The BIOPATTERN Grid – Implementation and Applications

P. Hu [1], L. Sun [1], C. Goh [1], B. Hamadicharef [1], E. Ifeachor [1,a], I. Barbounakis [2], M. Zervakis [2], N. Nurminen [3], A. Varri [3], R. Fontanelli [4], S. Di Bona [4], D. Guerri [4], S. La Manna [4], K. Cerbioni [5], E. Palanca [5] and A. Starita [5]

[1] School of Computing, Comm. and Electronics, University of Plymouth, UK
[2] Telecommunications System Institute, Technical University of Crete, Greece
[3] Institute of Signal Processing, Tampere University of Technology, Finland
[4] Synapsis S.r.l. in Computer Science, Italy
[5] Computer Science Department, University of Pisa, Italy

Abstract

The primary aim of this paper is to report the development of a new testbed, the BIOPATTERN Grid, which aims to facilitate secure and seamless sharing of geographically distributed bioprofile databases and to support analysis of biopatterns and bioprofiles to combat major diseases such as brain diseases and cancer within a major EU project, BIOPATTERN (www.biopattern.org). The main objectives of this paper are 1) to report the development of the BIOPATTERN Grid prototype and implementation of BIOPATTERN grid services (e.g. data query/update and data analysis for brain diseases); 2) to illustrate how the BIOPATTERN Grid could be used for biopattern analysis and bioprofiling for early detection of dementia and for assessment of brain injury on an individual basis. We highlight important issues that would arise from the mobility of citizens in the EU and demonstrate how grid can play a role in personalised healthcare by allowing sharing of resources and expertise to improve the quality of care.

1. Introduction

Grid computing is aimed to provide a global Information Communication Technology (ICT) infrastructure to facilitate seamless and secure sharing of geographically distributed resources (e.g. data, storage, computation, algorithms, applications and networks). Great efforts, resources and funding have been put into national, regional and international initiatives in grid infrastructure, grid core technologies and grid applications. The integration of grid computing and healthcare has formed a new exciting and specialist area called Healthgrid. Examples of healthcare applications include distributed mammography data retrieval and processing (e.g. the EU's MammoGrid [1] and the UK's eDiaMoND [2] projects), and multi-centre neuro-imaging (e.g. the USA's BIRN [3] and Japan's BioGrid [4]). There is a trend in modern medicine towards individualisation of healthcare and, potentially, grid computing can also play a role in this by allowing sharing of resources and expertise to improve the quality of care.

In this paper, we report efforts to exploit grid computing to support individualisation of healthcare to combat major diseases such as brain diseases within a major EU-funded, Network of Excellence (NoE) project, BIOPATTERN (www.biopattern.org). The Grand Vision of the project is to develop a pan-European, coherent and intelligent analysis of a citizen's bioprofile; to make the analysis of this bioprofile remotely accessible to patients and clinicians; and to exploit bioprofiles to combat major diseases such as cancer and brain diseases. A biopattern is the basic information (pattern) that provides clues about underlying clinical evidence for diagnosis and treatment of diseases. Typically, it is derived from specific

[a] Corresponding Author: Prof Emmanuel Ifeachor, School of Computing, Communications and Electronics, University of Plymouth, Plymouth PL4 8AA, U.K. Email: E.IFeachor@plymouth.ac.uk.

data types, e.g. genomics and proteomic information and biosignals, such as the electroencephalogram (EEG) and Magnetic Resonance Imaging (MRI). A bioprofile is a personal `fingerprint' that fuses together a person's current and past medical history, biopatterns and prognosis. It combines data, analysis, and predictions of possible susceptibility to diseases. It will drive individualisation of care.

The aim of the BIOPATTERN Grid is to facilitate secure and seamless sharing of geographically distributed bioprofile databases and to support analysis of biopatterns and bioprofiles to combat major diseases such as brain diseases and cancer.

