care cases. Section 8 describes a system able to validate the solution. Finally, Section 9 reports conclusions and future work.

2. Health Care case studies

In what follows we list a set of Health Care cases where Grid computing can be used to help speed-up a diagnosis or perform tasks that can only be executed in specialized centers. While a center is free to operate in isolation the Grid infrastructure can be called into play whenever needed.

- **Automated diagnosis** - A medical application can include a check-up, a specialist examination, bio-signal or genomic measurement. By using Grid computing, all the examinations and measurements can be linked using geographically-distributed computing resources, intelligent analysis software and algorithms, and databases via secure and reliable wired or wireless networks. The physician or the specialist can perform analysis and diagnosis with the aid of advanced intelligent automated diagnosis machines, software and tools.
- **Remote consultation** - Online medical information can be acquired by connecting to national or international medical information centers to help with the diagnosis. Remote consultation can also be set up through the network.
- **Remote surgery** - Remote surgery is also possible via a high quality audio/video link connected to a remote surgeon specialist. Surgery, specialist and consultant units, laboratory and instrument tests are all connected to distributed databases and high-performance computing resources.
- **Comparing clinical and research data** - Comparing the results coming from clinical data and research data is a very productive process. Both clinical and research data might be stored at different locations and have different formats and representations which can make the process of comparing it difficult. An integration pass can benefit from the on-going efforts in the fields of metadata description and Grid Digital Repositories [7]. Algorithms for searching and retrieving a knowledge base and for data mining can be properly defined based on the work carried out in the Grid environment.
- **Data Privacy** - Some of the distributed sources of data may have strong privacy rules. A highly reliable system must ensure privacy and security.

The Grid Infrastructures currently available provide various examples of such systems.

3. Requirements for Health Grid

Health Grid applications have specific requirements. We have identified six important main categories as specified in the following list:

1. **Security:** The security requirement includes authentication, authorization and auditing. Through the authentication process, it is possible to know who is accessing a Grid resources and whether this person is allowed to access the Grid network. Authorization allows applications to know what a user is allowed to do in the network or system and what kind of data he or she can access. Through auditing it is possible to record what a user did or tried to do.
2. **Privacy and confidentiality of data:** The privacy and confidentiality of medical data requirement can be achieved by removing sensitive patient data from file headers, using de-identification or encrypting data to guarantee its confidentiality and integrity. For instance, the name and personal data of a patient can be removed from the result of an examination and the information stored separately while linked together. Access to patient data will be stored in a highly secure way with strict control over people who can access this information. The result of the examination can instead be made publically available for comparison studies and other purposes.
3. **Legal and ethical:** The legal and ethical requirement can be achieved by obtaining explicit consent from patients and by gaining clearance to use data from ethics committees. Grid tools can help enforcing local policies and special access constraints on local data.
4. **Interoperability of data:** The interoperability between data stored in different centers and countries has to be taken into account when data are shared over heterogeneous networks and systems. Unfortunately standards for medical data are incomplete. It is fundamental to standardize data format, data structure and data model following existing examples and joining similar efforts in the Grid communities.
5. **Data access transparency:** Once medical data are available over a distributed infrastructure and several formats and protocols are used for their encoding and description, it is important to guarantee location independent transparent access to data. This will allow for a true interchange of

experiences, collaboration and cooperation among specialists sitting at different locations. In order to achieve this goal a storage infrastructure should allow for the preservation of metadata (contextual data about the information being stored) and the negotiation of supported capabilities. The metadata information could concern for example:

- the software that has generated the data
- their encoding
- their validity
- security and privacy information
- other logically connected data.

A framework processing such a description could allow for the negotiation of given capabilities with the data access services. When a specialist asks for the retrieval of a given data set the supporting framework will negotiate with the available storage services offering the wanted data set for the one supporting, for example, the requested specific data format, protocol, security and environment. The selection will then be made among the storage services satisfying the requirements.

6. **Quality of service:** The quality of service requirement concerns the ability to guarantee, for example, the necessary response time, and the required video or image quality. It can differ depending on the specific medical application. For example, remote surgery requires real-time high quality audio/video quality transmission.

