A SYSTEM ARCHITECTURE FOR THE BING
Brain Image Network Grid

Micael Pedrosa, Luís Alves
Institute of Electronics and Telematics Engeneering of Aveiro (IEETA), Campus de Santiago, Aveiro, Portugal

Ilídio Oliveira, José Maria Fernandes, João Paulo Silva Cunha
IEETA/Dept. of Electronics, Telecommunications and Informatics, University of Aveiro
Campus de Santiago, Aveiro, Portugal

Keywords: Medical Data, Brain Imaging, Grid-Enabled Scientific Repositories, Workflows, Semantic Data Structure.

Abstract: This paper presents a detailed architecture model for the Brain Imaging Network Grid (BING) that will be the main IT infrastructure of the recently created Portuguese Brain Imaging Consortium. The proposed architecture follows a service oriented philosophy and is designed to empower medical data sharing and processing, specifically brain images. Allowing the use of computationally intensive methods like feature extraction and retrieving of structured information, this system will take advantage of Grid computing new paradigm. In BING context, Grid infrastructure is the right option to provide the ability to seamless aggregate distributed computational power, extensive storage resources and high-bandwidth networking. The goal is to develop a system that simultaneously can provide basic data services, allow collaborative research between geographically distributed partners (e.g. analysis processes, workflows) and make use of the Grid computational power.

1 INTRODUCTION

There is a long tradition in the Portuguese scientific community of R&D in neurosciences where Brain Imaging (BI) is a sub-area. BI is in the frontier between neurology, engineering and physics. Multimodal medical imaging techniques, such as Magnetic Resonance Imaging (MRI and fMRI) and Spectroscopy (MRS), Single Photon/Positron Emitting Tomography (SPECT/PET) among others, are emergent medical research tools that can provide valuable information for diagnosis of brain diseases. High-resolution electroencephalogram (HR-EEG) and techniques for synchronizing and fuse its analysis results and several imaging techniques are also part of BI. Most of the BI related areas, from medical, engineering or physics are subjects of research in many R&D groups within the Portuguese scientific system.

This strong critical mass in BI of the Portuguese research community paved the way to a "bottom-up" process, organized initially around a group of young researchers in BI working in Portugal and abroad, to propose the creation of a BI centre and network to join all community efforts. From this process emerged the consortium of the universities of Aveiro, Coimbra, Minho and Porto which is now a reality, already funded in 81.3% for the first 5 years of operation of a BI centre, located in Coimbra (where neuroimaging equipment will be installed such as 3 Tesla MRI equipment), and a national IT infrastructure that is open to the participation of other national institutions.

The BI needs IT infrastructure to support distributed collaborations between different neuroimaging research member groups. In a previous paper (Cunha et al., 2007) the authors addressed the national wide deployment of the BI network at a organizational and physical deployment level (Figure 2): it will be constituted by a data provider (BI centre), located at the University of Coimbra, two integrated data processing and storage provider nodes, located at the Universities of Aveiro and Porto, and a basic and clinical neuroscience data access client node at the University of Minho.

Pedrosa M., Alves L., C. Oliveira I., Fernandes J. and Silva Cunha J. (2009).
A SYSTEM ARCHITECTURE FOR THE BING - Brain Image Network Grid.
In *Proceedings of the International Conference on Health Informatics*, pages 276-281
DOI: 10.5220/0001549202760281

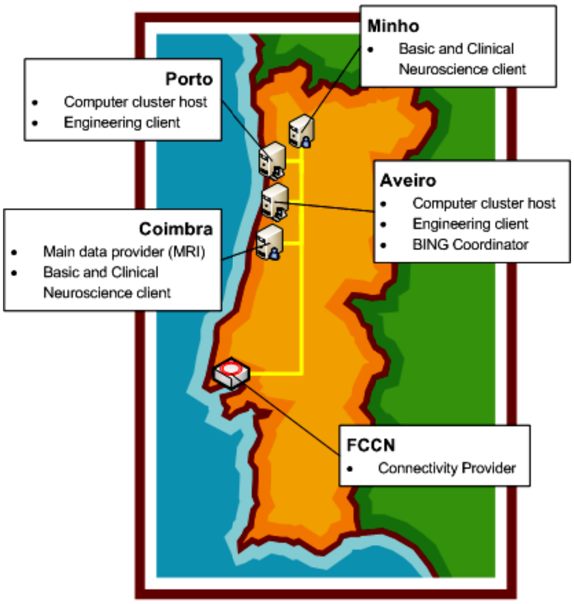

Figure 1 - Brain Imaging Network Grid.

