Virtual imaging laboratories for marker discovery in neurodegenerative diseases

Giovanni B. Frisoni, Alberto Redolfi, David Manset, Marc-Étienne Rousseau, Arthur Toga and Alan C. Evans

Abstract | The unprecedented growth, availability and accessibility of imaging data from people with neurodegenerative conditions has led to the development of computational infrastructures, which offer scientists access to large image databases and e-Science services such as sophisticated image analysis algorithm pipelines and powerful computational resources, as well as three-dimensional visualization and statistical tools. Scientific e-infrastructures have been and are being developed in Europe and North America that offer a suite of services for computational neuroscientists. The convergence of these initiatives represents a worldwide infrastructure that will constitute a global virtual imaging laboratory. This will provide computational neuroscientists with a virtual space that is accessible through an ordinary web browser, where image data sets and related clinical variables, algorithm pipelines, computational resources, and statistical and visualization tools will be transparently accessible to users irrespective of their physical location. Such an experimental environment will be instrumental to the success of ambitious scientific initiatives with high societal impact, such as the prevention of Alzheimer disease. In this article, we provide an overview of the currently available e-infrastructures and consider how computational neuroscience in neurodegenerative disease might evolve in the future.

Frisoni, G. B. *et al. Nat. Rev. Neurol.* advance online publication 5 July 2011; doi:10.1038/nrneurol.2011.99

Introduction

Research in neurodegenerative diseases is undergoing a radical transformation brought about by extraordinary growth in the volume, availability and accessibility of clinical and research imaging data, both in the form of public releases and within virtual research organizations. Traditional neuroimaging research typically involved small to mid-sized locally collected data sets ranging from dozens to hundreds of scans. Only a few imaging laboratories have the technical expertise and computational resources required to merge multiple large data sets and explore scientific questions relating to larger populations. Not only do neuroscientists face a steep learning curve to grasp their own particular computing ecosystem, in terms of operating system environment, basic scripting, programming, remote data transfers and remote computing, but also, because of divergence in the basic information technology (IT) setup, the principles of one ecosystem often do not adapt well to other laboratories. The commonplace replication and idiosyncrasies of toolsets and infrastructures among many sites greatly increases the complexity and overheads for neuroimaging projects, leading to issues such as the need to locally support IT-related technical staff, and difficulties in coordinating multisite studies.

Open access to large data sets, pioneered in genetics and physical sciences, has been implemented successfully by various initiatives in the neuroimaging field, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI)[1] and the NIH Pediatric Database (NIHPD).[2] Since 2004, all researchers who subscribe to these databases have been able to obtain full access to images and clinical data from people with varying degrees of cognitive deterioration that were originally collected to identify biomarkers of disease initiation and progression.[3,4] Currently, a number of large to very large data sets can be found in the public domain and freely downloaded, such as the 1000 Functional Connectomes Project,[5] the Human Imaging Database (HID),[6] the Open Access Series of Imaging Studies (OASIS),[7] the Bipolar Disorder Neuroimaging Database (BiND),[8] Multisite Imaging Research In the Analysis of Depression (MIRIAD),[9] and Efficient Longitudinal Upload of Depression in the Elderly (ELUDE).[10]

The gap between the pace of data generation and the capability to extract clinically or scientifically relevant information is rapidly widening. Sophisticated algorithms are available, and more are being developed, that allow the extraction of biologically relevant markers from images and clinical data requiring heavy computations. For instance, the extraction of the three-dimensional cortical thickness map, a marker of neurodegeneration, from a high-resolution structural MRI scan can take between 30 min and 22 h per scan on a single-core computer, and extraction of functional connectivity networks can take 20–120 min. At present, relatively few imaging laboratories worldwide have the expertise and resources required for such sophisticated high-throughput computational

Competing interests
The authors declare no competing interests.

IRCCS Fatebenefratelli, Via Pilastroni 4, 25125 Brescia, Italy (**G. B. Frisoni, A. Redolfi**). MAAT France, Immeuble Alliance, Entrée A 74160 Archamps, France (**D. Manset**). Montreal Neurological Institute at McGill University, Sherbrooke Street West 845, Montreal, QC H3A 2T5, Canada (**M.-E. Rousseau, A. C. Evans**). Laboratory Of Neuro Imaging, UCLA School of Medicine, Neuroscience Research Building, Suite 225, 635 Charles E. Young Drive South, Los Angeles, CA 90095-7334, USA (**A. Toga**).

Correspondence to: G. B. Frisoni gfrisoni@fatebenefratelli.it

Key points

- Image data sets of unprecedented size from healthy and pathologically aging individuals are posing new challenges related to availability and accessibility of data, computational resources, and visualization tools
- Scientific e-infrastructures based on grid computing, such as LONI, neuGRID, and CBRAIN, offer a suite of services to facilitate advanced computational analyses on brain images
- In the neurodegenerative disease field, such e-infrastructures are critical to foster the development of disease markers for early diagnosis and to track the course of the disease in clinical trials
- Steps have been taken towards convergence of the individual infrastructures into a worldwide, cloud-based global virtual imaging laboratory

Box 1 | e-Science and e-infrastructures

e-Science is defined as "computationally intensive science that is carried out in highly distributed network environments, or science that uses immense data sets that require grid computing."[42] The term 'e-Science' encompasses technologies that enable distributed collaboration, and was coined by John Taylor, the Director General of the UK Office of Science and Technology in 1999. In addition to computational neuroscience (Box 2) and bioinformatics, e-Science has been applied to social simulations, particle physics and earth sciences. Owing to the complexity of the software and the infrastructural requirements, e-Science projects almost invariably involve large teams coordinated by research laboratories, large universities or governments. e-Science requires specific environments, known as e-infrastructures, to manage and process data. These infrastructures exploit information and communication technology facilities and services, providing all researchers—whether working within their home institutions or in national or multinational scientific initiatives—with shared access to unique or distributed scientific facilities (including data, instruments, computing and communications).

Box 2 | Computational neuroscience

Computational neuroscience is defined as "the study of brain function in terms of the information processing properties of the structures that make up the nervous system."[43] This interdisciplinary science bridges the gap between neuroscience, cognitive science and psychology, and electrical engineering, computer science, mathematics and physics. The term 'computational neuroscience' was introduced by Eric L. Schwartz in 1990 following a conference on neural modeling, brain theory and neural networks. Computational neuroscience aims to describe the physiology and dynamics of functionally and biologically realistic neurons and neural systems. The resulting models encapsulate the fundamental features of the biological system on multiple spatiotemporal levels, ranging from membrane currents and protein and chemical coupling, through network oscillations, columnar and topographic architecture and structure, to learning and memory. The models can be used to frame hypotheses, which can subsequently be tested by biological or psychological experiments. In the field of neurodegenerative diseases, the aims of computational neuroscience are to develop unidimensional or multidimensional models of the brain changes that take place over time at the molecular, neuronal and glial, gray and white matter, and whole-brain levels.

imaging analyses in large databases. Clearly, the traditional way will no longer be efficient or sustainable when hundreds of scientists worldwide wish to perform these analyses on thousands of brain images.

