

Grid infrastructures for computational neuroscience: the neuGRID example

Alberto Redolfi[†], Richard McClatchey, Ashiq Anjum, Alex Zijdenbos, David Manset, Frederik Barkhof, Christian Spenger, Yannik Legré, Lars-Olof Wahlund, Chiara Barattieri di San Pietro & Giovanni B Frisoni

[†]Author for correspondence: Fatebenefratelli – Centro San Giovanni di Dio, Laboratory of Epidemiology & Neuroimaging, Via Pilastroni 4, I-25125 Brescia, Italy ■ Tel.: +39 030 350 1311 ■ Fax: +39 030 350 1313 ■ aredolfi@fatebenefratelli.it

Neuroscience is increasingly making use of statistical and mathematical tools to extract information from images of biological tissues. Computational neuroimaging tools require substantial computational resources and the increasing availability of large image datasets will further enhance this need. Many efforts have been directed towards creating brain image repositories including the recent US Alzheimer Disease Neuroimaging Initiative. Multisite-distributed computing infrastructures have been launched with the goal of fostering shared resources and facilitating data analysis in the study of neurodegenerative diseases. Currently, some Grid- and non-Grid-based projects are aiming to establish distributed e-infrastructure, interconnecting compatible imaging datasets and to supply neuroscientists with the most advanced information and communication technologies tools to study markers of Alzheimer's and other brain diseases, but they have so far failed to make a difference in the larger neuroscience community. NeuGRID is an European commission-funded effort arising from the needs of the Alzheimer's disease imaging community, which will allow the collection and archiving of large amounts of imaging data coupled with Grid-based algorithms and sufficiently powered computational resources. The major benefit will be the faster discovery of new disease markers that will be valuable for earlier diagnosis and development of innovative drugs. The initial setup of neuGRID will feature three nodes equipped with supercomputer capabilities and resources of more than 300 processor cores, 300 GB of RAM memory and approximately 20 TB of physical space. The scope of this article is highlights the new perspectives and potential for the study of the neurodegenerative disorders using the emerging Grid technology.

Evolution of brain imaging in neurodegenerative diseases

Brain imaging was regarded as an elective examination in patients with cognitive decline 15 years ago [1]. The practice parameters for diagnosis and evaluation of dementia defined by the American Academy of Neurology regarded computed tomography (CT) and magnetic resonance (MR) as 'optional' assessments [2,3]. Over time, imaging in dementia has moved from a negative, exclusionary role to one that added positive diagnostic and prognostic information. In the late 1990s, the traditional exclusionary approach was abandoned in favor of the inclusive approach [4,5]. Rapid advances in neuroimaging technologies such as PET, single photon emission CT, MR spectroscopy, diffusion tensor imaging and functional MRI have offered new vision into the pathophysiology of Alzheimer's disease (AD) [6] and, consequently, increasingly new powerful data-analysis methods have been developed [7].

Since the beginning of the 21st Century, the development of innovative techniques for region-of-interest-based volumetry, automated voxel-based morphometry, cortical thickness measurement, basal forebrain volumetry and multivariate statistics have emerged [7–9] and those measurements most feasible and accurate have started to be used in clinical settings. The availability to the neuroimaging community of large prospective image data repositories has led to the development of web-based interfaces to access data and online image analysis tools to assess longitudinal brain changes [10–13].

With the development of novel analysis techniques, the computational complexity of neuroimaging analysis has also increased significantly. Higher spatial resolution images and longer time scans are being acquired so that more voxels will need to be processed for each acquisition. The same applies to the computational resources required by algorithms, since these have become increasingly central processing

Keywords

- Alzheimer's disease
- Grid architecture ■ medical imaging ■ neuGRID
- parallel computing
- remote experiment

unit (CPU)-intensive. Unfortunately, many medical imaging facilities do not currently have the necessary computational resources – usually expensive and difficult to maintain – in order to satisfy the computational demand of the most advanced neuroimaging analysis. To exploit new neuroimaging algorithms, the current working paradigm is that scientists physically migrate with small personal datasets (of a few hundred images at most) to centers of excellence where they can find expertise and computational resources. Typically, a research fellow spends 3 months or more at an image analysis center where he/she learns the use of those algorithms on personal image data. Then he/she returns to the original research group where the procedure has to be installed in whole or in part and, finally, runs jobs either in house or remotely on the image analysis center servers and mainframes (FIGURE 1).

This scenario, which has so far been sustainable, will need to change radically in the near future when thousands of images will become available [13–17]. Under these circumstances, the neuroimaging community will need an efficient distributed infrastructure where high-performance computing and innovative customizable algorithms will be available, together with access to the large image datasets that are currently being collected worldwide (FIGURE 2) [18–20].

In the following section, we will describe some of the advantages of Grid infrastructures and outline in detail the architecture and structure of an innovative infrastructure that is currently under development, neuGRID [101], funded by the European Commission under the seventh Framework Programme. Finally, specific regard will be devoted to applications that might hugely benefit from Grid technology for the study of neurodegenerative diseases.

Recent decentralized neuroscientific infrastructures

A number of efforts at creating large image repositories exist worldwide and tools have been developed to manage the increasing wealth of image data with which biomedical scientists will soon need to cope. For example, in the USA, the largest ever project in the field of neurodegenerative disorders (the North American Alzheimer’s Disease Neuroimaging Initiative [ADNI]) [21] assesses early AD patients every 6 months for 4 years with different imaging techniques, allowing the collection of over 5000 brain imaging studies. In Europe, the feasibility of the adoption of the ADNI platform has been evaluated in the Foresight Study for the Development of a European NeuroImage Repository (FP6 ENIR) [22], in the data collection study pilot European ADNI [23] and in the Innomed-AddNeuroMed (FP6) [24].

AddNeuroMed has studied 900 persons at baseline, 3 and 12 months and facilitated the collection of 2500 brain imaging studies. Notably, these images are acquired using the ADNI protocol and, with the ADNI 5000 images, thus created a dataset of approximately 7500 brain imaging studies overall. Moreover, the ADNI data are public and access to that data will be granted to any scientist wishing to exploit it. The combination of larger datasets and larger scientific communities make research environments necessary with efficient archiving systems as well as powerful computational facilities.

Neurobase is another initiative that aimed to promote the federation of distributed information databases in neuroimaging, based on a distributed architecture [25]. Interestingly, the clinical and imaging data of Neurobase is not open to the public but only reserved to partners.

The US-NIH MRI Study of Normal Brain Development (NIHPD) has produced a definitive database of normal child brain development images [26,27]. Over 500 children from newborn to 18 years old were followed with repeated MRI and clinical/behavioral testing.

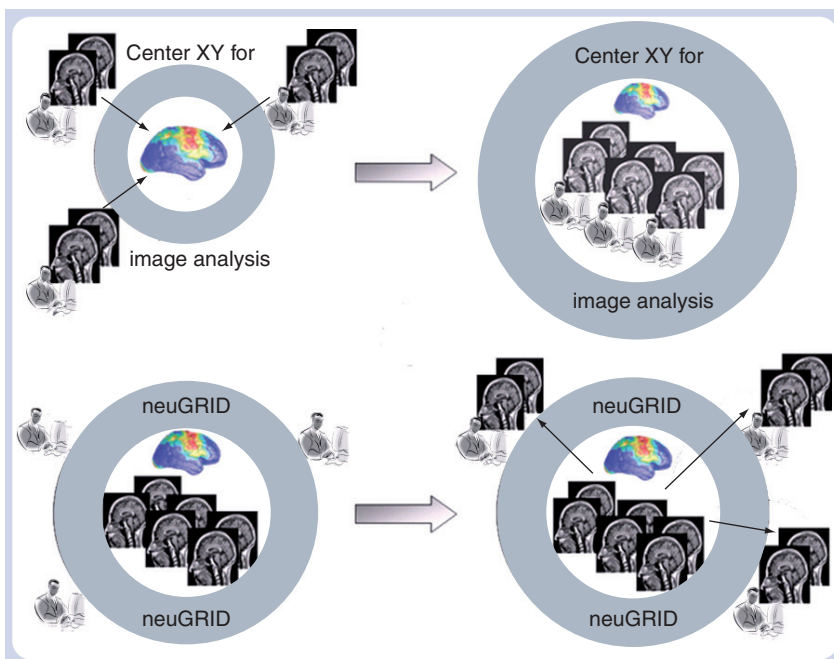


Figure 1. How science is carried out today in the field of computational neuroscience. Scientists physically migrate to imaging centers where they can find expertise and computational facilities (upper section of the panel). Distributed archive and distributed analysis will be available through the Grid infrastructures. Grids will bring the image analysis center in the laboratory for each individual neuroscientist (bottom section of the panel).

The database system developed in the course of this project (Online Research Imaging System [LORIS]) has been generalized for use in many other studies, such as the previously described Innomed-AddNeuroMed. The web-based system developed and used in this study is one of very few such systems that have been proven to work effectively. It has been used for the MRI Study of Normal Brain Development for almost a decade and was deployed in the Innomed-AddNeuroMed project with minimal technical challenges resulting from the transition from the pediatric to elderly populations. LORIS has been specifically designed to facilitate the collection of imaging as well as clinical/behavioral data in the context of longitudinal, multicenter studies. It is highly modular and configurable, such that new scans, test instruments, subprojects or workflows can be added easily and rapidly [27]. Owing to its simplicity and reliability, it has been chosen as the foundation of the neuGRID database.

In parallel with these increasingly sophisticated neuroimaging datasets, some first rudimentary environments have been developed for the advanced management and use of large repositories of biomedical images. The project Towards a Grid Environment to Process and Share DICOM Objects (TRENCADIS) developed a software platform comprising a set of services as a solution for interconnecting, managing and sharing medical (DICOM) data for the development of training and decision support tools [28]. The Web Interfacing Repository Manager (WIRM) is an open-source toolkit that allows the creation of web applications that facilitate the acquisition, integration and dissemination of biomedical data over the web [29]. The system is able to support the collaboration of computer scientists, neuroscientists, radiologists and other clinical professionals, enabling them to share experimental data and work together on a distributed software system for the study of the brain. WIRM was funded by the US National Institute of Mental Health, it is a Perl-based application server running on the Linux, Apache, MySQL and Perl (LAMP) platform and its access is public.

Other brain MRI datasets specific to the evaluation of the neuroimaging tools have been developed. The IBSR is a world wide web resource providing access to brain MRI data and segmentation results [102].