2 A GRID-ENABLED BING

From the beginning a Grid infrastructure, Brain Image Network Grid (BING) seemed the right option to provide the ability to seamless aggregate distributed computational power, extensive storage resources and high-bandwidth networking (Cunha et al., 2007). In addition, Grids also ensure a proper level of security, both at identity (digital certificates) and access (Virtual Organizations management) levels. At the same time successful examples exist in the use of Grid in medical image processing to handle the demanding requirements of large images storage and communication, and to enable complex analysis workflows (Breton, 2005) (Montagnat J., 2004) (Amendolia, 2005).

Grid computing is being promoted in the scientific community as an essential technology to solve large scientific challenges due to its ability to bring together distributed capabilities (computing resources, storage space, information sources and scientific instruments) and use them in a integrated virtual environment.

The application of Grid in life sciences can be generically labeled as HealthGrids. The HealthGrid vision advocates the use of Grid Infrastructures for Medical Research, Healthcare and Life Sciences, which implies the availability of Grid Services and the definition and adoption of international standards and interoperability mechanisms (Breton et al., 2006).

At a higher level, semantic-enabled Grids (Aisha and Lampros, 2006) can support services such medical data analysis protocols and workflows and promote the decoupling between Grid computational services. Grid will leverage the required computational and data resource to research and education, enabling the communities to publish, share and explore clinical cases.

2.1 BING Use Cases

BING infrastructure will need to accommodate several research profiles from simple data providers to high specialized research user with very specific requirements (e.g. computational power, several data types specificities, complex processing workflows).

Good examples can be associated to digital signal processes that may be required for image analysis (Taylor, 1997), for data visualization in client applications or for interpretation on cases analysis results. Although some of these use must be developed by informatics researchers (Poliakov, 2001), when possible the usage should be kept at user level as in medical validation of workflows and results.

From the analysis of some scenarios the following actors and use cases were identified:

The informatics engineer who requires a generic system, that provides extension points for development of data interpreters and adding them to the semantic part of the system.

The digital signal processing researcher that needs to describe and reuse algorithms, test new image processing methods (that may be integrated in the system with help of informatics engineers) and report outputs to clinical personal for evaluation.

The clinical expert whose research is at the clinical research semantic where concepts like patient studies, image data and reports should be logically detached from the BING engine details. For instance, validation of study results and consequent output from signal processing, comparative query of cases, identify from reports significant conclusions, associate patient information to a range of data inputs and results.

Data security access based on semantic rules providing security situations like patient anonymity and discard of non relevant information for a specific actor type.

Information and work share between actors, data classification and organization.

2.2 BING Services

In this context it was natural to assume a service-oriented solution where services can be customized to specific needs of each user without compromising the integrity and expansibility of the BING framework that from the initial stage – centred in providing data storage - to the future evolutions where we envisage that BING, through the definition of the basic services will be able offer a tear of services for a vast sort of clients, providing processing power for complex signal analysis and a standard access to heterogeneous data storages.

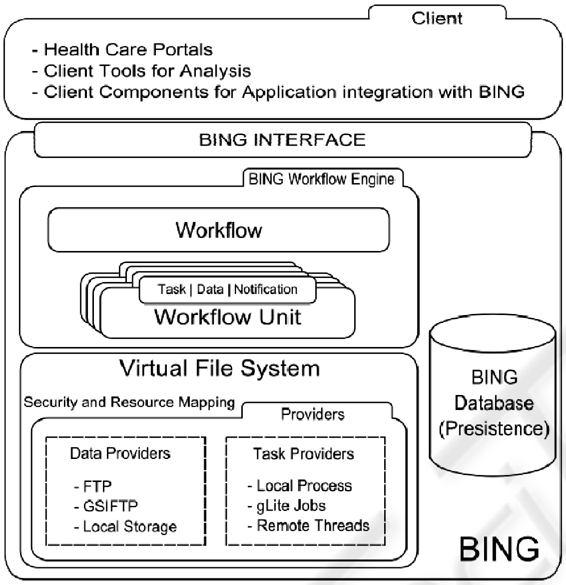

Figure 3 - BING Architecture Overview.