4. Related Works in Health Grid

Some Health Care Grid initiatives in the EU, USA and Asia have been identified in order to give an overview of the worldwide efforts in Grid research and development. In 2002 the European Community funded projects specifically in the Health Care field: these are MammoGrid [8], and GEMSS [9]. The UK e-Science program invested funding between 2001 and 2004 on Health Grid applications such as eDiaMonND [10], CLEF [11] and CareGrid [12].

The MammoGrid (European Federated Mammogram Database Implemented on a GRID Structure) project aimed at developing a pan-European distributed database of mammography images using Grid technologies. The evolution of the MammoGrid project was consistent with the evolution of the European Grid infrastructure from the EDG [13] to the EGEE [14].

The goal of the GEMSS (Grid-Enabled Medical Simulation Services) project was to create an innovative Grid middleware to support several medical service applications, including maxillofacial surgery simulation, neuro-surgery support, radio-surgery simulation, inhaled drug delivery simulation, cardiovascular system simulation and advanced medical image reconstruction. These simulation tools can assist medical practitioners in diagnosis, pre-operative planning and near real-time surgery support.

The eDiaMoND (on digital mammography) project aimed at delivering a prototype which could support breast screening in the UK through screening, computer-based training, epidemiology study and computer aided detection of breast cancer. It gave patients, physicians and hospitals fast access to a vast database of digital mammogram images. The database was also used for image analysis algorithms, searching mechanisms for data mining and imaging standardization techniques.

The CLEF (CLinical E-science Framework) project has created a scalable generic architecture based on Grid technology for capturing, integrating, interpreting and using clinical data with genomic data and images within practical clinical systems.

The CareGrid project has developed software for supporting decisions based on trust, privacy, security and context models in the Health Care application domain.

One topic not yet covered is how to provide support to Health Care applications through the Grid Storage Technology. In what follows we describe what is currently provided by Grid Storage Services, highlighting the features that could be exploited.

5. Grid Storage Technology

An application running on a Grid infrastructure needs to be able to transparently store and access data distributed at some storage centers. The Grid middleware is responsible for ensuring the transparency of operations over a multitude of hardware and software solutions used to provide reliable and performing storage.

The need for reliable, scalable, manageable and high performing network storage has led to a proliferation of commercial and customized solutions. Each computing center participating to a Grid makes its own choice for a storage system depending on many factors, such as particularly positive experience and good contacts with some vendors, budget considerations, local manpower to maintain and administer the system, and special contacts with reference sites that give guidelines on solutions to

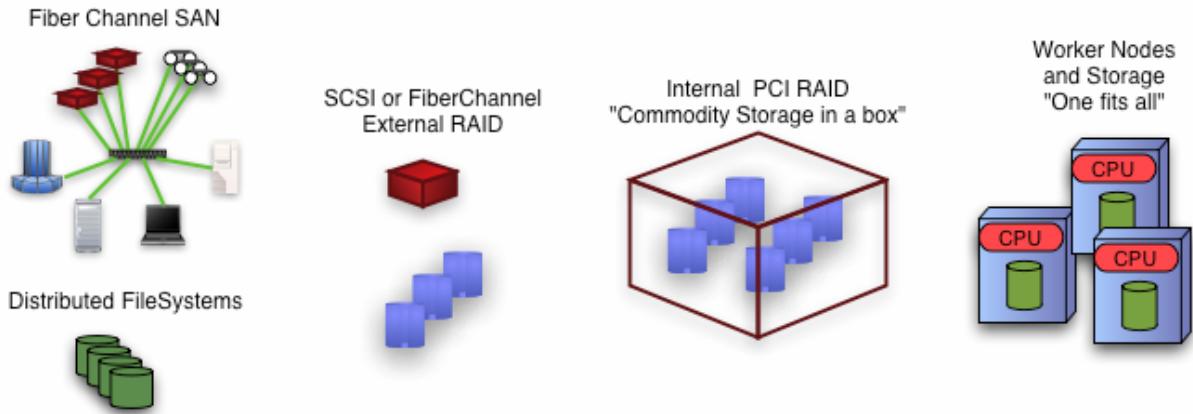


Figure 2. From complex expensive, systems [on the left] to inexpensive storage solutions [on the right].