The main objectives in this paper are 1) to report the development of a new Grid test bed, the BIOPATTERN Grid, for biopattern analysis and bioprofiling in support of individualisation of healthcare. We focus on the implementation of the BIOPATTERN Grid prototype and the development of grid services; 2) to illustrate the applications of the BIOPATTERN Grid, we present two pilot applications – use of the BIOPATTERN Grid for early detection of dementia and for assessment of brain injury.

The remainder of the paper is organised as follows. In Section 2, the BIOPATTERN Grid architecture and prototype are described. In Section 3, the detailed implementations of BIOPATTERN Grid services are presented. In Section 4, two pilot applications of BIOPATTERN Grid are illustrated. Section 5 concludes the paper.

2. BIOPATTERN Grid Architecture and Prototype

The architecture of BIOPATTERN Grid is divided in four layers as shown in Figure 1. The Grid Portal serves as an interface between an end user (e.g. a clinician) and the BIOPATTERN Grid network. The Grid services layer provides advanced services for data acquisition (including remote automatic data acquisition), data analysis & visualisation for brain diseases and cancer, and data query and/or information crawling services. The Grid middleware provides grid functionalities for security, resource management, information service, data management and data services support. The Globus Toolkit 4 (GT4) [5] is chosen to implement Grid middleware functions. Condor [6] is used for job queuing, job scheduling and to provide high throughput computing. The grid resources layer, contains computational resources, data resources, and knowledge resources (e.g. algorithms pool) and networks.

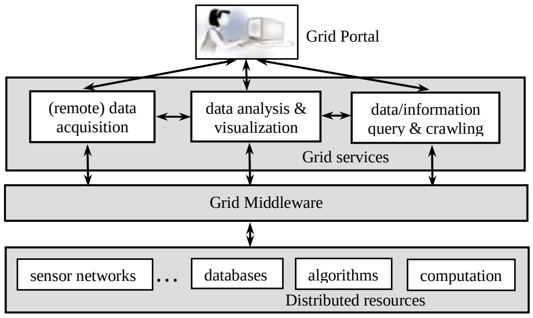

Figure 1. BIOPATTERN Grid Architecture

We have built a BIOPATTERN Grid Prototype within the BIOPATTERN Consortium to provide a platform for clinicians and researchers to share information in distributed bioprofile databases and computational resources. A major aim of the prototype is to facilitate analysis, diagnosis, and care for brain diseases and cancer. Currently, the prototype connects five sites –the University of Plymouth (UOP), UK; the Telecommunication System Institute (TSI), Technical University of Crete, Greece; the University of Pisa (UNIPI), Italy; Synapsis S.r.l. (Synapsis), Italy, and Tampere University of Technology (TUT), Finland (see Figure 2). Each site may hold bioprofile databases, Grid nodes, Condor pool, high performance cluster, algorithms pool, Grid portal, or an interface to remote data acquisition networks. For example, at University of Plymouth (UoP) node, it contains an algorithm pool including key algorithms for brain diseases analysis (e.g. fractal dimension algorithm for early detection of dementia and Independent Component Analysis based algorithm for assessment of brain injury); bioprofile databases which contain basic patient's clinical information, EEG data (awake EEG at resting state) for dementia, and EEG data for brain injuries); a web server which holds the BIOPATTERN Grid Portal; a condor pool, named PlymGRID, which contains up to 1400 nodes within the Plymouth campus.

3. Implementation of BIOPATTERN Grid Services

The BIOPATTERN Grid provides both high level and low level services. The high level services are implemented mainly for end users (e.g. clinicians or researchers), who have permissions to use specified grid-enabled services (e.g. to access distributed resources or to compare results from existing algorithms) without any grid knowledge (e.g. underlying grid techniques, locations of resources and detailed access constrains to resources). The low level services are implemented and provided mainly for grid services/applications developers to access BIOPATTERN Grid services directly, or to develop interfaces with BIOPATTERN Grid (e.g. the current AmI-Grid project within BIOPATTERN is seeking the integration of AmI's wireless data acquisition system with BIOPATTERN Grid).