As the interest in quantitative analysis of medical image data is growing, the need for the validation of techniques is increasing too. The BrainWeb portal contains a set of realistic

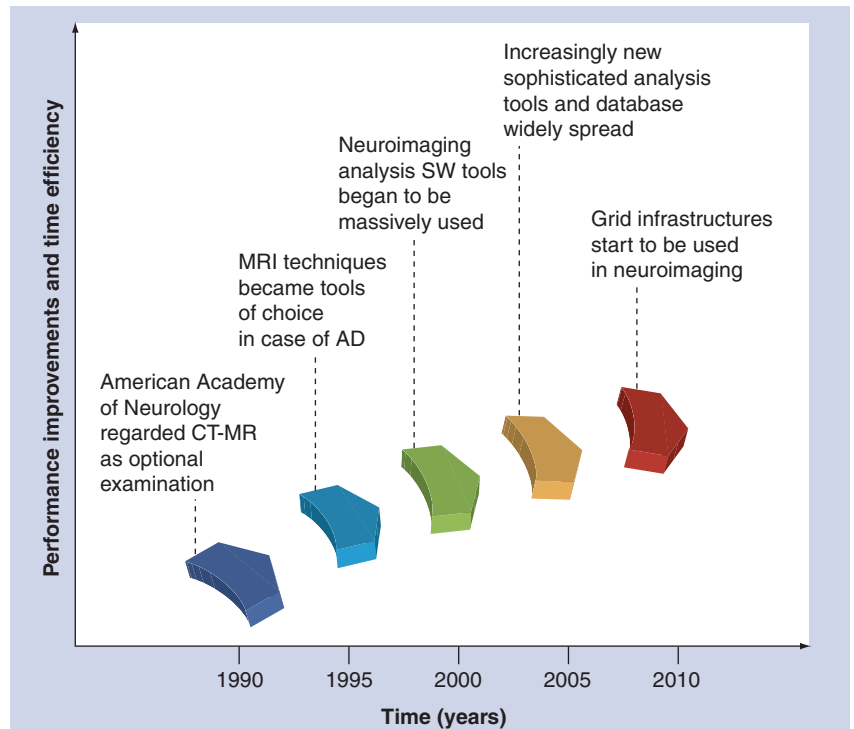


Figure 2. Evolutionary steps of neuroimaging analysis tools from the early 1990s to the actual exploitation of the Grid infrastructure. Grids provide innovative and more efficient data analysis among a large number of centers for early diagnosis and disease-marker discovery. AD: Alzheimer's disease; CT-MR: Computed tomography-magnetic resonance; SW: Software.

MRI data volumes produced by an MRI simulator [30]. These data can be used by the neuroimaging community to evaluate the performance of various image-analysis tools.

Notably, computing facilities are included in none of the aforementioned environments. TABLE 1 reports the initiatives mostly aimed at the management of neuroimaging data, but none of which offers computational facilities. Since 2001, many national institutions have started to fund projects to address the absence of advanced computing facilities available to neuroscience research, which has resulted in the NeuroGrid [31], PsyGrid [32], NeuroPsyGrid [33,34], NeuroLOG [35] and Bing [36] projects. All these initiatives, characterized exclusively by national extensions, were developed with the objective of validating small Grid-based networks, promoting data sharing and data combination via specific-domain ontology, exploiting new tools of analysis for different neuropathologies and defining new methodology of analysis.

The Biomedical Informatics Research Network (BIRN) is probably the best known of such efforts; BIRN aims to enable a software fabric for participating centers and to facilitate the collaborative use of domain tools and flexible

Table 1. Largest neuroimaging repositories developed worldwide.

Project	Aim	Date; funding body	Ref.
Alzheimer's disease Neuroimaging Initiative	ADNI aims to validate prospective imaging (MRI and PET) and biological markers for use in clinical trials of anti-amyloid drugs. It recruits and follows 600 patients for 5 years with very early to early Alzheimer's disease and 200 normal controls in approximately 60 clinical and research centers. Strong emphasis is devoted to accurate harmonization of the methods for collecting images and biological fluids	2004; US National Institute on Aging and Industry	[114]
Innomed-AddNeuroMed	The AddNeuroMed objectives are to find new biomarkers for AD. It recruits 900 patients and aims to collect 2500 brain images	2005; EU – FP6 (EFPIA study)	[115]
Pilot E-ADNI	The aim of this pilot project is to demonstrate that the core ADNI methodology (i.e., standardized and centralized collection of MRI, clinical data, blood and CSF samples) can be adopted by EADC	2006; Grant from the Alzheimer's Association	[116]
ENIR	ENIR aims to develop a large and shared European multidimensional repository of high resolution MRI of normal brains and brains with different neurodegenerative disorders (e.g., Alzheimer's, amyotrophic lateral sclerosis and Parkinson's disease) completed by clinical, genetic and neuropsychological data	2006; EU – INFRASTR-6 (FP6)	[117]
NIHPD	NIHPD uses MRI to further the knowledge of brain maturation in healthy, typically developing infants, children and adolescents; to correlate brain development with cognitive and behavioural development; to provide essential data for understanding the course of normal development as a basis for understanding diseased brain development. NIHPD enrolled over 500 children from newborn to 18 years of age and studied longitudinally at least three timepoints with a large battery of clinical and behavioural assessments. Data were collected at six Pediatric Study Centers	2001; NIH (NIBIB, NIMH, NIDA, NCI, NIA and NINDS)	[118]
Neurobase	This aims to establish the federation of distributed information databases in neuroimaging. This project consists of specifying how to connect and access distributed information databases through a datamining approach	2005; Grant from French Ministry of Research	[119]
ISBR	ISBR provides manually guided expert segmentation results along with brain MRI data. Its purpose is to encourage the evolution and development of segmentation methods	1996; National Institute of Neurological Disorders and Stroke	[102]
BrainWeb	A database of simulated brain imaging data, specifically useful for the evaluation of algorithm performance	Canadian Institutes of Health	[120]
<small>ADNI: Alzheimer's Disease Neuroimaging Initiative; CSF: Cerebrospinal fluid; EADC: European Centres of the Alzheimer's Disease Consortium; ENIR: European NeuroImage Repository; NIHPD: NIH MRI study of Normal Brain Development; NINDS: National Institute of Neurological Disorders and Stroke; ISBR: Internet Brain Segmentation Repository.</small>			

processing/analysis frameworks for the study of brain diseases [37,38]. The BIRN architecture is Grid based and comprised of approximately 20 Grid hosts around the USA. Notably, the BIRN research environment has not yet had a huge impact on the neuroscience community, despite having been in place for the past 8 years. BIRN has been designed as a horizontal facility that could accommodate the needs of a large number of different communities requiring biomedical informatics resources (designed from the general to the specific). The BIRN project was conceived with a very broad scope, based on the idea that many research institutes would start to share and collaborate on a common set of data. Different causes, including the large number of sites involved (19 universities and 26 research centers), different infrastructures, own data formats and intrinsic difficulties to motivate the sharing of the data, have all contributed to a significant slow down in the final objectives achievement and a significant loss of cohesion.

The Montreal Neurological Institute (MNI) is developing a pan-Canadian platform termed CBRAIN [103], using a Grid with a Service-Oriented Architecture (SOA) for distributed processing and exchange of 3D/4D brain imaging data. The CBRAIN initiative makes extensive use of web services to support new dynamic appliance provisioning as well as to interface with a number of existing processing tools. Using the innovation brought by this Grid platform, it should be possible to apply complex algorithm pipelines to large databases stored in different locations and visualize the results as 3D brain maps in real time.

At the University of California in Los Angeles (UCLA; LA, USA), the Laboratory of NeuroImaging (LONI) allows the specification and execution of complex pipelines utilizing existing processing and visualization tools onto given data sets. The LONI Pipeline project [39] has a powerful client-server architecture based on a Sun Grid Engine (SGE), which uses the Distributed Resources Management Application (DRMAA) that enables the sequencing and execution of heterogeneous software modules developed by numerous groups around the world. The execution can be performed locally or via connection to the LONI facilities. The graphical interface provides users with a simple, intuitive and flexible system. It handles all connections, conversions and other housekeeping between software modules. Finally, it connects to the LONI Image Data Archive (IDA), in which the ADNI dataset is stored, among others.

The initiatives reported in TABLE 2 demonstrate that Grid-based solutions might be an appropriate answer to the archiving and computational needs of neuroscientists [40–42]. Indeed, Grid-based solutions to share data and carry out advanced computational analyses are spreading fast but very few have international capacity.

NeuGRID [101] is the first e-infrastructure project at a European level including both European neuroimaging research centers and information technology (IT) companies (FIGURE 3). It aims to provide the European neuroscience community with services such as the collection/archiving of large amounts of imaging data paired with computationally intensive analyses. The neuGRID initiative takes a focused and centralized approach, which we hope will greatly increase the Grid's productivity. Rather than trying to connect existing disparate datasets and codebases together, a Grid-wide, coherent database and image-processing structures are implemented, moving directly from the need of a neuroscience community. Biomedical scientists will be able to identify neurodegenerative disease markers through the analysis of 3D brain MRI via the provision of sets of both distributed medical and Grid services. Once this has been successfully accomplished, the infrastructure could be generalized to other medical application areas (designed from the specific to the general).

NeuGRID's Grid

The neuGRID project started in February 2008 and will last for 3 years. The neuGRID infrastructure is designed to be expandable and it will be compliant with international standards for data collection [104] and data management. Three typical neuGRID scenarios are:

- Clinical data/images collected from subjects considered to be specifically suited to enter in the neuGRID infrastructure (e.g., US-ADNI or AddNeuroMed data).
- Clinical data/images previously collected in different research projects that will be archived in neuGRID (e.g., Pilot E-ADNI, Italian ADNI and other similar datasets acquired using the ADNI protocol).
- Clinical data/images previously collected in different research projects that will be analyzed but not archived in neuGRID (e.g., projects founded by 'Big Pharma' or biotechnology industries to test the efficacy of new drugs in some trials of Phase II or III).

Table 2. Grid-related initiatives in the neurosciences worldwide.