adopt. As a result, in the Grid we have a great deal of products deployed, all presenting good features and performance, but with a variety of interfaces and implementations. Furthermore, such solutions were designed either for wide area or local use. Figure 2 shows the storage technologies today in use from complex expensive systems to inexpensive storage solutions. In order to answer to the users' need for storage, a system administrator tends to use "anything" that is available. In these centres, the computing nodes of a farm can be equipped with a good amount of medium-quality disks or with small reliable storage systems. To reduce the system administration overhead, other sites might decide to go for commercial solutions that can be easily adapted to the environment that the system administrator has to deal with. For instance, this was the case for many sites choosing to install and support classic distributed filesystems, such as the Network File System (NFS) [15] or the Andrew File System (AFS) [16], or parallel file systems such as IBM GPFS (General Parallel File System) [17] or Lustre [18]. Large centres have stronger requirements. They must provide high capacity, very reliable, and durable storage. Most of the times, this is achieved with tape-based mass storage systems. High performance tape drives are available in a tape library, a system that comes with a robotic arm and an optical label reader able to find the needed tape in the library and mount it in a free tape drive for reading and/or writing. Proprietary or custom written software drives the tape library. Some centres have decided to adapt proprietary backup or Hierarchical Storage Management systems to manage the robotic libraries and serve user requests.

A Grid Storage Element (SE) [19] is the set of hardware and software solutions adopted in order to realize a storage Grid service. It hides the difference

among specific solutions and allows users to consistently and securely store files at a Grid site.

The EDG and EGEE EU funded projects provide a generic Grid infrastructure to several communities of scientists which include biology, Health Care and physics. Several working groups have been created to analyze the storage requirements of the user communities these projects serve [20]. The common agreement was that an SE is a basic storage resource which provides a uniform interface to storage in a Data Grid together with the following set of useful functionalities:

1. Grid users and applications migrate across multiple administrative domains. This has numerous diverse consequences: both a **transparent interface** for specific, very frequent operations and a **set of different communication protocols** need to be supported at the same time. Also and more important, **security protocols** used across domains might differ and must be equally and transparently supported. In particular, authentication and authorization mechanisms valid across domains must be honored.
2. Efficient methods for both local and wide area access must be provided. Reliable and high performing **data transfers protocols** supported across domains on wide area networks are fundamental.
3. Even though Grid-wide privileges should be guaranteed, **local policies and priorities** must be respected.
4. Beside the normal user who needs to store permanently his or her data, the Grid introduces a new set of users, the migratory users who use the local storage as intermediate storage to execute transient operations. Therefore a set of **storage classes** must be available and published in the

- information system so that applications and services can take advantage of them. A storage class is a type of storage space with specific attributes. For instance it might offer reliable, persistent storage or unreliable scratch space for temporary usage only.
5. Grid users also need to have available a set of operations for **managing and reserving storage space** and for **filesystem-like operations** (such as ls, mkdir, ln, file locks, etc.) as well as different **qualities of services** that need to be supported by the local storage solution.
 6. Storage services should also be able to differentiate between valuable and expendable data (**volatile** vs. **permanent data**) when expired reservations are selected for removal. Operations such as **transparent, automatic or forced migration to tertiary storage** (tapes) should be available. In order to avoid that a file in use is removed by a concurrent application, a mechanism called **pinning** is used to keep the system from removing files.
 7. Mechanisms for transparently **locating data** on any storage device must be provided for debugging reasons and for recovering from disasters, besides implementing specific application needs.
 8. Storage systems should also provide a mechanism to **advertise capacity, status, availability and content** to an information system.
 9. **Management and monitoring functions** for Grid global control of service behavior and functionality are important.
 10. **Support for multiple file access and discovery protocols** is requested for those (legacy) applications that need remote file location and access.
 11. **Data integrity mechanisms** allow through the usage of checksum and location independent information to ensure correctness of data even after replication or reprocessing.
 12. **Data privacy mechanisms** are ensured with the support of side services that allow for instance the storage of encryption keys on distributed servers so as to protect from malicious attempts at a specific site.