3.1 High Level Services

Three high level services have been implemented in the BIOPATTERN Grid. As shown in Figure 3, the services are clinical information query, clinical information update and EEG analysis. The BIOPATTERN Grid Portal serves as the interface between an end user (e.g. a clinician or a researcher) to the BIOPATTERN Grid. It allows an end user to access these services from any where with Internet access (via a web browser).

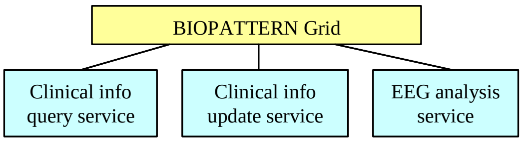

Figure 3. High level services provided by the BIOPATTERN Grid

3.1.1 Clinical Information Query Service

This service is designed based on a scenario that an end-user (e.g. a clinician) wants to query the information of a patient and to download the patient's EEG data for off-line analysis (the patient's information and EEG data are all distributed across the BIOPATTERN Grid network). To achieve this, the end-user has to go through certain steps via the BIOPATTERN Grid Portal, e.g. 1) Query for a specified patient; 2) Select EEG data files of the patient to download.

The query service is implemented in four levels from the top customer GUI to the bottom grid resources level as shown in Figure 4.

The customer GUI level provides interfaces to allow end-users to seamlessly access grid-enabled services with graphic visions. The query interface and file download interface are designed to enable users to access required clinical information without any grid knowledge, and to allow users to simply follow the defined steps to use this service. The functionalities provided at this level include handling HTTPS requests (e.g. obtainment of requests from end-users by filling e-forms), forwarding such requests to the lower level, etc. JSP and html are languages used to implement these interfaces, which are held by Tomcat containers.

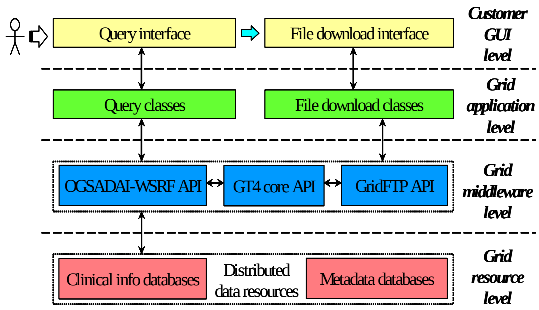

Figure 4. Implementation of clinical information query service

For the grid application level, two types of classes: query classes and file download classes, are responsible to translate specified requests from the interfaces in the customer GUI level into detailed action requisitions and further send those requisitions to grid servers for execution. The query classes can provide concurrent access of distributed databases via OGSADAI-WSRF in order to obtain information as descriptions of specified patients, EEG data locations, etc. The file download classes mainly offer file transfers from the grid nodes which contain specified EEG data to the grid nodes which can be accessed by end-users.

The grid middleware level contains various APIs, including OGSADAI-WSRF API, GT4 core API, GridFTP API, etc. All these APIs are responsible for dealing with action requisitions that come from the grid application level and realising such requisitions by accessing distributed data resources.

The grid resource level holds distributed data resources, which are described as low level services held by Globus containers. Data resources that support the query service includes clinical information databases (e.g. database for personal information of patients and EEG data), and metadata databases (e.g. database for descriptions of EEG data).

3.1.2 Clinical Information Update Service

This service is designed based on a scenario that an end-user (e.g. a clinician) needs to update an existing patient's clinical records (e.g. a new EEG recording has been taken and needs to be uploaded to the distributed bioprofile databases or a new patient's clinical information has to be added to the bioprofile databases). To achieve this, the end-user needs to go through the following steps via the Portal, e.g. 1) Add a new patient and his/her information, or query for an existing patient; 2) Upload EEG data files of the patient.