Project	Aim	Date; funding body	Country	Ref.
Birn	The Birn goal is to design and implement a distributed architecture of shared resources usable by all biomedical researchers in order to advance the diagnosis and treatment of disease	2001; National Institutes of Health's National Centre for Research Resources (NCRR)	USA; UK	[121]
MedGrid	The Medical Grid Project aims to provide cost-effectiveness in the dissemination and maintenance of high performance analysis system; robustness in the analysis by integrating different methodologies; and prompt reporting capability based on real-time processing	2003; ICT-Asia Program	Japan; The Philippines	[122]
NeuroGrid	NeuroGrid aims to enhance cooperation between clinical, imaging and e-scientists to create a Grid-based network of neuroimaging centers and a neuroimaging tool-kit. Sharing data, experience and expertise facilitate the archiving, curation, retrieval and analysis of imaging data from multiple sites and enable large-scale clinical studies	2005; Medical Research Council	UK	[123]
PsyGrid	The PsyGrid objective is to build an electronic database to hold anonymized clinical data regarding people presenting to NHS professionals with first episode psychosis	2007; Medical Research Council	UK	[124]
NeuroPsyGrid	NeuroPsyGrid aims to combine datasets, extend work on a shared metadata model, define an ontology of terms for psychosis and develop Grid-based data analysis on patients with first episode psychosis. The NeuroPsyGrid project is built on two overlapping projects: NeuroGrid and PsyGrid	2007; Medical Research Council	UK	[125]
Bing	The Bing goals are to develop an information technology infrastructure to support collaborative use and sharing of neuroimaging data collections, analysis and modeling software and visualization tools; provide a 'neuroscientist-friendly' web portal (termed Brainimaging.pt) to enable secure access to these tools and training in their use; and to encourage scientific collaborations among participants from different research institutions that work independently	2006; Fundação para a Ciência e Tecnologia	Portugal	[126]
NeuroLOG	neuroLOG objectives are the management and access of partially structured data; the control of workflows implied in complex computing process on Grid infrastructures; and the extraction and quantification of relevant parameters for brain pathologies	2007; National Agency for Research	France	[127]
NeuGRID	NeuGRID aims to provide a user-friendly Grid-based research e-infrastructure enabling the European neuroscience community to carry out research required for the study of degenerative brain diseases. The collection and archiving of large amounts of images will be paired with computationally intensive data analyses	2007; EU – FP7	Europe	[101]
CBrain	The CBrain goal is to develop a Canadian network of the five leading brain imaging research centers. The goal is to create a platform for distributed processing and exchange of 3D–4D brain imaging data. The expected result is a platform that will render the processing environment transparent to a remote user in order to apply complex algorithm 'pipelines' to large databases	2008; CANARIE funding awards	Canada	[103]

NeuGRID exploits the experience developed during the earlier FP5 MammoGrid project (this provided knowledge related to the middleware and upperware that allowed specific applications to 'talk' to the Grid while remaining independent) [43–48], and the AddNeuroMed study (the collection, archiving and retrieval of multicenter clinical data, biomedical images and computerized image analysis) from the FP6 InnoMed project.


The design philosophy underpinning neuGRID is one that embodies the principles of reuse, flexibility and expandability. A layered and service-oriented approach (SOA) is followed to address these requirements. This approach enables a separation of concerns between the details of the application services (brain imaging) and the Grid deployment infrastructure. Different services will be delivered to satisfy the specific requirements of neuroscientists but will be designed and implemented to be flexible in nature and reusable in application. In this manner, a set of generic services will glue a wide range of user applications to available Grid platforms. This will result in the production of a system that addresses specific applications but that retains a

large degree of underlying generality, thus being able to cope with the still rapidly changing Grid environment. A schematic diagram of this layered approach is shown in FIGURE 4.

NeuGRID's key research challenges are: the gridification of the most advanced neuroimaging algorithms starting from the Constrained Laplacian Anatomic Segmentation using Proximity (CLASP; cortical extraction software developed at Montreal Neurological Institute) cortical thickness extraction algorithm [49–51] as proof-of-principle (FIGURE 5), the development of a middle layer of generic services to make the infrastructure expandable to a huge number of neuroimaging pipelines and the ability to offer a Grid-based easy-to-use set of tools with which scientists could transparently perform analyses and collaborate internationally.

The neuGRID infrastructure consists of two main levels (plus eventually one additional layer), each of which has a specific role (FIGURE 6).

Level 0: this is the upper level of the Grid, which hosts the two main nodes of the entire Grid: the Data Coordination Centre (DCC) and the Grid Core Centre (GCC).



	NeuGRID	Country	Typology	Roles
1	IRCCS – Fatebenefratelli (Giovanni B Frisoni)	Italy	Research center	Project management and coordination, dissemination and training
2	Prodema Informatics AG (Christian Spenger and Alex Zijdenbos)	Switzerland	Small medium enterprise	Provision of services in the IT sector, joint research activities
3	University of the West of England (Richard McClatchey)	UK	Research center	User requirements analysis and implementation of a medical grid service layer (gluing service)
4	Maat G Knowledge SL (David Manset)	France/Spain	Small medium enterprise	Deployment of existing e-infrastructure and adaptation for neurodegenerative disease studies
5	VU Medisch Centrum (Frederick Barkland)	The Netherlands	Research center	Host of the MRI and clinical data. Definition of the user community requirements
6	Karolinska Institutet (Lar-Olof Wahland)	Sweden	Research center	Host of the AddNeuroMed MRI database. QA and QC of the AddNeuroMed MRI scans. Processing and analysis of clinical MRI
7	HealthGrid (Yannick Legre and Tony Solomonides)	France	Small medium enterprise	Dissemination and training privacy and ethical issues, distributed medical and Grid services

Figure 3. Institutions, main roles and key people of the neuGRID centers involved in the project.

IT: Information technology; QA: Quality assurance; QC: Quality control.

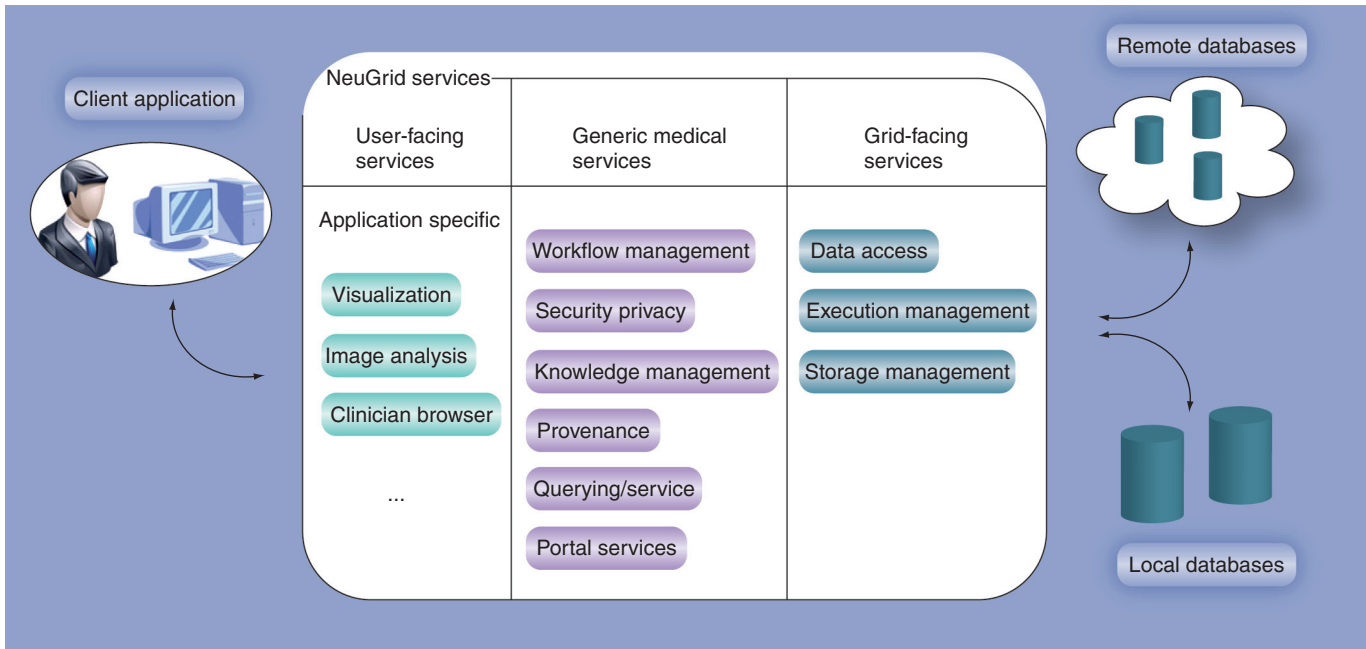


Figure 4. NeuGRID service-oriented architecture structure. User services, generic services and Grid-facing services have been defined. Via a well-characterized service-oriented architecture, other services can be added into the flexible and generalizable architecture.

The DCC's primary functions are the deployment, development and maintenance of the applications inside neuGRID. It coordinates operations such as: the standardization of acquisition protocols, the development of quality control procedures, the monitoring of data consistency, and the use, performance, evaluation and validation of the image-analysis algorithms. The DCC takes care of coordinating and maintaining the different downstream Data Archiving Computing Site (DACS) of level 1. Towards a later phase, the DCC will also participate in the Grid, providing additional storage space and CPU resources to neuGRID. The DCC is physically located in Prodema Informatics facilities in Wil, Switzerland.

The GCC is in charge of hosting, maintaining and running the Grid middleware services. The information system service is the core component of every Grid. Thanks to the information system, it is possible to obtain updated information regarding the status of the Grid; to retrieve updated information regarding free space on storage elements or the number of free CPUs of a DACS computing element; to discover the closest computing element for a worker node of a DACS core laboratory or to search for the best computing element matching a particular users' job requests. These services are the cornerstones providing the inner mechanics of the distributed infrastructure. The GCC is installed in Maat GKnowledge headquarters in Archamps, France.

Level 1: the DACS primary functions are the coordination and quality control of data acquired at the level 2 sites; and the provision of Grid resources (storage space, CPU and expertise). All the Grid data will be stored and archived at the level 0 and level 1 nodes. The DACS have been strategically selected so that pre-existing supercomputing facilities can be bridged and then used by the neuGRID system, allowing the execution of large-scale computing intensive data analysis. Powerful brain analysis tools will be available for the analysis of large datasets of thousands of MRI. The DACS is situated at IRCCS Fatebenefratelli in Brescia, Italy, at the Karolinska Institute in Sweden and at the VU Medical Center in The Netherlands. Each node of this project is planned to be equipped with a Grid cluster and supercomputer capabilities characterized by at least 100 processor cores, 100 GB of RAM memory and up to 6 TB of physical space. In addition, each neuGRID node could be easily expandable by different orders of magnitude. The DACS number could rise over the course of the project with the inclusion of new centers, assuring a gradual and constant development of the neuGRID platform, which will result in augmented computing facilities and enlarged user communities. Central storage is expected to rapidly increase reaching a theoretical total storage capacity of some petabytes by the end of 2010.

Beyond the level 0 and level 1 centers, there is one other potential level of the neuGRID platform: level 2. This would consist of the so-called Data Collection Sites (DCS), which should collect MR imaging sequences and nonimaging records. At this early stage of development the level 2 centers do not, as yet, belong to the neuGRID platform. However, once the procedure of data acquisition in the level 2 sites begins then the data should be sent to the DACS through a secure internet connections (i.e., Secure File Transfer Protocol/Secure Socket Layer). Later, these data will undergo quality assurance control by level 1 centers, they will be included in the neuGRID database and then they will be made public for the entire neuroimaging community. When feasible in terms of IT infrastructure, DCSs may also act as the lowest computing resource satellites in the Grid.

NeuGRID nodes exploit the GÉANT [105] EU Research Network. GÉANT provides neuGRID with the highest possible capacity present today (through a bandwidth of up to 10 Gbps) and offers one of the widest geographic coverage in the world. The Level 0 and level 1 of neuGRID sites have already been identified as relevant candidates for GÉANT connectivity.