The community of Grid researchers has invested a big effort in order to understand the needs and requirements for a Grid storage service. The definition of a standard storage interface has allowed for the virtualization of the storage resources and for a wide range of physical storage devices to be optimally utilized. The **Storage Resource Manager (SRM)** specification [21] is used today by many Grid storage services. Other emerging protocols and data

management related services are available and could be used in the context of Health Care. In the next section we provide an overview of the most common protocols used today in the Grid environment for storage services.

6. Grid Storage existing protocols

In what follows we list the existing Grid Storage protocols and their characteristics that are important to support Health Grid applications.

An SRM [21] is a middleware component whose function is to provide dynamic space allocation and file management on shared storage components on the Grid.

An SRM is a standardized interface [22] offered by a SE that satisfies most of the requirements explained in the previous section. It allows authorized clients to initiate requests for allocating and managing a given amount of space with a given quality. The space can be used to create new files or to access existing files stored in the underlying storage system using a set of negotiated protocols.

The granularity of the data that SRM can refer to is a file, since more specific granularities such as “partial files” or “objects” imply knowledge on the structure of the files being managed by the SRM. Such knowledge is application dependent and cannot be generalized. However, data sets and metadata information can be packaged in a file that SRM can manage.

To illustrate the functionality of an SRM, we illustrate a typical Health Care case study and describe how the SRM can support it. A physician would like to make available to his colleagues sitting at different locations a set of medical information (or files) through the Grid. In order to be successful the application the physician uses has to guarantee that the amount of space needed for storing the data is available on the Grid and to his or her application. Therefore, after authentication and authorization are performed, the application can issue SRM requests to allocate or reserve a certain amount of space on a SE with a certain quality (such as magnetic device or online disks) in order to allow for preservation policies to be applied. Additionally, the application is enabled to only use a certain set of data access or transfer protocols. For instance, the application manages medical images through proprietary access protocols. Therefore, via the SRM interface, the application can negotiate the provision of an area of space where the requested protocol can be used. As a result, the SRM provides a protocol dependent file handle that the application can use to store or access the data. The application initiates the store session that includes data encryption and storage of encryption keys on specific servers (like the

Hydra Keyserver [23]). Metadata information can also be stored through interfaces such as the SNIA XAM [24] described later. Access Control Lists can be defined on both the data namespace and the storage space to control access and allow or disallow operations on data and the space containing it. Other physicians at a different location on the Grid can check for the availability of the data in order to start further processing. Therefore, the SRM interface provides for a pinning mechanism, filesystem like operations and mechanisms for transparently locating data. Temporary copies of the data can be created at different locations to optimize data access. Data integrity methods used by the storage services guarantee the correctness and authenticity of the data and the associated metadata.

The space reserved by the application in the example above can have a lifetime that can be finite or infinite. After the lifetime is expired the space can be re-collected by the system and reused to satisfy other requests. Files also have a lifetime and, if not permanent, they can be removed automatically by the system. Additionally, files can have many copies in the system. Such copies have a lifetime (“pin” lifetime). When the lifetime expires the copies might be automatically removed by the system. This allows for further security to be imposed allowing a specific time window for given operations. Copies can have a protocol specific handle with a given lifetime. Clients can use such a handle to access the file’s copies in read mode or to overwrite an existing copy.

In order to logically organize and manage files, the SRM offers a **namespace** similar to the one of a UNIX filesystem. Under a root directory, users can create subdirectories and files. As a result, applications can refer to these files via UNIX-like paths.

The SRM interface is implemented using a Web Service. It uses SOAP over HTTP. In order to allow for authenticated and authorized access the Grid Security Infrastructure [25] is used.