Similar to the query service, the update service is also implemented in four levels, as shown in Figure 5.

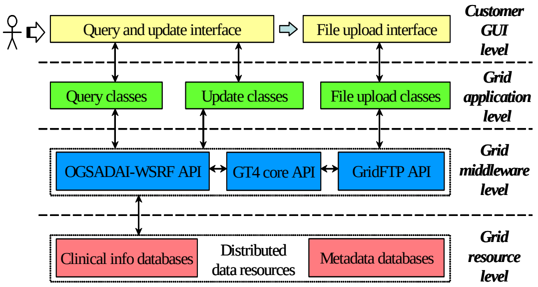

Figure 5. Implementation of clinical information update service

For this service, we have implemented both query and update interfaces in the customer GUI level in order to allow end-users to be able to easily insert information of new patients to the databases or to find a specified patient to update his/her records.

3.1.3 EEG Analysis Service

This service is designed based on a scenario that an end-user (e.g. a clinician) wants to use automated analysis services provided by BIOPATTERN Grid to help for the detection/diagnosis of dementia or to help for the assessment of brain injury (two services are provided currently).

To achieve this, end-users have to go through several steps via the portal, e.g. 1). select for analysis either for dementia or for brain injury; 2) query for a specified patient; 3) select the patient's EEG data files for analysis and submit analysis jobs; 4) choose the way of viewing analysis results (e.g. in canonograms and/or bargraphs for dementia).

The EEG analysis service is also implemented in four levels, as presented in Figure 6.

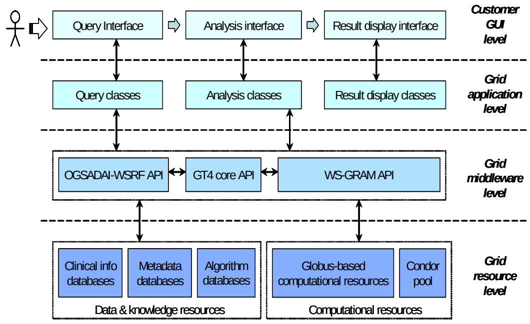

Figure 6. Implementation of EEG analysis services

In this service, the customer GUI level provides three user friendly interfaces, as query, analysis and result display interfaces, which are connected with their corresponding types of Java classes at the grid application level, as presented in Figure 6. For those Java classes, the analysis classes mainly offer generation of grid jobs (e.g. generation of Resource Specification Language files for job description), finding appropriate computational grid resources (e.g. looking for light-loaded resources based on job requirements and resource information), and submission of jobs, retrieving analysis results and forwarding the results to the high level for display. The results display classes enable end-users to view such analysis results in graphical display. The direct connection between result display classes and grid middleware level will be established in the future in order to support more complex functionalities, such as 3D visualisation.

Different to the other two high level services, the EEG analysis service holds both computational and data resources. The computational resources cover both High Throughput Computing (HTC) and High Performance Computing (HPC) resources. The Condor pool, currently plays a key role in HTC-based EEG analysis. The Globus-based computational resources support the HPC-based EEG analysis at present.

3.2 Low Level Services

There are two types of low level services which have been considered in the implementation, BIOPATTERN data services and computational services.

3.2.1 BIOPATTERN Data Services

The goal of the provision of the data services is to support certain features in clinical data access and integration over distributed heterogeneous data resources, such as database federation and data transformation.

Thus far we have mainly investigated into the area of database federation and implemented some basic data services, such as the generic query and update services, based on OGSADAI-WSRF, an OGSADAI distribution compliant with WSRF specifications.

Figure 7 presents a three-layer architecture for the implementation. In the architecture, the top layer houses all BIOPATTERN data services, which can provide functionalities, such as data access and transformation to permitted and authorized grid users based on specific requirements. The middle layer that holds the OGSADAI grid data service acts as an interface between the top and bottom layers to make the underlying data sources be accessible to those grid users.