In Europe and North America, e-Science infrastructures are being developed to fill the gap between data acquisition and information extraction (Box 1). Particle physics has a particularly well-developed e-Science infrastructure owing to its need for adequate computing facilities for the analysis of results and storage of data originating from the CERN Large Hadron Collider, but neuroimaging is quickly catching up.[3] Neuroimaging e-Science infrastructures such as Laboratory of Neuro Imaging (LONI) at the University of California, Los Angeles (UCLA),[11] neuGRID,[12] and CBRAIN[13] offer access to large databases, sophisticated algorithms for image analysis, computational resources, and statistical and data visualization tools.[14] Access to such novel infrastructures can be provided through web browsers or services or via Linux command line interfaces. The range of databases and algorithms is markedly variable, and computational resources are based on either a central server or cluster or a distributed grid infrastructure.

Presently, we are in the very early days of public services for computational neuroscience (Box 2), and the current infrastructures might undergo substantial reshaping in the near future. However, it is relevant for neuroscientists to be aware of what is available today, as these infrastructures can be to neuroscientists what the Large Hadron Collider is to physicists; that is, the laboratory where the most 'muscular' experiments can be run and audacious hypotheses can be tested. These scientific infrastructures can be instrumental to the success of extremely ambitious initiatives recently launched, such as Prevent Alzheimer's Disease by 2020 (PAD 2020),[15,16] a political and scientific effort aiming to achieve an effective treatment to prevent the disease in asymptomatic or mildly symptomatic cases.

In this Review, we provide an overview of the structure, services and current capabilities of the LONI, neuGRID and CBRAIN infrastructures. We provide an example of a scientific question that can be answered by running computationally demanding analyses in the context of these infrastructures, as well as outlining a possible scenario of what computational neuroscience in neurodegenerative diseases might look like in the near future. A glossary of some of the specialist terms used in the article is provided in Box 3.

Virtual imaging laboratories

Following the advent of MRI, it rapidly became clear that stereotactic imaging would be an exceptionally powerful tool to explore the brain and for clinical use (diagnosis, prognosis and disease tracking). Early neuroimaging efforts focused on the processes of image acquisition, data management and independent structural or functional analyses of normal development or specific cognitive disorders. Later efforts addressed clinically driven research hypotheses by means of integrated multimodal imaging. Until recently, however, the considerable demands for high-level neuroscientific, engineering, computational and technical expertise and the need for specialized hardware infrastructure have limited the scope of the applications to large monolithic and centralized research centers. Brain mapping is a multidisciplinary research field where basic, applied and clinical sciences converge to address important human health challenges. Integration of the power of sophisticated mathematical models, efficient computational algorithms and advanced hardware infrastructure provides the necessary sensitivity to detect, extract and analyze subtle, dynamic and distributed patterns distinguishing one normal brain from another, and a diseased brain from a normal brain.

The potential for integrated services offering neuroscientists all the major components for imaging experiments (that is, data, algorithms, computational resources, and statistical tools) has remained below threshold pending two developments: first, harmonization of image acquisition to allow the pooling of scans acquired from scanners of different model and manufacturer, and second, a novel policy of unrestricted data access. The ADNI effort[1] represents the successful implementation of such a policy. ADNI has been interested in gray matter atrophy as a marker of neurodegeneration in people with early Alzheimer disease (AD), and a number of protocols for the acquisition of high-resolution 1.5 T and 3.0 T structural MRI scans with similar signal-to-noise ratio and gray–white matter contrast were developed for 59 different scanners from the three main manufacturers (GE Healthcare, Philips Medical Systems and Siemens Medical Solutions). These protocols allowed the design of experiments that pooled scans acquired on scanners of different model and manufacturer. Harmonization efforts have been completed or are under way for other acquisition modalities in the context of other initiatives, which may soon lead to the creation of large multi-scanner data sets of spectroscopic MRI,[17] diffusion MRI,[18] and resting-state functional MRI (fMRI).[19]

Importantly, the public access policy of ADNI, which imposes no embargo period, thereby permitting virtually anyone in the world to download the whole image data set, has led to its extensive scientific use. At the time of writing, about 150 scientific manuscripts had been published on the ADNI data[1] by 933 investigators, at 177 research centers, from six economic sectors, in 35 countries.

An initial effort to promote the adoption of neuroimaging informatic resources, data and tools was started by the NIH through the public launch of the Neuroimaging Informatics Tools and Resources Clearing house (NITRC)[20] in October 2007. The mission of NITRC was to provide a user-friendly knowledge environment for fMRI and structural imaging analyses. The NITRC website hosts tools and resources, vocabularies, and databases for MRI research, thereby extending the impact of previously locally funded neuroimaging informatics contributions to a broader community.[21]

A further step forward was represented by the shift from centralized to distributed platforms. Two examples of these evolutionary infrastructural changes are the French NeuroLog project[22] and the Centre pour l'Acquisition et le Traitement de l'Image (CATI). NeuroLog was one of the first projects to invest in grid technologies for neurosciences. Its primary objectives were to extend the computing infrastructures deployed within French brain imaging centers, and to provide a country-wide platform dedicated to neuroscience and address the challenges raised by modern large-scale statistical studies.[23] CATI has recently been funded to provide assistance for acquiring, analyzing, organizing and sharing neuroimaging data among scientific and medical communities working on AD. The CATI initiative will offer a complete portfolio of image processing tools, including international standards like voxel-based and tract-based morphometry, as well as distributed database services. Via these services, the CATI initiative will mutualize the resources and offer valid support through experts. The tools and services of all these projects have been developed to adhere to the ADNI standards.[24]

These new scenarios have prompted the birth and growth of international service infrastructures to help scientists to cope with public data sets of unprecedented size. The three initiatives that will be described in the sections that follow (Table 1) share the common vision of offering a full range of imaging techniques to non-imaging neuroscientists by offering easy access to data, algorithms,

Box 3 | Glossary

Algorithm

A set of steps to accomplish a particular task implemented in a single software step or a series of steps.

Atomic modules

Individual modules that make up complex workflows.

Cloud computing

A type of distributed computing infrastructure (DCI), cloud computing is a web-based processing infrastructure, whereby shared resources, software and information are provided to computers and other devices (such as smartphones) on demand over the Internet.

Constrained Laplacian Anatomic Segmentation using Proximity (CLASP)

A fully automatic method to reconstruct the brain pial surface. This algorithm uses a complex classification method with statistical probabilistic anatomical maps and geometric deformable surface models. The gray matter surface is initiated from the white matter surface and is expanded to the boundary between gray matter and cerebrospinal fluid along the Laplacian force field.

Grid computing

A type of distributed computing infrastructure where the system is created by forming a virtual organization over geographically distributed heterogeneous clusters. Commonly used grid middleware include gLite and Globus.