The initial focus on data storage IT facilities is the natural consequence of the fact that BING users, regardless of their skills and intended usage, will need such service. It is essential to enable safe sharing and transparent access to digital data (acquisitions, reference models, etc.). This approach is essential to ensure that the data IT network will really connect the distributed research centers and be the leverage for cooperation where each member is able to contribute with specific skills namely: providing the raw neuroimaging data, developing bioengineering analysis methods, using the different integrated computational methods to their clinical neuroimaging research purposes.
The concept of the BING is strongly inspired on other international experiences (such as the Biomedical Informatics Research Network (Grethe et al., 2005)) and will be completely opened to grow easily and foster other nodes from any other research groups that may join this initiative in the future. In this particular aspect Grid technology is a natural ground to deploy such a collaborative and controlled network. It addresses specifically the requirements of Virtual Organizations, understood as teams of trusted organizations, willing to share their computational resources (processors and storage), knowledge (data and analysis workflows), best practices and also specific and expensive equipments (such as advanced medical scanners).

3 BING ARCHITECTURE

BING architecture can be described in a top-down approach in three main functional divisions, as illustrated Figure 4: BING Interface, BING workflow engine and Virtual File System. This architecture will serve client applications, health care portals, custom applications, client tools for data analysis and visualization.

3.1 BING Interface

Clients can contact the internal engine via BING Interface, a tier of Web Services provided by standard SOAP/WSDL. Its purpose is to offer a way in to the Virtual File System, a structured system that addresses all other parts of the engine and handles security access.

3.2 BING Workflow Engine

The central part of BING is the BING Workflow engine. The goal here is to supply a system independent of data storage nature and processing power providers that, besides direct access to processing operations enable the definition of sequences of processing operations as abstract workflows either as neuroimaging semantic level protocols or as simple batch processing instructions.

3.2.1 Wokflow Unit

The basic concept for operations workflow definition is the Workflow Unit (WU).
WU can be tasks, data references or BING events that can be created, sent and received by any compatible provider type. WU are not implementation but abstract interfaces with a minimal semantic associated depending on the type (task, data or notification).
The interface concept is a natural object –oriented concept that, besides being independent of

implementation issues it provides mechanism for extending and complementing the basic WU concepts. By this way it is possible to accommodate different software implementation solutions without compromising the workflow definition – namely by hiding specific state implementations and providing only clear access functions with clear semantics.
Tasks, data and notifications are all WU that can be used with any compatible provider type.

3.2.2 Workflows

The workflow concept in BING is not new and maintain a high level description of software architecture. A workflow can be seen as a succession of data WU and processing WU, which in the end will generate some output that will be stored in BING – typically an image processing result.
To describe the workflow and maintain an high level description, XML based solutions such as "Open Symphony Workflow" (opensymphony, 2008) can be a good solution as they are is flexible enough while providing some semantic for workflow validation through a Document Type Definitions (DTD) (Shiyong et al., 2005).
With visual client tools using already available operators and tasks extensions, medical researchers are able to compose a workflow with no advanced informatics knowledge.

3.3 Virtual File System

Virtual File System (VFS) is responsible for providing an abstraction of the information related services of the BING: organize instances of WUs, enable access to the BING providers.
Providers can be data store servers (e.g. like FTP, GSIFTP), or other processing hosts (e.g. location to Grid Nodes). The VFS will be responsible for keeping the mapping between the providers and the associated VFS folders and WUs. In BING, providers and WUs can have modality specific semantic such EEG, fMRI, MEG, CT (Productions, 2001) which may facilitate the semantic verification of workflows.
The VFS concept is similar to common file systems, more like a semantic file system (Craig A. N. Soules, 2004) (Prashanth Mohan, 2006) but in this case, instead of files, it stores WUs and every folder and WU reference are abstract structures mapped on actual data/processing facilities in different providers, local or remote.
Another important responsibility of the VFS is to provide a secure and reliable environment where users and permission access to both folders and WUs is controlled. This implies that VFS must preserve users and permissions configuration and provide anonymity of data when required (e.g. enable access to data without any patient information).
Given the VFS abstract nature, it will be subject to provider's access policies. Nevertheless it will enable a fine grained configuration in order to define access and security policies at the BING level.