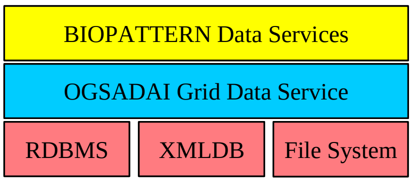

Figure 7. Low level data services architecture

The design of this architecture enables grid users to access different types of data resources within the BIOPATTERN Grid in a consistent and resource-independent way and also offers the flexibility for further development in areas like data integration, security, access constrains, etc.

In the implementation, both BIOPATTERN data services and OGSADAI grid data service are built as Web services and held by Globus containers.

3.2.2 BIOPATTERN Computational Services

The computational services are designed to support those computation intensive applications, such as medical image analysis and protein sequence comparison, and mainly implemented based on WS-GRAM.

On any BIOPATTERN grid nodes, a grid user can use this type of services to submit jobs to any permitted HPC and/or HTC resources and then get expected results back. The completion of a grid job may require several steps, i.e. 1) creation of the job; 2) stage in (e.g. transmission of input files); 3) execution of the job; 4) stage out (e.g. transmission of output files); 5) job cleanup.

Currently, all computational services are held by Globus containers located at certain grid nodes (e.g. nodes at UoP and TSI) in the BIOPATTERN Grid.

4. Applications of the BIOPATTERN Grid

In line with the Grand Vision of the BIOPATTERN project, the applications of the BIOPATTERN Grid are sought mainly around the areas of brain diseases and cancer. Currently we have developed two applications on early detection of dementia and assessment of brain injury. These applications are accessible via the BIOPATTERN Grid Portal and are developed for proof of the concept. It will be used as a vehicle for future clinical validation.

4.1 BIOPATTERN Grid for Early Detection of Dementia

Dementia is a neurodegenerative cognitive disorder that affects mainly elderly people. Several objective methods are available that may support early diagnosis of dementia. Among others, the EEG which measures electrical activities of the brains is regarded as an acceptable and affordable method in the routine screening of dementia in the early stages. Different biodata analysis methods (e.g. Fractal Dimension or FD) are developed which try to provide a biomarker (e.g. FD index) to indicate possible on-set of dementia. Using current clinical criteria, delay between the actual onset and clinical diagnosis of dementia is typically 3 to 5 years. Serial EEG recordings are normally suggested for elderly people in order to diagnosis dementia at an early stage [7].

Due to mobility of a citizen, a patient's serial EEG recordings can be located in different clinical centres and possibly across different countries. We envisage that grid computing can play a role in early detection of dementia when dealing with geographically distributed data resources.

To illustrate the concept of BIOPATTERN Grid for early detection of dementia, a hypothetical patient pool consisting of 400 subjects, each with three EEG recordings was created. These data are hypothetical representation of recordings taken at three time instances akin to longitudinal studies carried out in reality. The datasets are distributed at TSI, TUT and UoP sites. The FD analysis algorithm is used to compute the FD of each dataset.

Through the Portal, an end-user (e.g. a clinician) can select a patient, e.g. Mike, and the algorithm is used to perform the analysis. Upon submission, Mike's information, including his serial EEG recordings (located in UoP, TSI and TUT, respectively) are retrieved and analyzed. Results can be shown in canonograms (see Figure 8) where changes in the EEG indicating Mike's conditions. The canonograms (from left to right) show the FD values of the Mike's EEG taken at time instances of 1 (data at TSI), 2 (data at TUT) and 3 (data at UoP) respectively. The FD values for the left canonogram indicates Mike in a normal condition with high brain activity, whereas the FD value for the right canonogram indicates Mike in a probable Alzheimer Disease with low brain activity. The middle one shows the stage in between. Changes in the FD values provide some indication about the disease progression. This can help clinicians to detect dementia at an early stage, to monitor its progression and response to treatment.