Graphical User Interface (GUI)

A human–computer interface that uses windows, icons and menus, and can be manipulated by a computer mouse.

High Performance Computing (HPC)

A type of DCI that uses supercomputers and computer clusters to solve advanced computation problems.

Pipeline

Also known as a workflow, a pipeline is a software implementation with a well-defined input and output. For example, the input may be two three-dimensional MRI scans of a person's brain acquired 1 year apart, and the output may be the percentage change in the brain's volume over the year. A pipeline can consist of one or more algorithms and other software steps drawn from one or more toolkits that may also generate intermediate data.

UNIX

A multitasking, multi-user computer operating system.

Web portal

Web portals present information from diverse sources in a unified way. They offer many services, including e-mail, information and databases. Portals provide a consistent look and feel with access control and procedures for multiple applications and databases.

Table 1 | Core features of the three e-infrastructures

Feature	neuGRID	LONI	CBRAIN
Image data	On AD and aging	On AD and aging or user provided	On pediatric brain, AD and aging, or user-provided
Public brain atlases	None	17 multimodal human and animal brain atlases for a number of diseases, created through registration and warping, indexing schemes and nomenclature systems	Age-and-disease-appropriate three-dimensional probabilistic atlases
Image processing algorithms	For structural MRI analysis	For structural, functional and diffusion imaging analysis	For structural and functional MRI analysis, connectivity analysis
Statistical tools	R statistics	12 different tools covering data classification, linear and nonlinear regression, feature selection, and multivariate analysis	R statistics and the RMINC package; integrated voxel-based statistics and voxel-wise or vertexwise general linear models; for example, fMRIstat, SurfStat
Workflow management system	LONI Pipeline and Kepler	LONI Pipeline	CBRAIN Workflow Engine
Graphical user interface	Secure Web portal with LifeRay technology[12]	Pipeline Server interface installed on local computer[25]	Secure web portal, with HTML 5 and WebGL 3D visualization capabilities[44]

Abbreviations: AD, Alzheimer disease; HTML, HyperText Markup Language; LONI, Laboratory Of Neuro Imaging; WebGL 3D, three-dimensional Web Graphic Language.

computational resources, and statistical tools. A use case vignette will help the reader to appreciate the advantages of performing computational neuroimaging on these e-Science platforms.

LONI (USA)

LONI focuses on the development of image analysis methods and their application in health as well as in neurological and psychiatric disorders.[11] LONI hosts the large ADNI database (among many others), comprising clinical and genetic information as well as scans from 400 older people with mild cognitive impairment, 200 people with AD, and 200 healthy elders, all of whom are being followed semiannually for 3 years with high-resolution structural MRI, ^{18}F-fluorodeoxyglucose PET (FDG-PET) and, in the near future, amyloid PET, fMRI and diffusion tensor imaging. Algorithms for data analysis are accessible both independently and through the graphical LONI Pipeline,[25] a user-friendly workflow management system. The LONI Pipeline enables automated measurement of functional and morphometric analyses, dynamic assessment of volume, shape (for example, curvature) and form (for example, thickness) features, as well as the extraction and association between cognitive, genetic, clinical, behavioral and imaging biomarkers. For external investigators, LONI provides access to a large High Performance Computing (HPC) infrastructure, physically located at UCLA, for computationally intensive image analyses. Access to the LONI HPC resources to external investigators is granted on the basis of ad hoc scientific collaboration agreements. Access spans not only the Image Data Archive (IDA), but also all other published data sets.

neuGRID (Europe)

The neuGRID[12] platform makes use of grid services and computing, and was developed with the final aim of overcoming the hurdles that the average scientist meets when trying to set up advanced experiments in computational neuroimaging, thereby empowering a larger base of scientists. Funded by the European Commission, the prototype version was completed in January 2011. Although originally built for neuroscientists working in the field of AD, as is reflected in the currently available services, the infrastructure is designed to be expandable to services from other medical fields and is compliant with European Union and international standards for data collection, data management and grid abstraction. An expansion—Diagnostic Enhancement of Confidence by an International Distributed Environment (DECIDE)[26] —has also been funded by the European Commission to include image analysis tools for clinical users; that is, tools sensitive to the departure of single cases from a normative reference image database. This principle applies in pattern recognition[27] for the differential diagnosis of AD from frontotemporal dementia, dementia with Lewy bodies, or normal aging on the basis of FDG-PET, structural MRI scans, or ACM-AdaBoost,[28] an intelligent algorithm that can automatically segment the hippocampus on high-resolution structural MRI to map hippocampal atrophy, a recognized diagnostic marker of AD progression. Currently, neuGRID provides external investigators with access to its distributed infrastructure following ad hoc cooperation agreements.

CBRAIN (Canada)

CBRAIN[13] is a network of Canada's five leading brain imaging research centers linked within a platform for distributed processing and data sharing. The CBRAIN platform addresses issues of advanced networking, transparent access to remote computer resources, integration of heterogenous environments, tool usability, and web-based three-dimensional visualization by providing users with a comprehensive collaborative web portal enabling them to manage, transfer, share, analyze and visualize their imaging data. Because of its distributed nature and ease of use, the CBRAIN platform connects five Canadian brain imaging research centers not only to seven HPC centers spread across Canada and Europe, but also to multiple collaborating sites around the world. CBRAIN provides a generic framework into which almost any processing pipeline or e-Science tool can be