Another emerging protocol that is acquiring popularity is NFS v4 [26]. NFS has been a very popular and successful protocol that allows for transparent serving of files and directories from multiple storage servers on a LAN to multiple clients on the same LAN. Among the new features of NFS v4, we can list the following:

- Improved access and good performance on the Internet.
- Strong security - The authorization and authentication algorithms used can be dependent on the particular storage server and can be negotiated between server and client. This feature is useful when enforcing local policies on the data and to distinguish between Grid and non-Grid users, for instance.

- Enhanced cross-platform interoperability - Windows and Linux servers can both run the NFS v4 protocol serving files to clients transparently.
- Extensibility of the protocol - The protocol can be extended to include specific needed features. For instance, NFS v4.1 can be made aware of data or file sets (files logically grouped together).

Furthermore, NFS v4.1 provides support for metadata encapsulation, a very attractive feature for Health Care applications.

While SRM and NFS v4 address the problem of providing uniform interfaces to world-wide distributed Grid Storage Services, the Storage Networking Industry Association (SNIA) is promoting the definition of the eXtensible Access Method (XAM). This protocol proposes a way to handle reference information (also known as fixed content) at the level of a storage device. XAM-enabled devices allow for the timely retrieval of metadata information connected to data. Such metadata provide a way to relocate data across diverse local hardware platforms without compromising data integrity. Current products for storing and managing reference information propose vendor-specific data access and data management methods (e.g., for naming, retention, and deletion). This implies modifications to the application software when integrating with other storage products and applications [27]. These integration obstacles also limit the ability to share reference information among applications. The SNIA XAM protocol and standard interfaces enable for more functional and sophisticated products. For instance, based on the XAM-retrieved metadata content the storage service can trigger automatic data operations before serving the data to a user for access or manipulation. The XAM application programming interface allows for storing data in a fashion that does not depend on the specific storage system. In particular the following functionalities are important to applications and storage systems:

- **A Global location independent unique name is associated with reference information.** Data are uniquely identified via a location independent name that can be used also to check data integrity. This name is associated to reference information that describes in detail the referenced data. Reference information such as creation date, lifetime of the hosting storage technology, the type of data (like image and text) allow for implementing Information Lifecycle Management (ILM) practices [28] within, for instance, a XAM-based storage system itself.
- **Metadata is strictly coupled with its data.** This facilitates the data management, access and manipulation together with data interoperability

which is one of the main requirements for Health Care applications.

- **Storage systems are accessed via a standard, pluggable architecture.** Hardware and software upgrades can take place without impacting applications. Reference information is stored in a vendor-independent and location-independent manner.
- **XAM also provides a standardized set of management disciplines and semantics for fixed content.** Such disciplines regulate how to access metadata information and what rules regulate changes in data and metadata. Retention and expiration policies can be enforced through these rules.

SNIA/XAM provides an interface for storage and metadata management. However, additional standards should be developed to foster greater levels of functionality and integration for Health Care data:

- Standards to describe the data content
- Security information
- Common metadata to support data classification standards allowing storage-centric ILM practices that do not depend on the specific application that generated the data.
- Reference information-naming schemas including standards for embedding structured data into a flat object space (or metadata information into a flat file). These schemas will facilitate application integration when processing file systems, databases, and other data resources with significant internal structure.
- Interfaces for bulk movement of reference information such as migrating, backing up, and replicating reference information among XAM-type systems.

7. Solution

The EGEE project [29], co-funded by the European Commission and coordinated by CERN, brings together experts from over 32 countries with the common aim of building on recent advances in Grid technology and developing a service Grid infrastructure which is available to scientists 24 hours-a-day. The project provides researchers in academia and industry with access to a production level Grid infrastructure independent of their geographic location. The EGEE project also produces a grid middleware distribution called gLite which was born from the collaborative efforts of more than eighty people in twelve different academic and industrial research centers. gLite provides a leading-edge, best-of-breed framework for building grid applications.

The entire gLite distribution is composed of several services described in Figure 3.