4.2 BIOPATTERN Grid for assessment of brain injury

The second application of the BIOPATTERN Grid is used to assess the severity of brain injury by analysing evoked potentials (EPs) buried in raw EEG recordings. The assessment process consists of two stages: 1) an Independent Component Analysis (ICA) [8] algorithm is used to extract single trial evoked potential activity of clinical interest and discard irrelevant components such as background EEG and artefacts; 2) a LORETA (LOw Resolution Electromagnetic TomogrAphy) [9] is used to localise the source of brain activity. These methods are implemented in Matlab and compiled as executable (using the Matlab run-time library) and distributed over the Condor pool to speed up the time-consuming analysis. Figure 9 shows an example of results of such analysis via the portal, with topography maps (only for the two first ICA components) of one normal patient (top) and one patient (bottom). The head model, which is shown, is assumed to be an 8cm sphere and is represented with 8 slices (the lower slice is the left and the higher is the right). Through this service, a clinician can easily use the advanced EEG analysis algorithm (e.g. ICA-LORETA) to analyze a patient's raw EEG data for assessing brain injury.

5. Conclusion

In this paper, we have presented a new testbed, BIOPATTERN Grid which aims to share geographically distributed computation, data and knowledge resources for combating major diseases (e.g. brain diseases). We illustrated the development of BIOPATTERN Grid Prototype, the implementation of the BIOPATTERN Grid services and two pilot applications for early detection of dementia and for assessment of brain injury.

BIOPATTERN Grid is an ongoing project and results presented here are limited in scale. In the near future, the BIOPATTERN Grid prototype will be extended to include more grid nodes (both on Globus and on Condor), more computing resources (e.g. connecting with HPC clusters in UNIPI), more grid applications (e.g. to include cancer diagnosis and prognosis) and services (e.g. data transformation), integration with other projects within the BIOPATTERN, such as to integrate with AmI-Grid project [10] for connection with wireless data acquisition networks and to integrate with grid.it [11][12] to provide crawling services for information query. In the long term, we also seek the integration of the BIOPATTERN Grid with other regional or national Grid projects/networks (e.g. EU's EGEE).

Due to the nature of healthcare, the BIOPATTERN Grid will need to address several issues such as regulatory, ethical, legal, privacy, security, and QoS, before it can move from research prototype to actual clinical tool.

Acknowledgement

The authors would also like to thank Dr. C. Bigan from EUB, Romania for providing the original EEG data sets and Dr. G. Henderson for providing the algorithms for computing the FD index. We acknowledge the financial support of the European Commission (The BIOPATTERN Project, Contract No. 508803) for part of this work.

References

[1]. S. R. Amendolia, F. Estrella, C. D. Frate, J. Galvez, W. Hassan, T Hauer, D Manset, R McClatchey, M Odeh, D Rogulin, T Solomonides and R Warren, "Development of a Grid-based Medical Imaging Application", Proceedings of Healthgrid 2005, from Grid to Healthgrid, 2005, pp.59-69.

[2]. S. Lloyd, M. Jirotka, A. C. Simpson, R. P. Highnam, D. J. Gavaghan, D. Watson and J. M. Brady, "Digital mammography: a world without film?", Methods of Information in Medicine, Vol.44, No. 2, pp. 168-169, 2005.

[3]. J. S. Grethe, C. Baru, A. Gupta, M. James, B. Ludaescher, M. E. Martone, P. M. Papadopoulos, S. T. Peltier, A. Rajasekar, S. Santini, "Biomedical Informatics Research Network: Building a National Collaboratory to Hasten the Derivation of New Understanding and Treatment of Disease", Proceedings of Healthgrid 2005, from Grid to Healthgrid, 2005, pp. 100-109.