Table 2 | Image data sets available in the three infrastructures

Data set	Characteristics	Accessibility	Objectives and impact
LONI			
ADNI-1	200 healthy older people, 400 patients with MCI, and 200 patients with AD; structural MRI serial scans at 1.5 T every 6 months for 4 years in 100% and at 3 T in 50%, FDG-PET every 12 months in 50%, serial CSF markers in >60%, genome-wide scan in 100%, amyloid imaging with ^{11}C-PIB in a restricted group; all data are de-identified	Public	Develop a uniform standard method for acquiring longitudinal biomarker data to better understand and characterize AD progression; develop markers to track disease progression for use as surrogate outcomes in clinical trials of disease-modifying drugs
ADNI-GO	Expands ADNI-1 by 200 additional patients with early MCI; all patients undergo structural MRI scans at 3 T at four time points, amyloid imaging with a fluorinated ligand, resting-state fMRI in Philips scanners, diffusion MRI in GE scanners, and CSF studies	Public	Better explore earlier stages of MCI; to study novel imaging markers
ADNI-2	ADNI-2 will study and follow 500 additional individuals	Public	Extend the observation window of the MCI stage to earlier and later stages; to leverage an integrated combination of clinical–cognitive, CSF–plasma biomarker, MRI, amyloid–FDG-PET, and genetic measures for early diagnosis and disease tracking
Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL)	1,000 individuals aged over 60 years have been studied, 285 of whom (including controls and patients with MCI or AD) have been published through the LONI Image Data Archive; data consist of structural MRI and ^{11}C-PIB PET imaging, neuropsychological scores, and blood analyses	Public	Improve understanding of the causes and diagnosis of AD, develop markers to monitor disease progression, and formulate hypotheses and interventions with respect to lifestyle factors that might delay disease onset
International Consortium for Brain Mapping (ICBM)	Multisite project that developed probabilistic human MRI, fMRI, MR angiography, DTI and FDG-PET brain atlases from 452 individuals aged 18–90 years	Public	Continuing development of a probabilistic reference system for structural and functional, macroscopic (*in vivo*), and microscopic (postmortem) anatomy of the human brain
PAD/CRYO	Anonymized MRI data from three normal control patients paired with digitalized histological data	Public	Imaging–histological reference correlations
neuGRID			
ADNI	ADNI through LONI	Public	See LONI
CBRAIN			
NIH pediatric MRI data repository	Longitudinal structural MRI, MR spectroscopy, DTI and correlated clinical–behavioral data from around 500 healthy, normally developing children, ages newborn to young adult	Public	Foster a better understanding of 'normal' as a basis for understanding atypical brain development associated with a variety of disorders and diseases
AddNeuroMed	Serial, multicenter, 1.5 T structural MRI study of 250 healthy elders, and 250 AD and 250 MCI patients scanned at baseline, 3, 6 and 12 months, then annually for an additional 2 years; MR spectroscopy in humans and transgenic animal models of AD complement these data; proteomic, genomic and lipidomic data are available	Shared, proprietary	Improve experimental models of AD for biomarker discovery, and identify biomarkers for AD that are suitable for early diagnosis, prediction of the development of dementia in patients with MCI, and monitoring of disease progression for use in clinical trials and practice

Abbreviations: AD, Alzheimer disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; ADNI-GO, ADNI Grand Opportunities; CRYO, Cryosection Imaging; CSF, cerebrospinal fluid; DTI, diffusion tensor imaging; FDG, ^{18}F-fluorodeoxyglucose; fMRI, functional MRI; MCI, mild cognitive impairment; MR, magnetic resonance; PAD, Public Anonymized Dataset; PIB, Pittsburg compound B.

connected. Researchers can then launch their jobs through an easy-to-use web interface, and allow the platform to handle data transfers, job scheduling on HPC, and results. CBRAIN currently offers full computing resources only to investigators within its network of centers.

CBRAIN is funded by CANARIE,[29] a Canadian government-supported nonprofit corporation, which maintains a set of leased high-speed wide area network links, CAnet, and also develops and deploys advanced network applications and technologies for education and high-speed data transfer purposes. GBRAIN, the international extension of the CBRAIN platform, connects international brain research partners located in the UK and Germany.

Commonalities and specificities

Despite the common vision of opening up the imaging laboratory to the non-imaging specialist, the three infrastructures were designed and developed at different times and in different scientific contexts to address specific contingent needs. As a consequence, while they have many commonalities, they also have differences regarding the types of imaging data sets that they offer, algorithm pipelines and tools, computational resources, and related services.

The imaging data made available by LONI are focused on AD and aging, while CBRAIN also encompasses brain development. The neuGRID platform is not home to its own data set; rather, it allows processing of the ADNI data set that can be accessed through LONI (Table 2). Being the first of the platforms to emerge, LONI offers the largest range of algorithms for skull stripping, brain registration, segmentation, feature analysis, statistical analysis, and visualization. CBRAIN offers many Montreal Neurological Institute algorithms, as well as commonly used external packages such as Statistical Parametric Mapping (SPM),[30] which is adapted for batch processing of large databases. The neuGRID platform offers packages for preprocessing

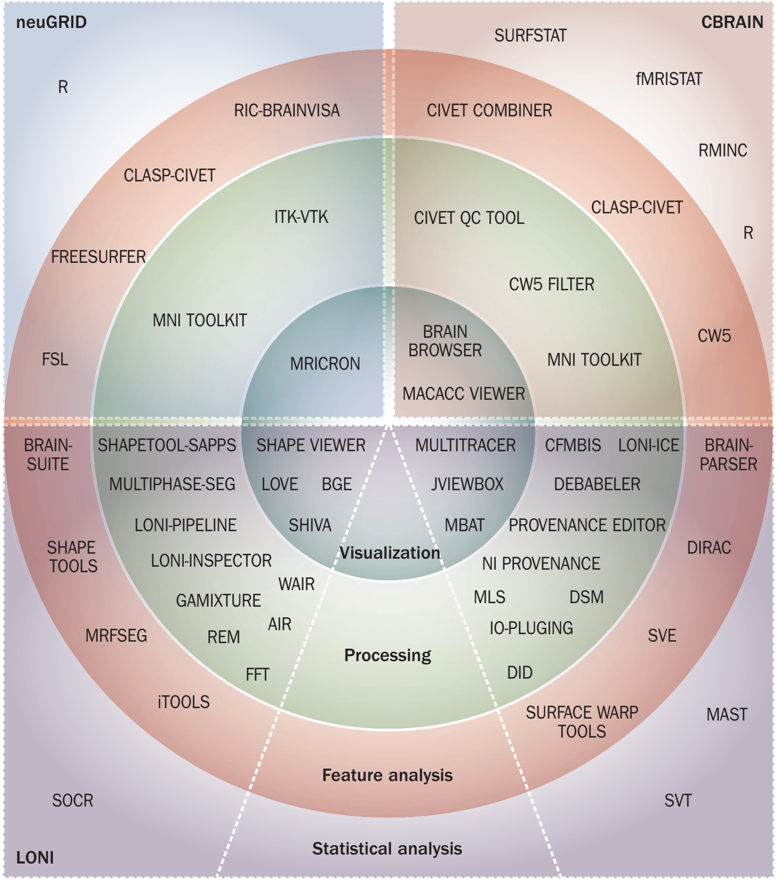


Figure 1 | Image-processing algorithms, suites and tools available in the LONI, neuGRID and CBRAIN infrastructures. The analysis tools provided by the three infrastructures are categorized into four classes. Visualization tools are applications that enable the visualization of medical images of different modalities (for example, MRI, PET, diffusion tensor imaging and functional MRI) and file formats (for example, .dcm, .nii, .hdr/img and .mnc). Processing tools are applications that enable transformation of the DICOM (Digital Imaging and Communications in Medicine) images into three-dimensional volume stacks, registration of three-dimensional stacks to templates, and reduction of inhomogeneities and magnetic field artifacts. Feature analysis tools are applications that enable quantitative assessment of properties of specific brain regions; for example, volumes, voxel classification or surface features. Statistical analysis tools are applications that enable the statistical assessment of the quantitative features extracted with feature analysis tools. Some of the statistical tools are applicable to single-subject analysis and others to group studies. An extensive description of the tools and exploded acronyms can be found in Supplementary Table 1 online. Abbreviation: LONI, Laboratory of Neuro Imaging.

and post-processing of structural brain scans (Figure 1, Supplementary Table 1 online).