NeuGRID architecture

In computing, a SOA represents a method in which a loose coupling of functions between operating systems, programming languages and applications [52] can be achieved. A SOA

separates functions into distinct units termed services [53]. The MammoGrid project proposed perhaps the first healthcare example of a Grid-based SOA [43]. The MammoGrid philosophy was to provide a set of services that were independent both of the front-end clinician-facing software and of the back-end Grid-facing software. In other words, a three-layer service architecture was proposed and implemented where the clinician was protected and isolated from the choice of Grid technologies. NeuGRID has followed the MammoGrid philosophy in developing a platform based on a SOA [54,55], in order to enhance its flexibility and interoperability and to promote its reusability, potentially across other medical applications. This architecture will allow neuGRID to be designed and implemented in such a way that its users do not require any advanced Grid know-how. This will be a great benefit since it has been shown that users often find it difficult to cope with the inherent complexities of Grid infrastructures.

Through the SOA, the neuGRID services are designed to satisfy all the specific requirements of neuroscientists. During the first 12 months of the project, detailed user requirements have been investigated, identified, collected and described in the three DACS neuroimaging centers, Karolinska Institute, Fatebenefratelli and VU Medical Center. At present, in neuGRID the Grid services are implemented on the EGEE/gLite middleware [106]. gLite provides a

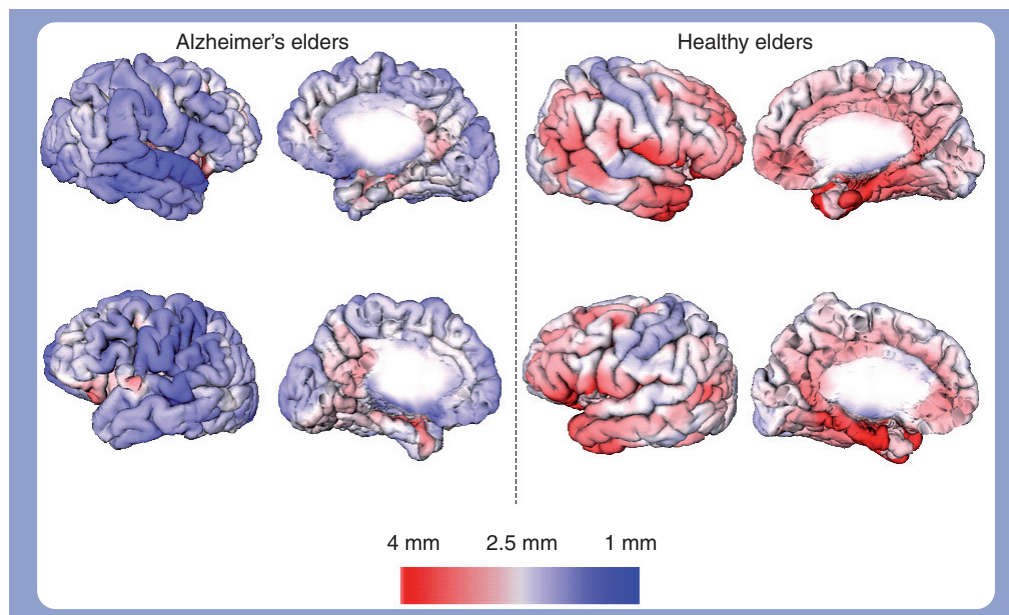


Figure 5. Gridified CIVET pipeline used to assay the cortical thickness marker in neuGRID.

Areas from red to blue denote increasing thinning of the cortex in a patient with Alzheimer's disease (left) and a normal older person (right).

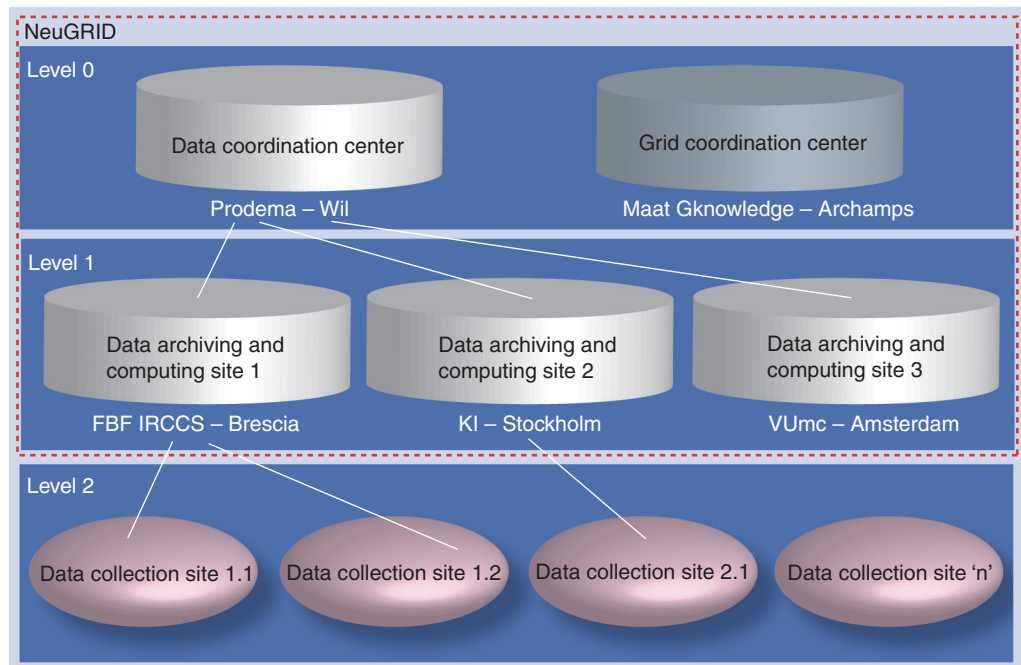


Figure 6. The neuGRID architecture is organized into distinct layers forming a well-organized hierarchical structure. The upper level 0 layer consists of the core nodes for Grid optimization processes. The intermediate level 1 layer is made of the data-analysis clusters. The lower level 2 layer consists of the centers responsible for image data collection.

framework for building Grid services by tapping into the power of distributed computing and storage resources across all the DACS.

However, most of the services developed within the neuGRID infrastructure will have the peculiarity of being interoperable, reusable, extensible and scalable [56,57], in order to ensure the highest flexibility.

The defined services are divided into three main groups (FIGURE 4):

- User-facing services
- Generic distributed medical services
- Grid-facing services

NeuGRID user-facing services

The user-facing services include those services that are accessed by the user for day-to-day activities. One example of a user-facing service, in this case specific to the neuroscience community, is the LORIS service. LORIS, developed for the NIH MRI Study of Normal Brain Development [27], is the clinical database on which neuGRID will be based. LORIS is a centralized system designed for the collection, management and processing of brain imaging data. It has been developed using a range of open source software such as Apache, PHP and MySQL and is characterized by an efficient database schema for storing brain images and

their associated metadata. In neuGRID, LORIS database will be extended to a Grid model and it will be made suitable for distributed data storage and high performance computational analysis. The resulting LORIS system will be able to collect, store, analyze and disseminate 3D DICOM image and related data across an arbitrarily large Grid network of participating sites using the same database schema. This objective will be achieved through a specific implementation and technical integration of the Grid and LORIS systems. Extensive validation studies will be conducted. In fact, LORIS must function in a manner transparent to the end user, producing results identical to the centralized system but without its potential limitations upon computational speed or storage capacity. More concretely, the LORIS service will be responsible for the data acquisition, capturing, annotating, visualizing, tracking the acquired datasets and sharing them with other users and services. It will also carry out quality control and validation on the acquired images (QC traceability service) and will apply the necessary format changes to make them useful for the neuroimaging community. Finally, the LORIS database will harmonize the main neuroimaging datasets (e.g., US-ADNI, AddNeuroMed and pilot E-ADNI) that will be collected within neuGRID. At this early stage, the harmonization of the ADNI and

AddNeuroMed protocols has already been completed. The harmonization was conducted by generating a specific neuGRID LORIS database schema. For this purpose a unified data dictionary was created. The data dictionary describes variables and data available for research that can be conducted through neuGRID. Specifically, it unifies the descriptions of the data acquired using the ADNI and AddNeuroMed protocols and results in an ontological mapping between the two projects.

NeuGRID generic distributed medical services

By abstracting Grid middleware specific considerations and customizations from clinical research applications, the neuGRID generic services provide functionality aimed at medical applications. These services will bring together sources of data and computing elements into a single view, making it possible to cope with centralized, distributed or hybrid

data and to provide native support for common medical file formats. Lower-level services will hide the peculiarities of any specific Grid technology from the upper-service layers, thereby providing application independence and enabling the selection of 'fit-for-purpose' infrastructures. The generic services will glue a wide range of user applications to the available Grid platforms, creating a foundation of cross-community and cross-platform services. The generic medical services are not tied to a particular application or Grid middleware, they could be used in any application domain and can be deployed potentially on any Grid infrastructure. These services are designed in such a way that a variety of applications and Grid middlewares could be supported. This level of abstraction will be reached through the use of a so-called gluing service that adopts the Simple API for Grid Application (SAGA) standardized protocol [107] for achieving middleware independence and platform interoperability.

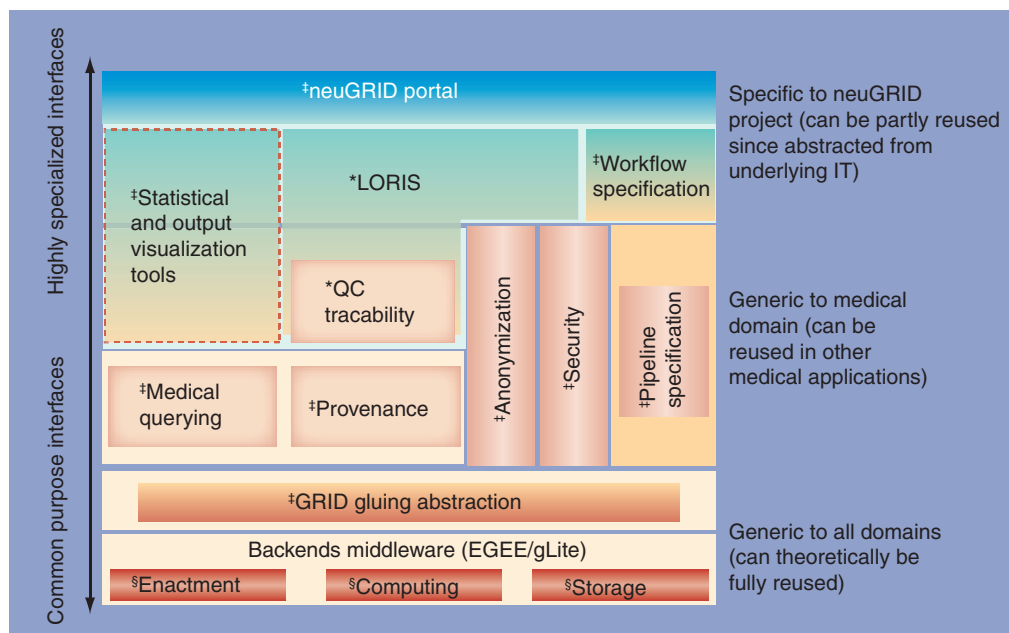


Figure 7. NeuGRID service-oriented layers architecture is made of three groups of services.