[4]. K. Ichikawa, S. Date, Y. Mizuno-Mastumoto, and S. Shimojo, "A Grid-enabled System for analysis of Brain Function", In Proceedings of CCGrid 2003 (3rd IEEE/ACM International Symposium on Cluster Computing and the Grid), May 2003.

[5]. Foster, "Globus Toolkit Version 4: Software for Service-Oriented Systems", Proceedings of IFIP International Conference on Network and Parallel Computing, 2005, pp. 2-13.

[6]. http://www.cs.wisc.edu/condor/hawkeye

[7]. G. T. Henderson, E. C. Ifeachor, H. S. K. Wimalartna, E. Allen and N. R. Hudson, "Prospects for routine detection of dementia using the fractal dimension of the human electroencephalogram", MEDSIP00, pp. 284-289, 2000.

[8]. T-W Lee, M. Girolami, TJ. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources". Neural Computation 1999;11(2):606-633.

[9]. R. D. Pascual-Marqui. "Review of methods for solving the EEG inverse problem" International Journal of Bioelectromagnetism 1999, 1: 75-86

[10]. M. Lettere, D. Guerri, R. Fontanelli. "Prototypal Ambient Intelligence Framework for Assessment of Food Quality and Safety", 9th Int. Congress of the Italian Association for Artificial Intelligence (AI*IA 2005) – Advances in artificial Intelligence, pp. 442-453, Milan (Italy), Sep. 21 - 23, 2005

[11]. Grid.it: "Enabling Platforms for High-Performance Computational Grids Oriented to Scalable Virtual Organizations", http://grid.it/.

[12]. K. Cerbioni, E. Palanca, A. Starita, F. Costa, P. Frasconi, "A Grid Focused Community Crawling Architecture for Medical Information Retrieval Services", 2nd Int. Conf. on Computational Intelligence in Medicine and Healthcare, CIMED'2005.