The three infrastructures offer computing power and storage capacity that benefit from the combination of distributed resources, such as the grid, regular HPC and public clouds, to increase the overall performance. All three are fairly generic platforms that can support any new package of broad interest to the scientific and clinical communities. It takes typically 2 days to 1 week to incorporate stable new processing packages and make them available to the user community. The three infrastructures are interconnected via GEANT/CAnet/Internet2 networks, which offer the possibility of efficiently exchanging massive data sets.

Use case: biomarker validation

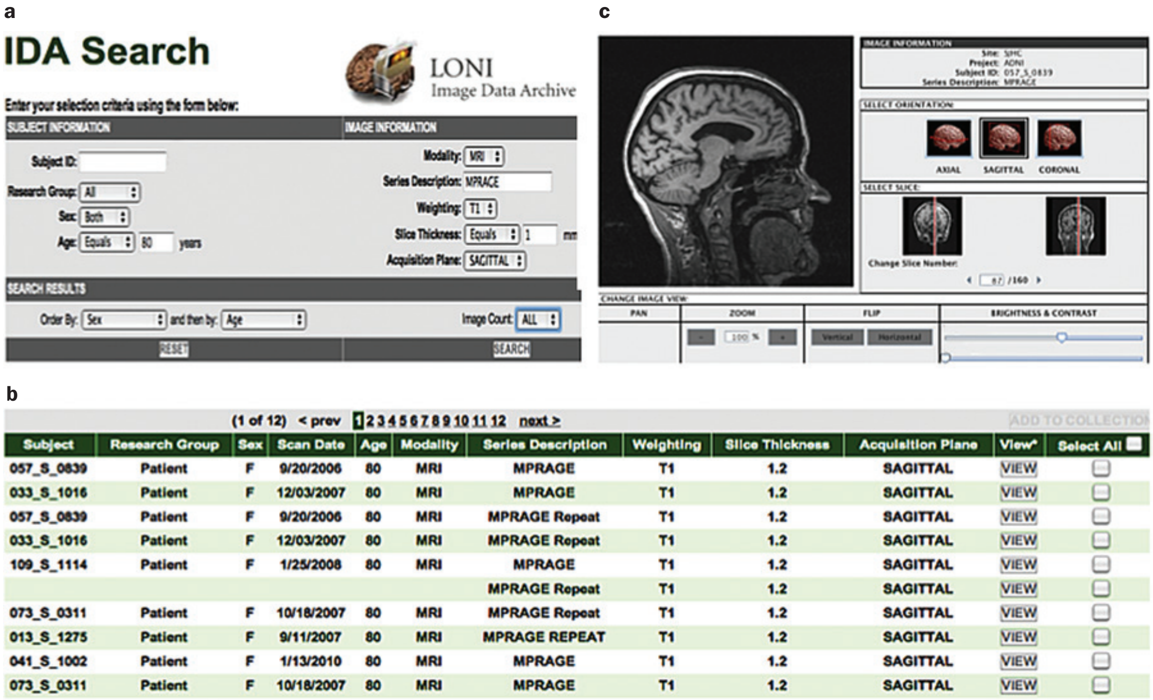
A neuroscientist wishes to make a head-to-head comparison of the available methods for estimating the thickness of the cortex in normal and pathological aging. Cortical thinning is a recognized marker of neurodegeneration,[31,32] a putative marker of disease progression,[33] and a reasonable surrogate outcome in clinical trials.

Three of the most popular automated algorithm pipelines for estimating cortical thickness are FreeSurfer, CIVET and RIC-BrainVISA. CIVET reconstructs the cortical thickness by identifying the gray–white matter and gray matter–CSF junctions.[34] FreeSurfer reconstructs the cortical surface through tessellation of the gray–white matter boundary following intensity gradients.[35] RIC-BrainVISA computes a Euclidean average distance from the outer gray matter mesh to the inner cortical white matter mesh.[36,37] The algorithmical differences are reflected into machine time, ranging from 0.5–24.0 h.

The neuroscientist is interested in the stability of the three algorithm pipelines to random noise in the image acquisition phase, and the sensitivity of the pipelines to age-associated and AD-associated structural changes of the cortical mantle. To this end, the neuroscientist accesses the high-resolution T1-weighted 1.5 T structural MRI scans of the ADNI-1 data set hosted by LONI, which comprises 9,250 individual brain images of 200 healthy older people, 400 patients with mild cognitive impairment, and 200 patients with AD, all of whom were scanned at baseline and every 6 months thereafter up to 48 months. Two back-to-back identical acquisitions were taken at each time point.

The neuroscientist selects images through an efficient database interface (the Image Data Archive, or IDA;[38] Figure 2), which exploits secure authentication and grants users immediate access to all data. After downloading scans from the e-Science database, the neuroscientist can specify pipeline analyses using an ad hoc workflow management system whereby the downloaded scans can be immediately accessed. By means of the intuitive visual programming graphic user interface (GUI; Figure 3), the neuroscientist can easily customize workflows and link modules, edit the flow of a predesigned pipeline, and replace modules. The workflow management system presents predefined modules and pipeline analyses to the researcher in organized tree structures. The module inputs and outputs are connected to form a complete pipeline. Specific inputs and outputs should be defined as a pipeline is created. The GUI allows the neuroscientist to submit jobs to the grid and at the same time monitor the execution of the launched jobs.

The neuroscientist is aware that the three algorithm pipelines possess diverse input and output requirements, utilize different file formats, run in specific environments as UNIX, Linux or IRIX file systems, and have limited capacities to read certain types of data usually developed in different laboratories. The input and output of individual modules of a pipeline may not be compatible with each other. However, the combination of different modules is no longer a problem because the interoperability issue has been solved in these emerging e-Science infrastructures:



Subject	Research Group	Sex	Scan Date	Age	Modality	Series Description	Weighting	Slice Thickness	Acquisition Plane	View*	Select All
057_S_0839	Patient	F	9/20/2006	80	MRI	MPRAGE	T1	1.2	SAGITTAL	VIEW	
033_S_1016	Patient	F	12/03/2007	80	MRI	MPRAGE	T1	1.2	SAGITTAL	VIEW	
057_S_0839	Patient	F	9/20/2006	80	MRI	MPRAGE Repeat	T1	1.2	SAGITTAL	VIEW	
033_S_1016	Patient	F	12/03/2007	80	MRI	MPRAGE Repeat	T1	1.2	SAGITTAL	VIEW	
109_S_1114	Patient	F	1/25/2008	80	MRI	MPRAGE	T1	1.2	SAGITTAL	VIEW	
						MPRAGE Repeat	T1	1.2	SAGITTAL	VIEW	
073_S_0311	Patient	F	10/18/2007	80	MRI	MPRAGE Repeat	T1	1.2	SAGITTAL	VIEW	
013_S_1275	Patient	F	9/11/2007	80	MRI	MPRAGE REPEAT	T1	1.2	SAGITTAL	VIEW	
041_S_1002	Patient	F	1/13/2010	80	MRI	MPRAGE	T1	1.2	SAGITTAL	VIEW	
073_S_0311	Patient	F	10/18/2007	80	MRI	MPRAGE	T1	1.2	SAGITTAL	VIEW	