*User-facing service.

+Generic service.

§Service that is always on the Grid middleware.

The first group is the user-facing services. A user who belongs to the neuroimaging community could immediately exploit these well-known tools. The second group of services (generic services) can be used by any application and can be encapsulated in any Grid middleware. The third group of services always sits on the Grid middleware and will provide the necessary infrastructure to execute the workflows. These Grid-facing services are middleware specific and depend on the particular middleware being used to provide the distributed Grid infrastructure. NeuGRID is a highly generalized architecture, in fact gradually moving from the bottom to the upper part of the diagram, the Grid architecture shifts from a generic and flexible interface, reusable in many other projects, to a more highly specialized and specific user-facing interface regarding exclusively the field of neuroimaging analysis.

IT: Information technology; LORIS: Online Research Imaging System; QC: Quality control.

The first of the generic services is the portal service [108]. This service has been designed to provide a single point of access in the neuGRID system. It will hide the complexity of the underlying low-level architecture. It will allow users to easily identify themselves, access the services, browse data, launch analyses, submit jobs and visualize results. The user can access most other supplied services through the portal service. The neuGRID portal architecture has a web Single Sign On (SSO). This is a mechanism that allows a user to share its authentication between all the different services. One of the most mature and widely used SSO technologies is the Central Authentication Server (CAS) [58]. CAS is an authentication system created by Yale University (CT, USA) to provide a trusted way for an application to authenticate a user. From the user's perspective, only one account is used to access all the neuGRID applications and one single login is required in order to access all the websites. The portal service has two components, one component provides the front-end interfaces and the second component links the back-end services (the other generic services) to the front-end. Before trying to access the back-end services through the portal, the user has to authenticate himself or herself through the neuGRID security interface.

The neuGRID platform is intended to handle a large quantity of sensitive data coming from heterogeneous sources. It is extremely important to ensure that medical information is not made available without appropriate ethical clearance. Such legal and ethical requirements mean that an efficient and common level of anonymization (anonymization service) must be used throughout the project. Anonymization should therefore be considered at two levels:

- Pseudonymization: this is a procedure whereby the most personal-related data within a data record are replaced by artificial identifiers (e.g., hash values) that map one-to-one to a patient. The artificial pseudonym always allows tracking back of data to its origins (but only by the original treating physician)
- Face scrambling: this is the process whereby a specific algorithm is applied so that the face is removed from an MRI, thus preventing the possibility of a subject being recognized. Face scrambling will be executed if subjects are acquired with high 3D MRI resolution.

To become a successful research infrastructure, it is clear that such issues will need to be addressed carefully. A Java applet and

scrambling software will be implemented in neuGRID in order to perform the aforementioned tasks in accordance with the European Parliament Directive [109] and the Member States' regulations.

The workflow and pipeline specification service, commonly known as the pipeline service, constitutes one of the key elements in providing generic services for research analysis. Neuroimaging pipelines allow neuroscientists and clinicians to apply series of automated transformations and processes on brain images for decision support purposes using complex and nested workflows. These processes are very computer intensive and deal with large amounts of data. Grid-enabled neuroimaging pipeline services are either proprietary or under research and neuroscientists have to currently rely on command line scripts to design and execute their pipelines. This service will enable scientists to create and design workflows in a user-friendly fashion, calling and combining automatically different algorithms (e.g., FMRIB Software Library, BrainSuite, FreeSurfer, MNI tools, Automated Image Registration, Analysis of Functional NeuroImages and many other programs) in a single environment. Once the inputs retrieved by LORIS and parameters specified by users have been declared, a main source of analysis can be drawn via this service. This service will enable researchers to plan and Grid-enable their pipelines for enactment over a Grid. The pipeline service is being designed to allow multiple workflow authoring tools to define standardized workflow specifications including the LONI pipeline workflow management system and other scientific workflow applications [59] (such as Kepler [60]). In order to achieve a high level of abstraction and interoperability, a language-independent application programming interface (API) will be defined, representing the way by which the different workflow authoring tools will request the service. The authored workflows generated will be transformed into a common format that could be enacted in multiple execution environments. Furthermore, users will be able to download specific workflows to their workspace and edit them by modifying algorithms, changing input data sources and other settings. From within the portal service, the user has the opportunity to submit the workflow for enactment and visualize the results. NeuGRID final users will browse for available workflows through the querying service. Finally, to execute the user-authored

pipeline in a Grid environment and to fully exploit the Grid capabilities, the pipeline will be optimally planned and transformed into a Grid-executable state. This transformation is often referred to as 'planning' or 'Grid-aware distribution'. In a Grid, multiple sites may deploy subsets of tasks while some other tasks may require specialized hardware requirements and may need to be staged to other sites. Some sites may be preferred for selection owing to input or output data location considerations, while others may be preferred due to availability of suitable computing resources. Hence, a number of decisions need to be considered before a workflow is enacted. All these tasks are addressed by the pipeline service.

The querying service is a general purpose browsing service that can query and browse data that are stored in a file or relational database. The querying service will provide an easy to use high level querying capability for the LORIS data store, provenance data and other Grid databases. The neuGRID querying service has been designed to provide a mechanism that facilitates users in obtaining meaningful results quickly and efficiently. The service has been implemented to query disparate data resources located in the three DACS distributed LORIS databases. The querying service will also be a point of contact to query the provenance as well as other databases in Grid. In order to access and query data through the querying service, the user in neuGRID has to invoke the neuGRID access system, which may consist of either simple user/password declaration or via a public key infrastructure system. Users can also view results from previous job executions, can access the output data from the LORIS and provenance databases and can download this data into their workspaces.

The provenance service is a generic service that will capture, store and provide access to analysis data for improved decision making. The provenance service will also enable users to reconstruct their workflows and validate the final results as well as intermediate results. Provenance is the process of tracking the origins of data, their history, their reconstruction and validation, as and when required by the users. The provenance service will help the users by facilitating access to past executions or histories of their analyses before a new study set is initiated. The process of neuroimaging analysis will involve a number of steps; an execution failure at any stage may lead to undesired execution results. These may be caused by

incorrect pipeline specifications, inappropriate links between pipeline components, execution failures owing to the dynamic nature of the Grid and other causes. The provenance service provides the means for capturing and maintaining workflow specifications and execution information in a provenance database. Users will be able to query in a transparent way, any provenance information including the specific software routines and operating systems that were used. Consequently, the provenance service allows to the neuroimaging community proper data interpretation, high fidelity replication and reuse.

Before any workflow is passed over to neuGRID gLite environment, it is handed over the gluing service, which has an important role to play in the architecture before execution. The gluing service is a standard way of accessing Grid services without tying generic services and applications to a particular Grid middleware. This is a solution that extends and enhances the reusability of already deployed services across domains and applications. The gluing service hides the encapsulation of Grid-middleware complexities from the neuGRID services. Using the gluing service, the workflows, queries and other jobs can be submitted to any Grid environment that could be built upon gLite, Globus [110], Unicore [111] or indeed any other kind of distributed middleware. In effect, the gluing service is the gateway that shields the services from being locked in to a particular Grid and makes them truly generic and reusable. The neuGRID gluing service aims to provide a mechanism to access any deployed Grid middleware in an easy-to-use manner and a simplified approach for enabling clients to 'gridify' their applications without installing and maintaining Grid-specific libraries.

NeuGRID services

Usually, once the gluing service passes over the workflow to a Grid middleware, it is managed and scheduled on the distributed resources for enactment. The common Grid services are an enactment service, computing service, storage service and some other Grid services that make it possible to exploit the computing power of Grid. Therefore, when a workflow has been specified, associated data elements have been identified and it has been transformed into an executable format and the workflow is ready for enactment. This enactment process has to be managed via the related enactment service, which deals with the execution of the

user-defined workflows on the Grid resources. The workflow enactment service takes a workflow specification as an input, queries the data that are required to run the workflow, mimics the Grid workflows according to some performance or efficiency considerations and dispatches the workflows for enactment on the Grid environment. Finally, when a job or workflow is being enacted in a Grid-execution environment, its progress and corresponding input and output operations are being monitored and logged into the provenance database. Ultimately, the results of the workflow execution are stored both in the LORIS and in the provenance data stores and from there, they are handled through the output visualization of the user-facing service.

The neuGRID service architecture here depicted encompasses different levels connected to each other. The neuGRID architecture emphasizes the possible separation of the client application domain from the underlying Grid infrastructure. In this way, the neuGRID platform can be considered free-distribution aware and Grid agnostic (FIGURE 7).

NeuGRID performance tests overview

One of the highest expectations of this project is the gain in performance, scalability and flexibility from executing the powerful cortical thickness algorithm within the neuGRID computing infrastructure. Indeed, the Grid will not only provide a secure, distributed and efficient environment for running such complex computing and data intensive analysis, but it will also enable its execution on a larger dataset shared across the several DACS centers in Europe. As soon as the neuGRID deployment phase is finished, performance and feasibility tests will be extensively carried out with the aim to control the integration of the different system components, supporting the deployment of the platform over the physical infrastructure and conducting extensive performance and validation tests throughout the different levels and sites infrastructure.

The neuGRID performance will be evaluated in simulations mimicking existing real-world hardware and networking infrastructures so that it will address potential pitfalls and bottlenecks such as bandwidth limitations between nodes, heterogeneous compute resources, different scenarios for data archiving, access and replication. Simulations of various job executions, scheduling scenarios, analysis of the system capacity to optimally handle the job frequency and various

queuing and caching approaches will also be studied and evaluated. Moreover, to test the relevance of the selected models, neuGRID has planned data challenges as well as functional tests. Data challenges will make use of all the available ADNI volumetric images (more than 7500 T1–3D MP-RAGE scans) and the accessible neuGRID processing power, in order to conduct the largest and most complex possible analysis. These tests should drive the ongoing developments and will validate neuGRID's computing model and its gridified software (CIVET). It should also highlight neuGRID infrastructure's limitations, and give indications for the infrastructure's exploitation and sustainability. Subsequently, functional tests will illustrate how 'usable' the system is. In other words, data challenges will provide accurate measurements of the platform, infrastructure performance and possible limitations; while the functional tests series will validate the neuGRID services from the user standpoint.

It is clear by now that the execution and performance of complex and intensive pipelines like the CIVET will be significantly improved according to the added value of the Grid technology: 'huge computing power resources serving severely time constrained simulations'.