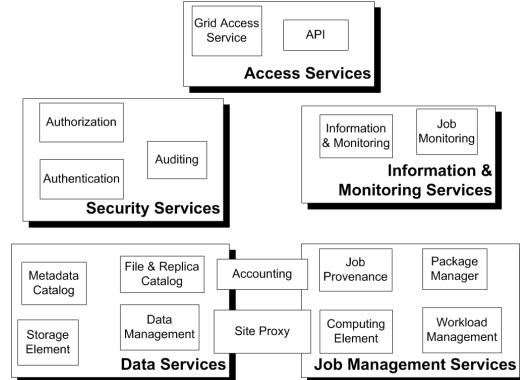


Figure 3: gLite services.

The gLite middleware and services have been used to build Digital Repository infrastructures in fields such as environment and biodiversity. An example is given by the D4Science [30] project involving participants such as the European Space Agency and the WorldFish Center. The gCube [31] middleware allows for building an on-demand distributed virtualized environment that includes specialized services to store and access data and metadata and to manipulate them.

The gLite services that are of interest for Health Grid applications are the Security, Data Management and Storage Services. We examine them and highlight their importance in the Health Care context.

The basic services to allow for authorization, authentication and auditing are the Grid Security Infrastructure based on X509 certificates and the Virtual Organization Management Server (VOMS) that allows for the assignment of specific roles to individuals or for the definition of specific groups [32]. Grid services can then use such definitions to allow or disallow access to resources based on Virtual Organization defined policies. The Hydra Keyserver developed at CERN allows for the management of split keys that can be used for encryption and decryption of data.

From a Data Management point of view, the AMGA metadata catalogue [33] allows for the management of metadata information associated to specific sets of data. Metadata can be associated to Grid universal data identifiers that the LCG File Catalogue (LFC) [34] associates to physical data available on a set of Storage Elements in the Grid.

The Disk Pool Manager (DPM) [35] is an SRM-enabled Grid storage service that provides the

functionalities described above. In [36, 37] it is described the experiment of serving medical data images stored in Digital Imaging and Communications in Medicine (DICOM) [38] format over the Grid and make them available in a trusted way to a set of hospitals participating in the experiment.

SNIA/XAM enabled technology has still not been exploited for Health Care purposes. We intend to provide a prototype where XAM enabled storage elements will be used to manage Health Care metadata information.

8. Validation of Solutions

The interoperability between various standards and the exchange of data amongst different resources constitutes an important issue in Health Care applications. Standards allow institutes to transmit clinical data and images from the electronic patient folder to different hospital sites requiring these elements for diagnosis, medical studies and treatment. Medical image examination is a subset of the medical folder that includes images, prescription, reports of specimen analysis and other elements.

A system called ETICS (eInfrastructure for Testing, Integration and Configuration of Software) [39, 40, 41, 42] has been designed to automate the investigation of the interoperability between various standards and the exchange of data among different resources. It provides a service for software projects and infrastructures by integrating well-established procedures, tools and resources in a coherent framework and adapting them to the special needs of distributed software. These tools help improving the interoperability of software applications and middleware components, evaluating software by comparing it with published standards and validating the portability of the software across different operating systems.

The ETICS system is addressed to projects developing Grid and distributed software to help them define processes, methodologies and tools to implement static and dynamic analysis (such as evaluating the adherence to user defined coding conventions, making custom analysis like IPv6 compliance, evaluating the code coverage for unit tests). Nevertheless, the ETICS system is equally oriented towards non-Grid projects.

In [41] an implementation of how the ETICS system supports interoperability issues is described taking into consideration the DICOM standard.

9. Conclusions

In this work we have identified the Grid services that could facilitate the integration and interoperation of Health Care data and frameworks world-wide. While many of the current Health Care Grid projects address issues such as data location and description on the Grid and the security aspects the problems connected to data storage, integrity, preservation and distribution have been neglected. We have identified the currently available Grid storage services and protocols that can come in handy when dealing with those problems. We are planning to make available a prototype infrastructure using EGEE gLite as a proof of concept for implementing our ideas.

In the future, we envisage using the existing EGEE gLite infrastructure deploying the additional Grid services identified in this paper in order to provide a Grid platform for the distribution and sharing of medical information. We have identified a few medical communities that have expressed interest in exposing their data into this proposed Grid infrastructure.