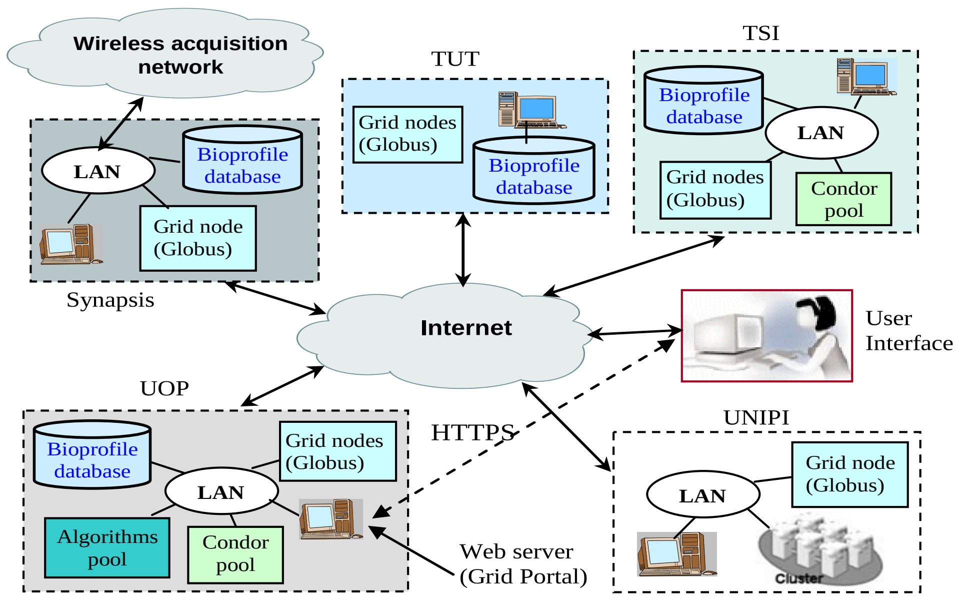

Figure 2. BIOPATTERN Grid Prototype

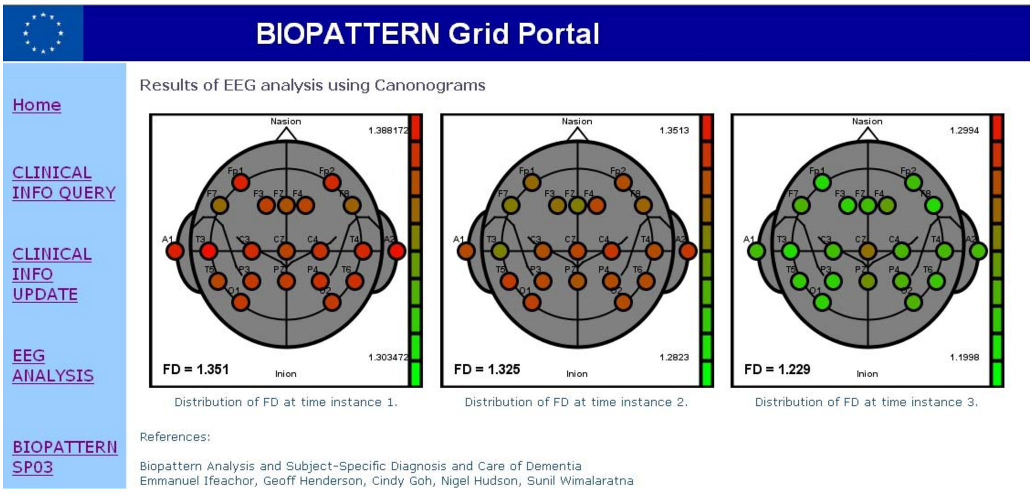

Figure 8. Canonograms showing the distribution of FD values for EEG analysis for dementia

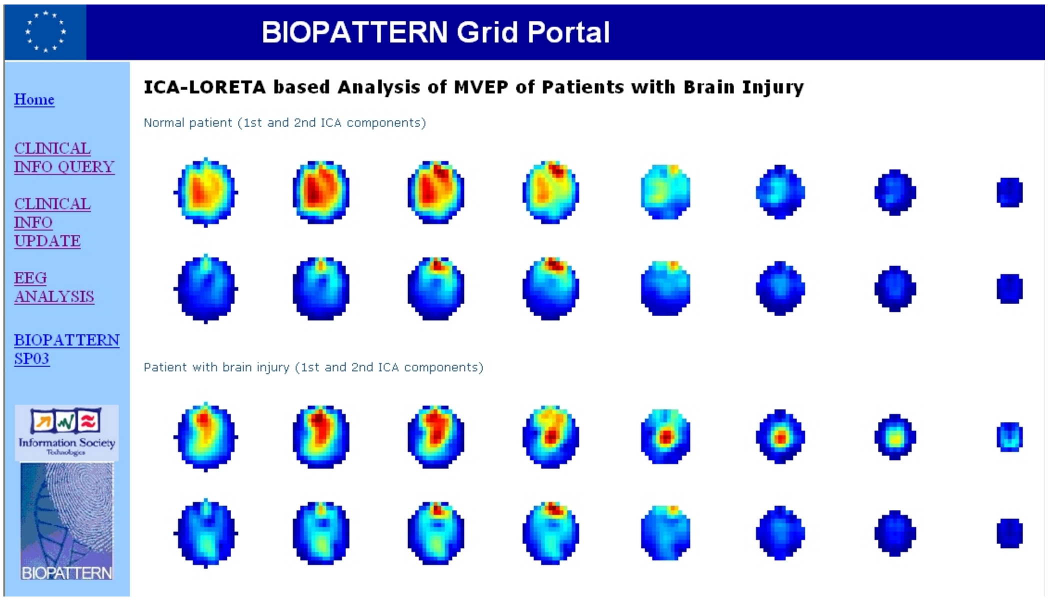

Figure 9. Topography maps of normal subjects and patients with brain injury