Figure 2 | End-to-end data sharing, databasing and data choosing with the IDA database.[38] **a** | Data search interface. **b** | Data selection and retrieving interface. **c** | Image viewer tool to make a quick quality control assessment of the selected images. Abbreviations: ADNI, Alzheimer's Disease Neuroimaging Initiative; IDA, Image Data Archive; LONI, Laboratory of Neuro Imaging.

the workflow management system takes care of many of the above problems, such as the conversion of different file formats (.dcm, .mnc and .nii) and the transparent management of inputs and outputs during any pipeline submission and execution. In addition, the neuroscientist knows that handling, organization and storage of the massive intermediate data output generated by workflows can prove difficult. After launching the jobs and using the neuGRID resources, the neuroscientist obtains their results in less than 7 weeks, compared with 10 years on a single mono-core computer. The 5 TB of result data are available for the user through a secure File Transfer Protocol (sFTP) connection.

Although the three algorithm pipelines produce cortical thickness data with heterogeneous formats, in the e-Science platform the neuroscientist can find a visualization tool that is compatible with all three formats. The visualization tool reads the result files and displays cortical surface maps in the same coordinate space and color reference system (Figure 4), allowing a direct visual comparison. After the visual inspection of a sample of maps, the neuroscientist runs a number of statistical tests with the 'R' software on all or part of the dataset, aiming to test the study hypotheses. Similar to the processing of the raw data, the statistical analyses produce algorithm-specific maps that can be transparently visualized.

Interoperability demonstrator

The interoperability demonstrator provides exemplar implementation of the capability of diverse platforms to work together. The rationale underpinning the demonstration is the possibility not only to exchange data, but also to generate meaningful results among LONI, neuGRID and CBRAIN. It consists of a 'super workflow' involving the synchronized and complementary use of distributed computing infrastructures and resources of the three platforms. The demonstrator will execute the CIVET cortical thickness extraction pipeline on three separate but compatible image data sets (to preserve the possibility of meaningfully merging the scientific results), each hosted in one of the three infrastructures. This demonstrator will be the first such challenge to be run across international neuroscience research infrastructures, with different technologies and environments, involving more than 2,000 central processing unit cores per execution cycle, and resulting in the largest computational analysis ever attempted in the field, with no fewer than 10,000 MRI scans being processed in parallel. The estimated image processing time is 3 days.

The super workflow is specified in a shared and harmonized authoring environment, the LONI Pipeline[39] graphical user interface,[40] which can talk to the three distributed computing infrastructures—CBRAIN, LONI and neuGRID—thanks to outGRID, an international cooperation project funded by the European Commission.[41]

The demonstrator will be achieved through active contribution from the three main infrastructures. CBRAIN will provide the CIVET cortical thickness extraction pipeline and access to computational resources, LONI will provide the workflow management system (LONI Pipeline) interface and access to computational resources, and neuGRID will use its integration middleware to enable all three infrastructures to interconnect and access its grid computing resources. At the time of writing, the

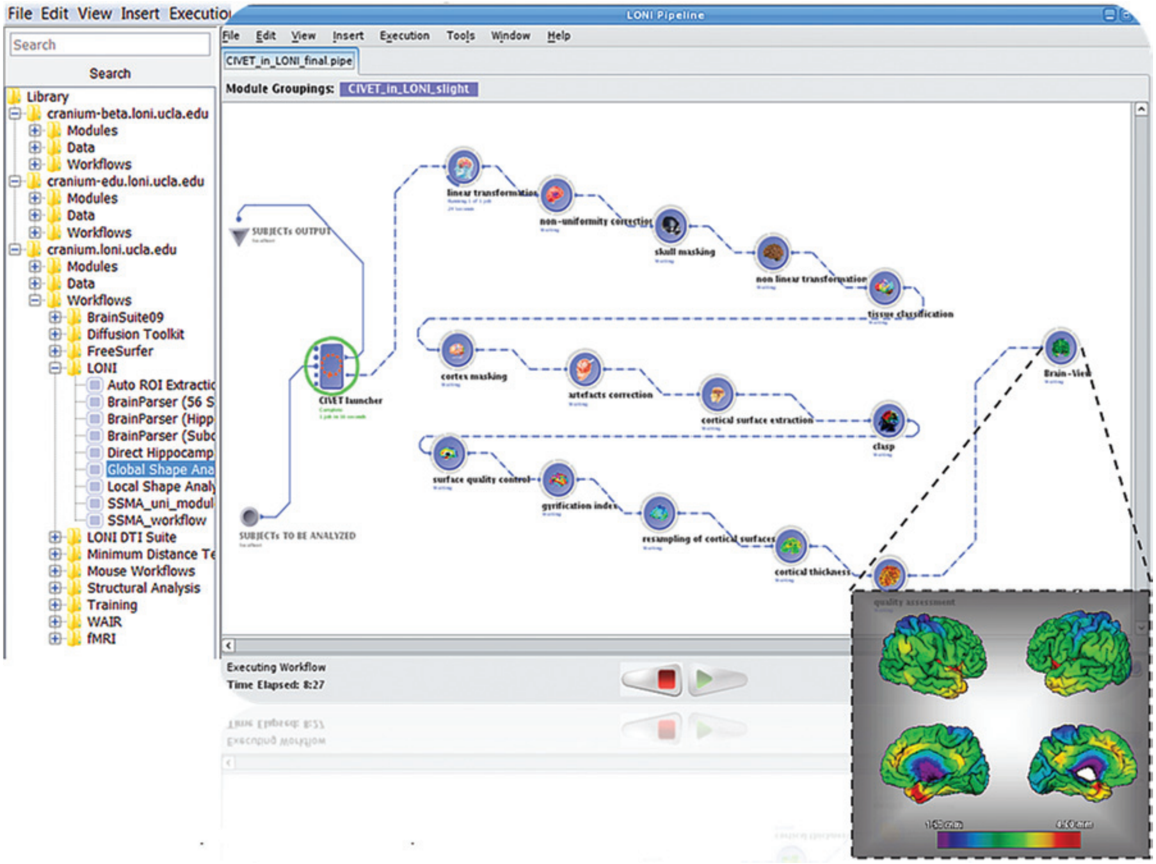


Figure 3 | Graphical representation of the CIVET cortical thickness extraction algorithm exposed through the LONI Pipeline Environment. The workflow is characterized by many atomic modules such as: MRI nonuniformity correction; linear registration; skull masking; tissue classification; cortical surface extraction; Constrained Laplacian Anatomic Segmentation using Proximity (CLASP); and cortical thickness estimation and visualization. Each module can be customized according to specific user needs. Abbreviation: LONI, Laboratory of Neuro Imaging.