NeuGRID limitations

Considering the interdisciplinary nature of the neuGRID project, in which the IT and neuroscientific fields are closely in touch, several risk factors could hinder the development of the Grid infrastructure. Therefore, the most probable pitfalls consist of:

- Poor network infrastructure. Initially, the Italian DACS connectivity was not suitable for exchanging large amounts of data (e.g., replication of MRI scans with an individual scan size of 400 MB) among the different nodes of the Grid. The problems faced were requesting an upgrade of the connection to the Italian Academic and Research Network (GARR) [112] and obtaining a bandwidth comparable with those used by the other neuGRID DACS centers.
- Difficulties with making the algorithms Grid compliant. This is only a potential risk with low probability. However, once the technology evaluations have been performed, the potential problems would be identified and corrected.
- CPU and storage resource usage exceeding neuGRID capacity. NeuGRID will be fully functional but it will have limited CPU and

storage resources. Although this limit will not be saturated shortly it will be necessary to provide additional resources.

- Results variation in the image processing algorithms as a function of different hardware, and possible discrepancy when these algorithms are running on or off the Grid. This is a very likely and well-known problem. The variability due to differences in numerical precision has to be expected since different machines are present in neuGRID. In this case, the source of the problem should be identified, its impact assessed and, if possible, the problem addressed.
- New user's difficulties: the neuGRID environment will require new skills for researchers. Scientists will need to learn how to work in new environments, conceiving and leveraging powerful new instruments. Even though neuGRID will develop user-friendly interfaces, scientists will, at least for the foreseeable future, operate on the cutting edge of what is possible. To address this problem, training activities will aim to provide neuroscientists and algorithm developers the necessary skills to access and run the infrastructure.
- Grid infrastructure maintenance; strategic and financial sustainability will be addressed starting from the beginning of the third year of funding when the needs of the big pharmaceutical companies regarding imaging markers for neurodegenerative disorders or for precompetitive research and multicenter clinical trials will be surveyed. It is expected that industrial interests and support will contribute to long-term sustainability of this infrastructure. In the meantime, neuGRID will also look into other running projects and initiatives to identify potential strategic alliances for enlarging its network and developing strengths. Some efforts should be made with Montreal Neurological Institute, Laboratory of NeuroImaging (LONI), Worldwide ADNI initiatives (US-ADNI, Australian ADNI and Japanese ADNI), European Alzheimer Disease Consortium (EADC), Dementia Research Group and many others.

However, neuGRID applies a risk management method in order to identify and monitor risks that could have a large impact on the project and mitigation plans have been set up to reduce the risk probability and their impacts.

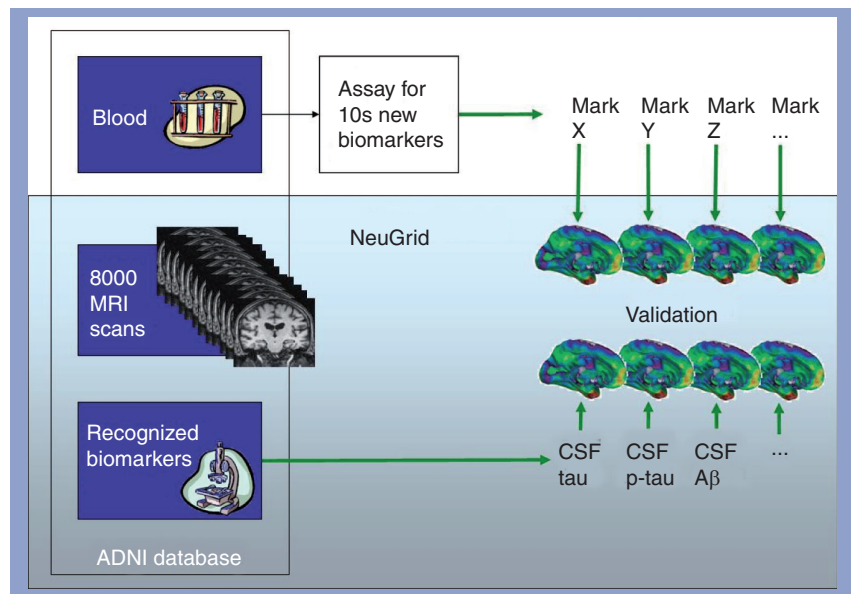


Figure 8. Validation of new biomarkers for Alzheimer's disease using neuGRID. Use case 1: through the neuGRID platform, the validation of new biomarkers for Alzheimer's disease will accelerate considerably mainly owing to wider datasets and higher computational resources never available before. For details on this case, please refer to text 'Grid applications in neuroscientific research'.

Amyloid- β : A β ; ADNI: Alzheimer's Disease Neuroimaging Initiative.

Grid applications in neuroscientific research

Grid infrastructures should enable a new research environment for the scientific community and future generations of researchers. It is critical for a service aimed to address global research needs, that the data included will be consistent with international standards [104], allowing pooled analyses or comparative studies. Furthermore, thanks to the federation of the ADNI, AddNeuroMed and all the other datasets, it will be possible to constitute an unprecedented environment for the validation of imaging and biological markers for AD and other neurodegenerative disorders.

A Grid application in clinical neuroscience is represented in FIGURE 8. A typical case is when a neuroscientist believes that a series of newly found biological markers assayed in blood could be more intimately related to the primary cause of Alzheimer's disease than any other known marker, such as tau, amyloid- β (A β), or others. As a consequence, the new markers might be more effective for an early diagnosis of Alzheimer's disease and could be more accurate to track the efficacy of new drugs too. From the AddNeuroMed and US-ADNI consortia, the neuroscientist receives blood samples of many patients with early AD and some samples of healthy older persons, together with the usual

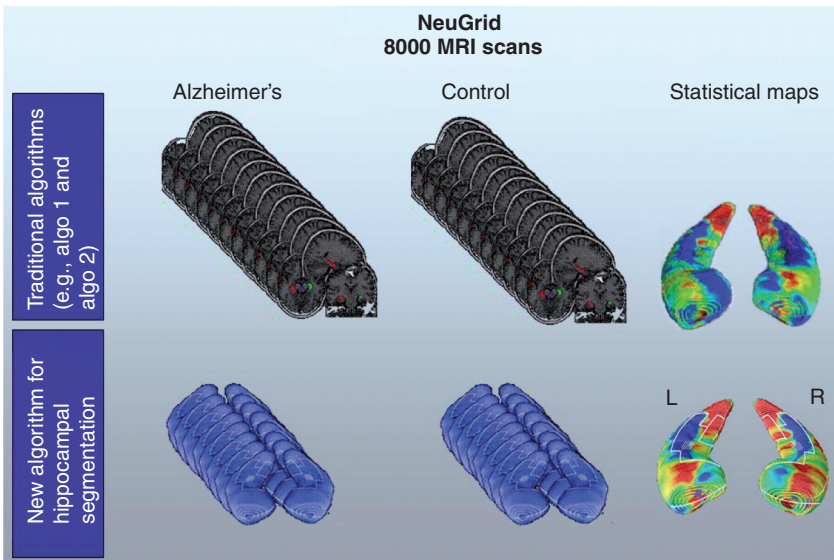


Figure 9. Validation of new computational neuroanatomy algorithms using neuGRID. Use case 2: the neuGRID platform will provide information technology developers and neuroscientists a powerful platform to test new image analysis algorithms.

recognized markers of neurodegeneration (neuropsychological tests, high resolution structural MR, fluorodeoxyglucose-PET, and cerebrospinal fluid tau and $A\beta$). Through discriminant analysis, the neuroscientist finds that some of the new markers might be more accurate to separate diseased from nondiseased patients. At this point, the neuroscientist might contact one neuroimaging Grid platform to validate the power of the newly discovered markers throughout imaging indicators of disease progression and *ad hoc* time-consuming algorithms. Complex correlative analysis would then be carried out on the structural MRI scans of cortical thickness with the new biological markers found. A potential critical issue could be represented by the data consistency within neuGRID platform. The approximately 2500 high definition structural MRI scans of the AddNeuroMed project, as previously argued, will be consistent with the approximately 5000 images of the North American ADNI, allowing full interoperability of the two datasets and the consequent joint analysis. In this manner, all the data acquired and stored within neuGRID will be consistent with international standards, allowing pooled analyses or comparative studies. This will constitute an unprecedented environment for the validation of imaging and biological markers for Alzheimer's and other neurodegenerative disorders. In doing so, using the Grid capabilities in a few hours the neuroscientist could demonstrate that a new specific marker is more strictly related to cortical thickness than any other classical recognized marker.

More generally, neuGRID will meet the needs of the academic neuroscientists' community, deploying repositories with consistent clinical/imaging data and advanced computing facilities to be used in research and care applications. In addition, neuGRID is particularly suitable to meet the needs of the scientific community for the expansion of research in the whole area of progressive brain diseases, focusing especially on those of old age. In fact, the availability of gridified algorithms widely accessible to the neuroscientific community will spur research lines aimed at elucidating the mechanism of action of a number of environmental and genetic factors putatively involved in accelerated brain aging, AD and other neurodegenerative diseases. This new research environment, capitalizing on new computing and communication technologies, will promote greater breadth and depth to new international collaborations amongst researchers. NeuGRID will also meet the needs of the pharmaceutical and biotechnology companies regarding the study and the characterization of new imaging markers for neurodegenerative disorders also improving the way in which they can gauge the efficacy of drug tests.

Another possible application of the Grid infrastructures is illustrated in FIGURE 9. Biomedical algorithm developers could have a powerful test-bed as well as the opportunity to access a large community of neuroscientists that might exploit their products. Developers of new algorithms for brain analysis, able to extract information on brain features to be used as disease markers, could have a larger dataset to test the performance of their algorithm than the traditional ones currently available. Moreover, successful algorithms will be able to be used and exploited by clinical neuroscientists immediately, once algorithms have been tested and optimized. The assessment of reproducibility and the validation of accuracy of the results obtained with a new algorithm are of importance to developers, as well as researchers or clinical end users. Using the Grid platforms, the programmer is able to validate the algorithm against established standards, as well as test the sensitivity and inner behavior of the algorithm against normally occurring variations in input data using an imaging database of unprecedented size. Using the computational power of a Grid, the developer would rapidly, and in a highly structured fashion, carry out a series of experiments, which up until the appearance of Grid infrastructures, would not have been feasible. By optimizing the parameter values of the algorithm across the database accessible in the Grid environment, the developer could prove

that the algorithm outperforms any pre-existing method and establish its utility as a tool in the study of neurodegenerative disease.

By exploiting Grid technology and focusing on the processing of information, neuGRID will also contribute to the building of a bridge between Grid IT technology and medical applications. This will effectively facilitate the transfer of the Grid power from computer science laboratories into the research world of medicine. The benefits that will follow include large increases in the peak capacity and the total computing power, as well as new ways for research and clinical communities to share and analyze very large data sets. NeuGRID should be considered as a concrete example of a novel shared use of computing and data resources across diverse technological applications and national domains. Even if the matter is still far from being completely solved, the current IT development is maturing quickly enough to support the emergence of this deployed e-infrastructure and, in the long term, its dissemination in the research community.