References

- [1] I. Foster, C. Kesselman, and S. Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations," In *International Journal of Supercomputer Applications*, vol. 15, no. 3, pp. 200–222, 2001, <http://www.globus.org/research/papers/anatomy.pdf>.
- [2] V. Breton, K. Dean and T. Solomonides, "The Healthgrid White paper," In Proceedings of Healthgrid 2005 (From Grid to HealthGrid), IOS Press, vol. 112, pp. 249-321, 2005.
- [3] W. Leister, S. Mazaher, J.I. Vestgården, B. Ø. Johansen, B. Nordlund, "Grid and related technologies," Number 1000, Norsk Regnesentral, 2004, ISBN 82-536-508-4.
- [4] A. Abbas, "Grid Computing: A Practical Guide to Technology and Applications," In *Charles River Media*, 2004.
- [5] I. Balasingham, H. Ihlen, W..Leister, P. Røe, and E. Samset, "Communication of Medical Images, Text, and Messages in Inter-Enterprise Systems: A Case Study in Norway," In *IEEE Transactions on Information Technology in Biomedicine*, Vol. 11, N. 1, pp: 7-13, Jan 2007.
- [6] The Service Level Agreement, <http://www.sla-zone.co.uk/>.
- [7] P. Watry, and R. Larson, "Cheshire 3 framework white paper: implementing support for digital repositories in a data grid environment, " In Proceedings of Challenges and Technologies of Local to Global Data interoperability, pp. 60-84, 20-24 June 2005,
- [8] R. Warren, T. Solomonides, C. del Frate, I. Warsi, J. Ding, M. Odeh, R. McClatchey, C. Tromans, M. Brady, R. Highnam, M. Cordell, F. Estrella, and S. R. Amendolia, "Mammogrid - A Prototype Distributed Mammographic Database for Europe," In *Clinical Radiology*, vol. 62, no. 11, pp. 1044-1051, June 2007.