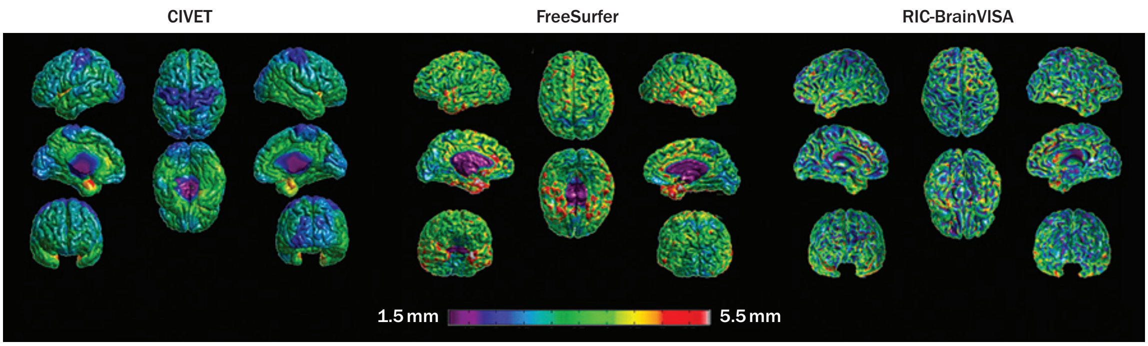


Figure 4 | Maps of mean cortical thickness in the Alzheimer's Disease Neuroimaging Initiative dataset obtained with CIVET, FreeSurfer and RIC-BrainVISA, and displayed with the same visualization tool.

demonstrator was earmarked for launch around July 2011. The results will be published and accessible on the outGRID website.[41]

Development of future services

An effort is ongoing to capitalize on the significant overlaps and redundancies among LONI, CBRAIN and neuGRID and to develop seamless and user-transparent interoperability (Table 3). This is a long-term multinational project that will lead to the development of a global virtual imaging laboratory. The aim is to offer computational neuroscientists a virtual space accessible through an ordinary browser, where image data sets and related clinical variables, algorithm pipelines, computational resources, and statistical and visualization tools will be transparently accessible to users irrespective of their own physical location. A single sign-on system will guarantee user-friendly but still privileged access to non-public resources.

Currently, the three infrastructures implement or integrate different technologies, formats and standards,

Table 3 | Hardware and connectivity features of the three e-infrastructures

Feature	neuGRID	LONI	CBRAIN
Infrastructure topology	Distributed	Centralized	Distributed
Accessibility	Hybrid	Private	Public
Paradigm used	Grid	HPC	Grid/HPC
Facilities	Three data analysis and computing sites	CRANIUM HPC and data center	Seven HPC centers and two main data centers
Physical server locations	Brescia (Italy); Stockholm (Sweden); Amsterdam (The Netherlands); Archamps (France)	UCLA (USA)	Montreal, Sherbrooke, Quebec City, Vancouver, Calgary, Toronto (Canada); Julich (Germany)
Storage capacity	7 TB (at FBF, VUmc, KI and MAAT) plus distributed storage	4 PB (at UCLA/LONI)	0.5 PB plus 0.5 PB distributed storage
Core computational resources	500 CPU cores	4,800 CPU cores	Over 45,000 CPU cores
Computational engine (middleware)	Grid (gLite)	Sun Grid Engine (SGE)	Data and Compute Grid (CBRAIN middleware)
External computational resource extension	EGI expansion (10,000 cores)	Not applicable	Juropa (Julich) HPC integration (26,000 cores)
Local computational resource extension	Desktop Fusion file sharing	File sharing	Data Providers
Network provider	GEANT	Internet2	CANET
Bandwidth	1 GB/s	20 GB/s (load balanced)	10 GB/s

Abbreviations: CPU, central processing unit; CANET; Collaborative Automotive NETwork; EGI, European Grid Infrastructure; FBF, Provincia Lombardo Veneta Ordine Ospedaliero di San Giovanni di Dio—Fatebenefratelli, Brescia, Italy; GB, gigabytes; GEANT, Gigabit European Advanced Network Technology; HPC, High Performance Computing; KI, Karolinska Institute, Stockholm, Sweden; LONI, Laboratory Of Neuro Imaging; MAAT, MAAT France, Archamps; PB, petabytes; TB, terabytes; UCLA, University of California, Los Angeles; VUmc, VU University Medical Center, Amsterdam, The Netherlands.

making it impossible to execute a given workflow from one infrastructure to the other, or even to interconnect resources management layers to allow pipeline environments to talk to each other's computing resources. The interoperability effort will leverage on the possibility of defining and executing pipelines through schematic representations hiding away all implementation details. Interoperability will be facilitated by the implementation of Web 2.0 technologies and applications (for example, LifeRay, AJAX and Java technologies) that facilitate participatory information sharing, interoperability, researcher-centered design and collaboration among the three infrastructures.

A notable service feature will be represented by the metadata and provenance information that will be made available to neuroscientists following image-processing experiments. Provenance is the process of tracking the origin and history of processed data, offering the possibility to easily reconstruct workflows, rerun previous executions, and validate intermediate and final results. Currently, provenance services in the three infrastructures rely on different schema and technologies that will also need to be made interoperable in the future.

The initial impetus for the interoperability exercise has been provided by outGRID,[41] setting the foundations for much larger research and development programs in the future that should lead to full interoperability. The outGRID demonstrator has led the way in the concept of virtual imaging laboratory interoperability.

Conclusions

Like many other fields, neuroimaging research is affected by the gap between the availability of digital data and tools to extract meaningful information. The availability of public image databases of unprecedented size has given rise to the need for research infrastructures that enable neuroscientists to access, query, process and statistically analyze these databases. e-Infrastructures have been and are being developed in Europe and North America, offering computational neuroscientists a suite of services. These infrastructures are seeking convergence towards a worldwide infrastructure that will constitute a global virtual imaging laboratory. Such an experimental environment will be instrumental to the success of ambitious scientific initiatives with high societal impact, such as PAD 2020.[16]

Review criteria

Articles were selected on the basis of the authors' personal knowledge and the following PubMed searches: "(Grid[ti] OR Virtual[ti]) AND laborator*[ti]) AND (((MR OR MRI) OR data*[ti]) OR (image OR imaging)))"; "Computational (infrastructure* OR analyses[ti]) AND Alzheimer"; "(Computing OR Computational) AND (infrastructure* OR analyses[ti]) AND Brain"; and "(Computing OR Computational infrastructure*) AND ("Alzheimer dementia"[ti] OR "frontotemporal dementia"[ti] OR "frontal lobe dementia"[ti] OR "frontotemporal lobar degeneration"[ti] OR "dementia with Lewy bodies"[ti] OR "Lewy body dementia"[ti] OR "Parkinson dementia"[ti])". Reference lists of the identified papers were examined for further leads. The search was limited to full-text manuscripts published in English over the past 10 years. The final selection was based on relevance, as judged by the authors.