3.4 Persistence

The Processing and workflow Engine will need a persistence mechanism in order to store both WU information and maintain the VFS information. The natural choice for a preliminary approach will be a relational database that 1) enable fast development for a specialized software engineer team, 2) are recognised for their performance in query relational and structured data. An example of a similar problem can be found applied to semantic file systems in (SELENG, 2007).

3.5 BING and Grid Middleware

The BING can be seen as part of the Grid middleware where the Grid framework essentially supply basic data and processing providers. While maintaining a BING specific semantic, the Grid Infrastructure can be used to access computing power, an aggregation of distributed storage and heterogeneous resources.
This solution does not compromise the interoperability concern commonly referred on Grid conceptions (eDiaMoND, 2008) (Foster et al., 2001). It is important that the various stakeholders of a Virtual Organization can transparently use all the resources available (data, sensors, etc.) in collaborative terms. Interoperability is achieved by the use of a standards-based architecture, commonly named as Grid middleware. The Grid middleware dynamically shares, manages and controls physical and logical resources geographically separated. In some sense, BING relies on this functionalities for IT infrastructure while providing services more specific to BI support infrastructure, similarly to the the Biomedical Informatics Research Network (Grethe et al., 2005).

3.6 Grid Middleware Providers

One good candidate to be the BING Grid middleware is the gLite (gLite, 2008). The gLite middleware has been introduced by EGEE project as a result of contributions from many other projects including LCG, European DataGrid and VDT (gLite, 2007). This middleware is based on GLOBUS2, Condor and many other services developed in projects like those above. gLite orchestrates multiple service layers such as Access, Security, Information and Monitoring, Data, and Job Management Services (gLite, 2007). It also accomplishes the Open Grid Services Architecture (OGSA) Standard introduced by the Open Grid Forum (OGF) (Foster et al., 2002). This software is Open licensed and portable.

The gLite job management is guaranteed by the use of a gLite User Interface (UI) node that will operate as a BING gateway to the Grid layer. This node provides the proxy creation and, by the use of the WMProxy (Workload Manager Proxy) Service, access to the Workload Management System (WMS). WMProxy accepts job submission requests described with the Job Description Language (JDL) (Pacini, 2006) and control requests such as job cancellation, job status, job output retrieval, etc (Maraschini A., 2006).

For data providers the gLite middleware supply complementary data management services like the Storage Resource Manager (SRM) and can support different data access protocols and interfaces like GSIFTP, RFIO and gsidcap allowing advanced functionality such as dynamic space allocation and file management on shared storage systems (gLite, 2007).

Some early results from our group (Andrade et al., 2007b) support that gLite might be a good option. On a vertical application (from user interface, storage and processing) on Multi-voxel Non-linear fMRI Analysis (Andrade et al., 2007a) it was possible to have a positive assessment on the gLite both on the processing power (report a 7 fold reduction in time consumption on EGEE Grid environment) and on the application development.

4 DISCUSSION

This paper presents a detailed architecture model for the Brain Imaging Grid (BING) that will be the main IT infrastructure of the recently created Portuguese Brain Imaging Consortium. The architecture was designed with flexibility in mind providing extension points for semantic information search and query.

Currently a prototype is under development integrating extensions to the basic WU concepts: *LocalGliteTask* WU to execute gLite jobs, *LocalTask* WU to map local processes, *FtpData* WU to access to data in FTP and *LocalData* WU to map local data storage. These are being used in BING workflow engine development in order to support simple workflows. Simultaneously we are developing the basic processes to handle different data formats, image modalities in order to be able to support the complex brain image data model at the VFS level.

ACKNOWLEDGEMENTS

This work was partially funded by the projects BING (GRID/GRI/81833/2006) and Geres-Med (GRID/GRI/81819/2006) of the F.C.T.

We would like also to thank our project colleagues António Sousa Pereira (University of Aveiro), Miguel Castelo-Branco (University of Coimbra), Nuno Sousa (University of Minho), Aurélio Campilho (University of Porto) and their institutions for the support to the BING project.

REFERENCES

Aisha, N. & Lampros, K. S. (2006) Discovering HealthGrid Services. *Services Computing, 2006. SCC '06. IEEE International Conference on.*

Amendolia, S. R., et al. (2005) Deployment of a grid-based medical imaging application, IOS Press.