- [9] "GEMSS (Grid-Enabled Medical Simulation Services) Project," <http://www.it.neclab.eu/gemss/>.
- [10] "UK e-Science eDiaMoND Project," <http://www.ediamond.ox.ac.uk/>.
- [11] "CLEF -integrating information for the clinical e-Scientist," <http://www.clinical-escience.org/start.html>.
- [12] Y. Liu, J. Bacon, and R. Wilson-Hinds, "On smart-care services: Studies of visually impaired users in living context," In Proceedings of the First International Conference on the Digital Society (ICDS'07), pp. 32-40, IEEE Press, January 2007.
- [13] "European Data Grid (EDG) Project," <http://www.edg.org/>.
- [14] "EGEE (Enabling Grid for E-sciencE) Project," <http://www.eu-egee.org>.
- [15] P. S. Yeager, "A distributed file system for distributed conferencing system," Thesis, University of Florida, 2003.
- [16] AFS File System, <http://lxr.linux.no/linux/Documentation/filesystems/afs.txt>.
- [17] IBM General Parallel File System, <http://www-03.ibm.com/systems/clusters/software/gpfs/index.html>.
- [18] Lustre File System, <http://www.sun.com/software/products/lustre/>.
- [19] G. A. Stewart, G. A. Cowan, B. Dunne, A. Elwell, and A. P. Miller, "Optimising LAN access to grid enabled storage elements," In *Journal of Physics: Conference Series*, vol. 119, 2008.
- [20] Baseline Services Working Group, <http://lcg.web.cern.ch/LCG/peb/BS/>
- [21] A. Shoshani, P. Kunszt, H. Stockinger, K. Stockinger, E. Laure, J.-P. Baud, J. Jensen, E. Knezo, S. Occhetti, O. Wynge, O. Barrig, B. Hess, A. Kowalski, C. Watson, D. Petrvick, T. Perelmutov, R. Wellner, J. Gu, and A. Sim, "Storage Resource Management: Concepts, Functionality, and Interface Specification," GGF 10, The Future of Grid Data Environment, 9-13 March 2004, Humboldt University, Berlin Germany.
- [22] A. Sim, A. Shoshani, P. Badino, O. Barrig, J.-P. Baud, F. Donno, M. Litmaath, T. Perelmutov, D. Petrvick, E. Corso, L. Magnoni, J. Gu, S. De Witt, J. Jensen, M. Haddox-Schatz, B. Hess, A. Kowalski, and C. Watson, "The Storage Resource Manager Interface Specification Version 2.2," OGF-GSM, 24 May 2008.
- [23] D. Gilliam, "Summary report on enterprise security workshop," In Proceedings of the Eleventh IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, pp. 43-46, 2002.
- [24] Storage Networking Industry Association, *Information Management – Extensible Access Method (XAM) – Part 1: Architecture*, Version 1.0, 2 April 2008, http://www.snia.org/forums/xam/technology/specs/XAM_Arch_v1.0_Apr6.pdf.
- [25] S. Campana, A. Delgado Peris, F. Donno, P. Mendez Lorenzo, R. Santinelli, and A. Sciafa', "Toward a Grid Technology Independent Programming Interface for HEP Applications," In Proceedings of CHEP2004, Sept. 27 - Oct. 1, 2004 Interlaken, Switzerland.
- [26] NFS v4, <http://www.nfsv4.org/>.
- [27] L. Wood, "Getting a Fix on Fixed Content", 9 June 2006, Enterprise Storage Forum.com, <http://www.enterprisestorageforum.com/continuity/features/article.php/3612266>.
- [28] Information Lifecycle Managment Solutions, http://www.solix.com/information.lifecycle.management_solutions.htm.
- [29] EGEE Middleware Architecture, August 2004, <https://edms.cern.ch/file/7476451/1.0/Architecture.pdf>.
- [30] D4Science, <http://www.d4science.eu/>.
- [31] gCube, <http://www.gcube-system.org/>.
- [32] VOMS, <http://voms.forge.cnaf.infn.it/>.
- [33] AMGA, <http://cern.ch/amga/>
- [34] J.-P. Baud, J. Casey, S. Lemaitre, C. Nicholson, G. Stewart, "LCG Data Management: From EDG to EGEE," http://www.gridpp.ac.uk/papers/ah05_dataman.pdf.
- [35] LCG Disk Pool Manager (DPM): <https://twiki.cern.ch/twiki/bin/view/LCG/DpmAdminGuide>.
- [36] J. Montagnat, *et al.*, "Bridging clinical information systems and grid middleware: a Medical Data Manager", In Proceedings of HealthGrid conference (HealthGrid'06), pp: 14–24, Spain, June 2006.
- [37] J. Montagnat, T. Glatard, D. Lingrand, R. Texier, "Exploiting production grid infrastructures for medical images analysis", In Proceedings of the First Singaporean-French Biomedical Imaging Workshop (SFBI'06), Singapore, Oct 2006.
- [38] Digital Imaging and Communications in Medicine, <http://medical.nema.org/>.
- [39] M. Bégin, *et al.*, "Build, Configuration, Integration and Testing Tools for Large Software Projects: ETICS", In Springer Verlag Lecture Notes in Computer Science (LNCS) Series, LNCS 4401, pp: 81-97, 2007.
- [40] M. Bégin, *et al.*, "Analysis of Requirements for Automated Interoperability Testing", In Proceedings of the Tenth World Conference on Integrated Design & Process Technology, IDPT 2007, Antalya, Turkey, June 3-8, 2007.
- [41] S. Da Ronco, G. Diez-Andino Sancho, L. Dini, E. Ronchieri, and M. Selmi, "Integrated Solutions For Information Sharing In Health Care Applications," Proceedings of the 41st Hawaii International Concerence on System Sciences, IEEE Computer Society, pp. 254, January 7-10, 2008, Big Island, Hawaii.
- [42] E. Ronchieri, M.-E. Begin, and C. Vistoli, "A Study of Multi-Node Mechanism to Interoperability Issue," Proceedings of the Eleventh World Conference on Integrated Design and Process Technology (IDPT 2008), June 1-6, 2008, Asia University, Taichung, Taiwan.