Conclusion & future perspective

The penetration of IT in medical sciences will undoubtedly continue to address clinical demands and to provide increasingly functionality [61–66]. It is expected that the exploitation of Grids could thoroughly change the way in which neuroscience is carried out today. Scientists will no longer need to migrate to advanced image analysis centers as computational facilities will be readily accessible via a standard internet connection and large databases will not pose issues of space saturation, owing to the Grid connectivity. Grid-based infrastructures, such as neuGRID, will bring the image analysis center to the neuroscientist, as opposed to the other way round. Neuroscientists will find a virtual space where the world's largest brain MRI databases of individuals with memory disturbances and Alzheimer's disease will be stored together with pertinent clinical variables integrated with biological data. Neuroscientists will be able to take advantage of all the primary neuroimaging tools inside the platform and by exploiting a user-friendly graphical interface they will create and launch analytical workflows to test scientific hypotheses.

NeuGRID could become one of the most usable and accessible Grid infrastructure for computational neuroscience, providing a well managed, easy-to-use set of tools with which scientists can perform analyses with a considerable saving of time. If neuGRID lives up to its promise, these tools will provide seamless access

to Grid resources, offering end users unprecedented power at their fingertips with the simplicity of a web service interface. Furthermore, Grid platforms can reduce the latency of obtaining results, increase researchers' productivity and speed up the discovery of new knowledge. Once neuGRID has been concluded, a Grid-based platform providing a consistent portfolio of services to different end-user communities such as clinical neuroscientists, brain image algorithm developers and, prospectively, pharmaceutical industries and non-neuroscience biomedical communities can be exploited.

Future perspectives relate to worldwide interoperability of the available e-infrastructures for neuroscientific analyses and initially involve the LONI's SUNs Grid Engine and MNI's CBRAIN. LONI's cranium Grid engine is already operational in the USA, while CBRAIN is presently under development. The interoperability of neuGRID with LONI will be facilitated by the clinical variable database structure of neuGRID being that of ADNI. There remain some challenges in integrating LONI with neuGRID but these are currently being addressed. The interoperability of neuGRID with CBRAIN is facilitated by three critical technological circumstances: the adoption of a SOA in neuGRID, which promotes interoperability, the use of common database software (LORIS) in both neuGRID and CBRAIN and the physical network connecting Europe to Canada being an ultra broadband link (CANet4 [113]) with the capacity of running computationally complex and data-intensive studies. Together, these technological choices should simplify and facilitate the interoperation of neuGRID with CBRAIN. Future efforts should be directed towards laying the foundation for a large research and development initiative aimed at achieving full interoperability among these infrastructures. Only by calling into action all the major players of this field will it be possible to achieve a critical mass to effectively advance the ultimate diagnosis and treatment of neurodegenerative diseases.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Overview

- The availability of image databases of unprecedented size and the spread of sophisticated central processing unit-intensive algorithm pipelines for image analysis encourage the development of scientific e-infrastructures.
- These will foster the development of innovative instruments for the early diagnosis of highly prevalent and deadly diseases such as Alzheimer's and lead to the identification of markers that will facilitate the development of causative drugs.
- NeuGRID is an EC-funded effort arising from the needs of the Alzheimer's disease imaging community aiming to become one of the most usable and accessible Grid infrastructure for computational neuroscience. NeuGRID will allow the collection and archiving of large amounts of imaging data paired with Grid-based algorithms and adequately powered computational resources.

Conclusion

- Grid infrastructures, such as neuGRID, will make sophisticated analysis techniques and imaging data available to a larger number of neuroscientists, drastically reduce the processing time of neuroimaging computation and allowing the querying of larger and more representative databases.

Future perspective

- Interoperability between European and international infrastructures should be sought in the future through large coordinated efforts.
- Both basic research and the pharmaceutical industry will benefit from neuGRID in order to study markers for Alzheimer's disease and other neurodegenerative diseases.

Bibliography

Papers of special note have been highlighted as:

- of interest
- of considerable interest

1. van Straaten ECW, Scheltens P, Barkhof F: MRI and CT in the diagnosis of vascular dementia. *J. Neurol. Sci.* 226, 9–12 (2004).
2. Practice parameter for diagnosis and evaluation of dementia. Report of the quality standards subcommittee of the American Academy of Neurology. *Neurology* 44, 2203–2206 (1994).
3. Frisoni GB, Scheltens P, Galuzzi S *et al.*: Neuroimaging tools to rate regional atrophy, subcortical cerebrovascular, and regional cerebral blood flow and metabolism: consensus paper of the EADC. *J. Neurol. Neurosurg. Psychiatr.* 74, 1371–1381 (2003).
- **Updated review of tools with demonstrated accuracy for detecting regional atrophy, cerebrovascular and brain function.**
4. Blow N: Neuroscience tools: brain insights. *Nat. Methods* 5(11), 981–987 (2008).
5. Scheltens P, Fox N, Barkhof F *et al.*: Structural magnetic resonance imaging in the practical assessment of dementia: beyond exclusion. *Lancet Neurol.* 1(1), 13–21 (2002).
6. Knopman DS, DeKosky ST, Cummings JL *et al.*: Practice parameter: Diagnosis of dementia (an evidence-based review): report of the quality standards subcommittee of the American Academy of Neurology. *Neurology* 56, 1143–1155 (2001).
- **First to publish on the importance of the magnetic resonance to diagnose dementia.**
7. DeCarli C: The role of neuroimaging in dementia. *Clin. Geriatr. Med.* 17(2), 255–279 (2001).
8. Teipel SJ, Meindl T, Grinberg L *et al.*: Novel MRI techniques in the assessment of dementia. *Eur. J. Nuc. Med. Mol. Imaging* 35, S58–S69 (2008).
9. Ashburner J, Csernansky JG, Davatzikos C *et al.*: Computer-assisted imaging to assess brain structure in healthy and diseased brains. *Lancet Neurol.* 2, 79–88 (2003).
10. Otte A, Halsband U: Brain imaging tools in neurosciences. *J. Physiol.* 99, 281–292 (2006).
11. Frisoni GB: Structural imaging in the clinical diagnosis of Alzheimer's Disease: problems and tools. *J. Neurol. Neurosurg. Psychiatr.* 70, 711–718 (2001).
12. Shattuck DW, Prasad G, Mirza M *et al.*: Online resource for validation of brain segmentation methods. *Neuroimage* 45(2), 431–439 (2009).
13. Clifford RJ, Bernstein M, Fox NC *et al.*: The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* 27(4), 685–691 (2008).
14. Van Horn JD, Toga AW: Is it time to re-prioritize neuroimaging databases and digital repositories? *Neuroimage* 47(4), 1720–1734 (2009).
- **First to suggest the importance for the neuroimaging community to reprioritize large-scale database for the neuroimaging community.**
15. Hasson U, Skipper JI, Wilde MJ *et al.*: Improving the analysis, storage and sharing of neuroimaging data using relational databases and distributed computing. *Neuroimage* 39, 693–706 (2008).
16. Foster I: The Grid: a new infrastructure for 21st Century science. *Phys. Today* 55(2), 42–47 (2002).
17. Banatre JP, Campadello S, Danelutto S *et al.*: Future for European Grids: grids and service-oriented knowledge utilities vision and research directions 2010 and beyond. ©NGG Group 2006 Next Generation Grids Expert Group report 3. Office for Official Publications of the European Communities, Luxembourg 1–60 (2006).
18. Glatard T, Lingrand D, Montagnat J *et al.*: Impact of the execution context on Grid job performances. *Proceedings of the Seventh IEEE International Symposium on Cluster Computing and the Grid, IEEE Computer Society*, 713–718 (2007).
19. Germain C, Breton V, Clarysse P *et al.*: Grid-enabling medical image analysis. *J. Clin. Monit. Comput.* 19(4–5), 339–349 (2005).
20. Foster I: The Grid: computing without bounds. *Sci. Am.* 288(4), 78–85 (2003).
21. Mueller SG, Weiner MW, Thal LJ *et al.*: Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement.* 1(1), 55–66 (2005).
22. Roberto V, Zappia A, Testa C *et al.*: Neuroimaging services on the net. *Comput. Based Med. Syst. 18th IEEE Symp.* 61–63 (2005).
23. Frisoni GB, Henneman WJP, Weiner MW *et al.*: The pilot European Alzheimer's Disease Neuroimaging Initiative of the European Alzheimer's Disease Consortium. *Alzheimers Dement.* 4, 255–264 (2008).
24. Lovestone S, Francis P, Strandgaard K: Biomarkers for disease modification trials – the innovative medicines initiative and AddNeuroMed. *J. Nutr. Health Aging* 11(4), 359–361 (2007).