1. Alzheimer's Disease Neuroimaging Initiative (ADNI) [online], http://adni.loni.ucla.edu/ (2011).
2. The NIH MRI Study of Normal Brain Development [online], https://nihpd.crbs.ucsd.edu/nihpd/info/index.html (2011).
3. [No authors listed] The scientific social network. *Nat. Med.* **17**, 137 (2011).
4. Frisoni, G. B. & Weiner, M. W. Alzheimer's Disease Neuroimaging Initiative special issue. *Neurobiol. Aging* **31**, 1259–1262 (2010).
5. Functional Connectomes Project [online], http://fcon_1000.projects.nitrc.org/ (2011).
6. Human Imaging Database [online], http://www.birncommunity.org/tools-catalog/human-imaging-database-hid/ (2010).
7. OASIS [online], http://www.oasis-brains.org/ (2011).
8. Bipolar Disorder Neuroimaging Database (BiND) [online], http://sites.google.com/site/bipolardatabase/ (2009).
9. Multisite Imaging Research In the Analysis of Depression (MIRIAD). *Biomedical Informatics Research Network* [online], http://www.birncommunity.org/data-catalog/multisite-imaging-research-in-the-analysis-of-depression-miriad/ (2010).
10. Efficient Longitudinal Upload of Depression in the Elderly (ELUDE). *Biomedical Informatics Research Network* [online], http://www.birncommunity.org/data-catalog/efficient-longitudinal-upload-of-depression-in-the-elderly-elude/ (2010).
11. Laboratory of Neuro Imaging, UCLA (LONI) [online], http://www.loni.ucla.edu/ (2009).
12. neuGRID [online], http://www.neugrid.eu/pagine/home.php (2011).
13. CBRAIN [online], http://cbrain.mcgill.ca/ (2008).
14. Redolfi, A. *et al.* Grid infrastructures for computational neuroscience: the neuGRID example. *Future Neurol.* **8**, 703–722 (2009).
15. Khachaturian, Z. S. & Khachaturian, A. S. Prevent Alzheimer's Disease by 2020: a national strategic goal. *Alzheimers Dement.* **5**, 81–84 (2009).
16. PAD 2020: The Campaign to Prevent Alzheimer's Disease by 2020 [online], http://www.pad2020.org/ (2009).
17. Jessen, F. *et al.* A multicenter ^{1}H-MRS study of the medial temporal lobe in AD and MCI. *Neurology* **72**, 1735–1740 (2009).
18. Teipel, S. J. *et al.* Longitudinal changes in fiber tract integrity in healthy aging and mild cognitive impairment: a DTI follow-up study. *J. Alzheimers Dis.* **22**, 507–522 (2010).
19. Dosenbach, N. U. *et al.* Prediction of individual brain maturity using fMRI. *Science* **329**, 1358–1361 (2010).
20. NITRC [online], http://www.nitrc.org/ (2011).
21. Luo, X. J., Kennedy, D. N. & Cohen, Z. Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC) resource announcement. *Neuroinformatics* **7**, 55–56 (2009).
22. NeuroLog [online], http://neurolog.polytech.unice.fr/doku.php (2007).
23. Montagnat, J. *et al.* NeuroLOG: a community-driven middleware design. *Stud. Health Technol. Inform.* **138**, 49–58 (2008).
24. Frisoni, G. B. Alzheimer's Disease Neuroimaging Initiative in Europe. *Alzheimers Dement.* **6**, 280–285 (2010).
25. LONI Pipeline [online], http://pipeline.loni.ucla.edu/ (2011).
26. Diagnostic Enhancement of Confidence by an International Distributed Environment (DECIDE) [online], http://www.eu-decide.eu (2011).
27. Klöppel, S. *et al.* Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. *Brain* **131**, 2969–2974 (2008).
28. Morra, J. H. *et al.* Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. *Neuroimage* **43**, 59–68 (2008).
29. CANARIE [online], http://www.canarie.ca/ (2009).
30. SPM [online], http://www.fil.ion.ucl.ac.uk/spm/ (2011).
31. Du, A. T. *et al.* Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia. *Brain* **130**, 1159–1166 (2007).
32. Avants, B. B., Cook, P. A., Ungar, L., Gee, J. C. & Grossman, M. Dementia induces correlated reductions in white matter integrity and cortical thickness: a multivariate neuroimaging study with sparse canonical correlation analysis. *Neuroimage* **15**, 1004–1016 (2010).
33. Risacher, S. L. *et al.* Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Curr. Alzheimer Res.* **6**, 347–361 (2009).
34. Kim, J. S. *et al.* Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. *Neuroimage* **27**, 210–221 (2005).
35. Fischl, B. *et al.* Automatically parcellating the human cerebral cortex. *Cereb. Cortex* **14**, 11–22 (2004).
36. Kochunov, P. *et al.* Relationship among neuroimaging indices of cerebral health during normal aging. *Hum. Brain Mapp.* **29**, 36–45 (2007).
37. Kochunov, P. *et al.* Can structural MRI indices of cerebral integrity track cognitive trends in executive control function during normal maturation and adulthood? *Hum. Brain Mapp.* **30**, 2581–2594 (2009).
38. The LONI Image Data Archive. *Alzheimer's Disease Neuroimaging Initiative* [online], https://ida.loni.ucla.edu/login.jsp?project=ADNI (2011).
39. Dinov, I. *et al.* Neuroimaging study designs, computational analyses and data provenance using the LONI pipeline. *PLoS ONE* **5**, pii: e13070 (2010).
40. LONI Pipeline Processing Environment. *LONI Software* [online], http://www.loni.ucla.edu/Software/Pipeline (2009).
41. outGRID [online], http://www.outgrid.eu/site/pagine/home.php (2011).
42. eScience. *Wikipedia* [online] http://en.wikipedia.org/wiki/E-Science (2011).
43. Computational neuroscience. *Wikipedia* [online] http://en.wikipedia.org/wiki/Computational_neuroscience (2011).
44. CBRAIN—Credits [online], https://portal.cbrain.mcgill.ca (2008).

Acknowledgments

The authors thank all the partners in FP7 outGRID, FP7 neuGRID, LONI–ADNI and CBRAIN for collecting and providing information on main National and International virtual imaging laboratories. Special thanks go to Richard McClatchey, University of the West of England, Bristol, UK. G. B. Frisoni and D. Manset are supported by FP7 neuGRID and FP7 outGRID funded by the European Commission (FP7/2007-2013) under grant agreement no. 211,714 and no. 246,690, DG INFSO, e-Infrastructures. A. Toga was supported by NIH LONI–ADNI projects. A. Evans was supported by CANARIE Inc. and CBRAIN/GBRAIN projects.

Author contributions

G. B. Frisoni developed the architecture of the manuscript. G. B. Frisoni and A. Redolfi drafted a first version, which was completed, edited, and reviewed for important intellectual content by D. Manset, M.-É. Rousseau, A. Toga and A. C. Evans.

Supplementary information

Supplementary information is linked to the online version of the paper at www.nature.com/nrneurol