Andrade, R., Oliveira, I., Fernandes, J. M. & Cunha, J. P. (2007a) A Grid framework for non-linear brain fMRI analysis. *Stud Health Technol Inform,* 126**,** 299-305.

Andrade, R., Oliveira, I., Fernandes, J. M. & Cunha, J. P. (2007b) Multi-voxel Non-linear fMRI Analysis: A Grid Computing Approach. *IberGRID* Santiago de Compostela.

Breton, V., Blanquer, I., Hernandez, V., Legré, Y. & Solomonides, T. (2006) Proposing a roadmap for HealthGrids. HAL - CCSD.

Breton, V., Dean, K., et al. (2005) The Healthgrid White Paper. *Stud Health Technol Inform,* 112**,** 249-321.

Craig A. N. Soules, G. R. G. (2004) Toward automatic context-based attribute assignment for semantic file systems. Pittsburgh, Carnegie Mellon University.

Cunha, J. P. S., Oliveira, I., Fernandes, J. M., Campilho, A., Castelo-Branco, M., Sousa, N. & Sousa Pereira, A. (2007) BING: The Portuguese Brain Imaging Network GRID. *IberGRID* Santiago de Compostela.

Ediamond (2008) eDiaMoND – Diagnostic Mammography National Database Project

Foster, I., Kesselman, C., Nick , J. & Tucke, S. (2002) The Physiology of the Grid: An Open Services Architecture for Distributed Systems Integration. .

Open Grid Service Infrastructure WG, Global Grid Forum.
Foster, I., Kesselman, C. & Tuecke, S. (2001) The Anathomy of the Grid. *Intl J. Supercomputer Applications,* 15.
GLITE (2007) gLite User Guide.
GLITE (2008) gLite: Lightweight Middleware for Grid Computing.
Grethe, J. S., Baru, C., Gupta, A., James, M., Ludaescher, B., Martone, M. E., Papadopoulos, P. M., Peltier, S. T., Rajasekar, A., Santini, S., Zaslavsky, I. N. & Ellisman, M. H. (2005) Biomedical informatics research network: building a national collaboratory to hasten the derivation of new understanding and treatment of disease. *Stud Health Technol Inform,* 112**,** 100-9.
Maraschini A., E. A. (2006) WMPROXY SERVICE. IN INTEGRATION, J. M. E. A. (Ed.) *EGEE User's Guide.* EGEE.
Montagnat J., B. F., Benoit-Cattin H., Breton V., Brunie L., Duque H., Legré Y., Magnin I. E., Maigne L., Miguet S., Pierson J. -M., Seitz L., Tweed T. (2004) Medical Images Simulation, Storage, and Processing on the European DataGrid Testbed. *Journal of Grid Computing,* 2.
Opensymphony (2008) OSWorkflow.
Pacini, F. (2006) JOB DESCRIPTION LANGUAGE - ATTRIBUTES SPECIFICATION (submission through WMProxy Service). IN MIDDLEWARE, J.-. (Ed.). EGEE.
Poliakov, A. V. A. A., Evan M and Corina, David P and Ojemann, George A and Martin, Richard F and Brinkley, James F (2001) Server-Based Approach to Web Visualization of Integrated 3-D Medical Image Data. *Proceedings, American Medical Informatics Association Fall Symposium.*
Prashanth Mohan, R., Venkateswaran S and Dr. Arul Siromoney (2006) Semantic File Retrieval in File Systems using Virtual Directories. *International Conference on High Performance Computing (HiPC 2006).* Bangalore, India
ProductionS, T. W. N. Y. A. D. G. (2001) The Secret Life of the Brain.
Seleng, M.-L., Michal - Balogh, Zoltan - Hluchy, Ladislav (2007) RDB2Onto: Approach for creating semantic metadata from relational database data. *INFORMATICS 2007* Bratislava, Slovak Society for Applied Cybernetics and Informatics.
Shiyong, L., Yezhou, S., Mustafa, A. & Farshad, F. (2005) On the consistency of XML DTDs. *Data Knowl. Eng.,* 52**,** 231-247.
Taylor, J. G. (1997) Problems in the analysis of brain imaging data. Digital Signal Processing Proceedings, 1997. DSP 97., 1997 13th International Conference on.