25. Nielsen FA: The Brede database: a small database for functional neuroimaging. *Neuroimage* 19(2), (2003).
26. Waber DP, De Moor C, Forbes PW *et al.*: The NIH MRI study of normal brain development: performance of a population based sample of healthy children aged 6 to 18 years on a neuropsychological battery. *J. Int. Neuropsychol. Soc.* 13, 729–746 (2007).
27. Evans AC; Brain Development Cooperative Group: The NIH MRI study of normal brain development. *Neuroimage* 30(1), 184–202 (2006).
28. Blanquer I, Hernandez V, Segrelles D: TRENCADIS – A WSRF Grid MiddleWare for managing DICOM structured reporting objects. *Stud. Health Technol. Inform.* 120, 381–391 (2006).
29. Jakobovits RM, Rosse C, Brinkley JF: WIRM: an open source toolkit for building biomedical web applications. *J. Am. Med. Inform. Assoc.* 9(6), 557–570 (2002).
30. Kwan RKS, Evans AC, Pike GB: MRI simulation-based evaluation of image-processing and classification methods. *IEEE Trans. Med. Imaging* 18(11), 1085–1097 (1999).
31. Geddes J, Lloyd S, Simpson A *et al.*: NeuroGrid: using Grid technology to advance neuroscience. *Comput. Based Med. Syst. 18th IEEE Symp.* (23–24), 570–572 (2005).
32. Ainsworth J, Harper R: The PsyGrid experience: using web services in the study of schizophrenia. *Int. J. Healthc. Inf. Syst. Inform.* 2(2), 1–20 (2007).
33. Kola J, Harris J, Lawrie S *et al.*: Towards an ontology for psychosis. *Cogn. Syst. Res.* DOI: 10.1016/j.cogsys.2008.08.005 (2008).
34. Whyte A: Neuroimaging data landscapes: annex to SCARP case study 1. *Digital Curation Centre*, University of Edinburgh, UK (2008).
35. Montagnat J, Gaignard A, Lingrand D *et al.*: NeuroLOG: a community-driven middleware design. *Stud. Health Technol. Inform.* 138, 49–58 (2008).
- **Detailed definition and description of neuroLOG Grid middleware according to the neuroscience community.**
36. Cunha JPS, Oliveira I, Fernandes JM *et al.*: BING: the Portuguese Brain Imaging Network Grid. *IberGRID*, 268–276 (2007).
37. Murphy SN, Mendis ME, Grethe JS *et al.*: A web portal that enables collaborative use of advanced medical image processing and informatics tools through the Biomedical Informatics Research Network (BIRN). *AMIA 2006 Symp.* 579–583 (2006).
38. Keator DB, Grethe JS, Marcus D *et al.*: A national human neuroimaging collaboratory enabled by the Biomedical Informatics Research Network (BIRN). *IEEE Trans. Inf. Technol. Biomed.* 12(2), 162–172 (2008).
39. Rex DE, Ma JQ, Toga AW: The LONI pipeline processing environment. *Neuroimage* 19(3), 1033–1048 (2003).
40. Montagnat J, Bellef F, Benoit-Cattin H *et al.*: Medical images simulation, storage, and processing on the European DataGrid testbed. *J. Grid Comput.* 2(4), 387–400 (2005).
41. Bagarinao E, Matsuo K, Tanaka Y *et al.*: Enabling on-demand real-time functional MRI analysis using Grid technology. *Methods Inf. Med.* 44, 665–673 (2005).
42. Bagarinao E, Nakai T, Tanaka Y: Medical Grid: using Grid technology for brain studies. *Philipp. Inf. Technol. J.* 1(1), 3–7 (2008).
- **Exhaustive description of how and for which purposes medical imaging applications can be Grid enabled.**
43. Amendolia SR, Estrella F, Hassan W *et al.*: MammoGrid: a service oriented architecture based medical Grid application. *Grid Coop. Comput.* 3251, 939–942 (2004).
44. Amendolia SR, Estrella F, Del Frate C *et al.*: Deployment of a Grid-based medical imaging application. *Stud. Health Technol. Inform.* 112, 59–69 (2005).
45. Estrella F, Del Frate C, Hauer T *et al.*: Resolving clinicians' queries across a Grid's infrastructure. *Methods Inf. Med.* 44(2), 149–153 (2005).
46. Estrella F, Hauer T, McClatchey R, Odeh M, Rogulin D, Solomonides T: Experiences of engineering Grid-based medical software. *Int. J. Med. Inform.* 76(8), 621–632 (2007).
47. Warren R, Solomonides AE, Del Frate C *et al.*: MammoGrid: a prototype distributed mammographic database for Europe. *Clin. Radiol.* 62(11), 1044–1051 (2007).
48. McClatchey R, Buncic P, Manset D *et al.*: The MammoGrid Project Grids Architecture. *Computing in High Energy and Nuclear Physics CHEP03*, (2003).
49. Lerch JP, Pruessner JC, Zijdenbos A *et al.*: Focal decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy. *Cereb. Cortex* 15, 995–1001 (2005).
50. Lee JK, Lee JM, Kim JS *et al.*: A novel quantitative cross-validation of different cortical surface reconstruction algorithms using MRI phantom. *Neuroimage* 31(2), 572–584 (2006).
51. MacDonald D, Kabani N, Avis D *et al.*: Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *Neuroimage* 12(3), 340–356 (2000).
52. Erl T: *Service-Oriented Architecture: Concepts, Technology, and Design*. Prentice Hall PTR, Upper Saddle River, NJ, USA (2005).
53. Newcomer E, Lomow G: *Understanding SOA with Web Services Edition 1.0*. Addison-Wesley Professional, Canada (2004).
54. Foster I: Service-oriented science. *Science* 308, 814–817 (2005).
55. Bell M: *Introduction to Service-Oriented (SOA): Service Analysis, Design and Architecture*. Wiley & Sons, Hoboken, NJ, USA (2008).
56. Foster I, Kesselman C, Tuecke S: The anatomy of the Grid enabling scalable virtual organizations. *Int. J. Supercomput. Appl.* 15(3) (2001).
57. Olive M, Rahmouni H, Solomonides T *et al.*: SHARE road map for HealthGrids: methodology. *Int. J. Med. Inform.* 78(Suppl. 1), S3–S12 (2009).
58. Naito H, Kajita S, Hirano Y *et al.*: Multi-tiered security hierarchy for web applications using central authentication and authorization service. International Symposium on: *Applications and the Internet Workshops, SAINT Workshops* (2007).
59. Fox G, Gannon D: Workflow in grid systems. *Concurr. Computat. Pract. Exp.* 18(10), 1009–1019 (2006).
60. Ngu AHH, Bowers S, Haasch N *et al.*: Flexible scientific workflow modelling using frames, templates, and dynamic embedding. *LNCS J.* 566–572 (2008).
61. Andrade R, Oliveira I, Fernandes JM: Multi-voxel non-linear fMRI analysis: a Grid computing approach. *IberGRID*, 1–10 (2007).
62. Jacq N: *From Genes to Personalized Health Care: Grid Solutions for the Life Sciences*. IOS Press, The Netherlands (2007).
63. Temal L, Dojat M, Kassel G *et al.*: Towards an ontology for sharing medical images and regions of interest in neuroimaging. *J. Biomed. Inform.* 41, 766–778 (2008).
64. Nakagawa S, Kosaka T, Date S *et al.*: A Grid computing infrastructure for MEG data analysis. *Neurol. Clin. Neurophysiol.* 36, 1–4 (2004).
65. Gurcan MN, Pan T, Sharma A *et al.*: GridIMAGE: a novel use of grid computing to support interactive human and computer-assisted detection decision support. *J. Digit. Imaging* 20(2), 160–171 (2007).
66. Lippman H, Kruggel F: Quasi-real time neurosurgery support by MRI processing via Grid computing. *Neurosurg. Clin. N. Am.* 16, 65–75 (2005).

Websites

101. NeuGRID Project
www.neuGrid.eu
102. Internet Brain Segmentation Repository
www.cma.mgh.harvard.edu/ibsr
103. Canadian Brain Imaging Research Network Project
http://cbrain.mcgill.ca
104. Digital Imaging and Communications in Medicine
http://dicom.nema.org
105. GEANT Project
www.geant.net
106. EgEE/gLITE Project
http://glite.web.cern.ch/glite/
107. SAGA Project
http://saga.cct.lsu.edu
108. NeuGRID portal service
http://neuGRID.healthgrid.org/
109. Directive 95/46/EC of the European Parliament
www.spamlaws.com/f/docs/95-46-ec.pdf
110. GLOBUS Project
www.globus.org
111. UNICORE Project
www.unicore.eu
112. GARR the Italian academic and research network
www.garr.it
113. CANet4 Project
www.canarie.ca/canet4
114. Alzheimer's Disease Neuroimaging Initiative
www.adni-info.org
115. Innomed-AddNeuroMed
https://addneuromed.net/main.php
116. Pilot E-ADNI
www.centroalzheimer.it/E-ADNI_project.htm
117. European NeuroImaging Repository
www.enir.eu/tasks/tasks.php
118. The NIH MRI study of normal brain development
www.bic.mni.mcgill.ca/nihpd/info/data_access.html
119. An Information System for Managing Distributed Knowledge and Data Bases in Neuroimaging
www.irisa.fr/visages/demo/Neurobase/index.html

120. BrainWeb: simulated brain database
www.bic.mni.mcgill.ca/brainweb
121. Biomedical informatics research network
www.nbirn.net/index.shtml
122. MedGrid Project
www.medgrid.org
123. NeuroGrid Technology for Neuroscience
www.neurogrid.ac.uk
124. PsyGrid: UK mental health research into the future
www.psygrid.org
125. NeuroPsyGrid Project
www.neuropsygrid.com
126. Brain Imaging Network Grid
www.brainimaging.pt/paginas_frontoffice/frontoffice_home.php
127. NeuroLOG Project
http://neurolog.polytech.unice.fr/doku.php?id=neurolog

Affiliations

- Alberto Redolfi
Fatebenefratelli – Centro San Giovanni di Dio, Laboratory of Epidemiology & Neuroimaging, Via Pilastroni 4, I-25125 Brescia, Italy
Tel.: +39 030 350 1311
Fax: +39 030 350 1313
aredolfi@fatebenefratelli.it
- Richard McClatchey
University of the West of England, The Centre for Complex Cooperative Systems, Frenchay Campus, Coldharbour Lane, Bristol BS16 1QY, UK
Tel.: +44 117 328 3176
Fax: +44 117 328 2734
richard.mcclatchey@uwe.ac.uk
- Ashiq Anjum
University of the West of England, The Centre for Complex Cooperative Systems, Frenchay Campus, Coldharbour Lane, Bristol BS16 1QY, UK
Tel.: +44 117 328 3176
Fax: +44 117 328 2734
ashiq.anjum@cern.ch
- Alex Zijdenbos
Prodema Medical, Industriestrasse 6B, PO Box 51, 9620 Bronschhofen, Switzerland
Tel.: +1 514 804 4108
Fax: +1 708 810 3309
alex@prodemamedical.com
- David Manset
maat Gknowledge, Immeuble Alliance Entrée A, 74160 Archamps, France
Tel.: +33 450 439 602
Fax: +33 450 439 601
dmanset@maat-g.com
- Frederik Barkhof
VU University Medical Center, Department of Radiology, De Boelelaan 1118, 1081 HV Amsterdam, The Netherlands
Tel.: +31 204 440 365
Fax: +31 204 440 397
f.barkhof@vumc.nl
- Christian Spenger
Prodema Medical, Industriestrasse 6B, PO Box 51, 9620 Bronschhofen, Switzerland
Tel.: +41 719130580
Fax: +41 71 913 0589
christian.spenger@prodema.ch
- Yannik Legré
HealthGrid, 36 rue Charles de Montesquieu, F-63430 Pont-du-Château, France
Tel.: +33 473 405 155
Fax: +33 473 405 002
yannik.legre@healthGrid.org
- Lars-Olof Wahlund, Karolinska Institutet, Stockholm, Department of Neurobiology, Caring Sciences & Society, Division of Clinical Geriatrics Novum 5th floor, 141 86 Stockholm, Sweden
Tel.: +46 858 585 419
Fax: +46 858 585 470
lars-olof.wahlund@ki.se
- Chiara Barattieri di San Pietro Fatebenefratelli – Centro San Giovanni di Dio, Laboratory of Epidemiology & Neuroimaging, Via Pilastroni 4, I-25125 Brescia, Italy
Tel.: +39 030 350 1311
Fax: +39 030 350 1313
barattieri@fatebenefratelli.it
- Giovanni B Frisoni, MD Fatebenefratelli – Centro San Giovanni di Dio, Laboratory of Epidemiology & Neuroimaging, Via Pilastroni 4, I-25125 Brescia, Italy
Tel.: +39 030 350 1261
Fax: +39 030 350 1313
gfrisoni@fatebenefratelli.it