

NeuroGrid: Collaborative Neuroscience via Grid Computing

John Geddes¹, Sharon Lloyd², Andrew Simpson², Martin Rossor³, Nick Fox³, Derek Hill⁴, Joseph Hajnal⁵, Stephen Lawrie⁶, Andrew McIntosh⁶, Eve Johnstone⁶, Joanna Wardlaw⁶, Dave Perry⁶, Rob Procter⁷, Philip Bath⁸, and Ed Bullmore⁹

¹Department of Psychiatry, University of Oxford, ²Oxford University Computing Laboratory, ³Institute of Neurology, University College London, ⁴Centre for Medical Image Computing, University College London, ⁵Imaging Sciences Department, Imperial College London, ⁶Department of Psychiatry, Edinburgh University, ⁷School of Informatics, University of Edinburgh, ⁸Division of Stroke Medicine, University of Nottingham, ⁹Addenbrookes Hospital, Cambridge

Abstract

Advances in neuroimaging have already led to breakthroughs in the clinical management of neurological disorders, with current developments holding comparable promise for neuro-psychiatric disorders. There are, however, key problems to be overcome before the large-scale clinical studies that will realise fully the potential benefits of modern imaging techniques in the diagnosis and treatment of brain disorders can be conducted. We provide an overview of the NeuroGrid project, which aims to tackle some of these key problems.

1. Introduction

Current neuroimaging research is characterised by small studies carried out in single centres. While neuroimaging research groups regularly share algorithms for image analysis (for example, many groups make their algorithms available for download over the Web), the widespread sharing of data tends to be the exception rather than the rule. In those situations where data is shared, subtle differences between centres pertaining to the method of image acquisition normally inhibit reliable quantitative analysis of aggregated data. Furthermore, data curation in neuroimaging research tends to be poor, making aggregation of data between or within sites difficult, if not impossible [1]. Imaging techniques are increasingly being used to detect features that can refine a diagnosis, to phenotype subjects, to track normal or often subtle pathophysiological changes over time, and to improve our understanding of the structural correlates of the clinical features. The identification of true disease-related effects is obviously crucial, with, unfortunately, problems being caused by confounding and artefactual changes in the complex procedures involved in image acquisition, transfer and storage.

There are two basic approaches to the extraction of detailed information from imaging data: quantitative assessment and qualitative assessment, with both posing key challenges. We consider each in turn.

In quantitative assessment, sophisticated, largely automated and computationally intensive image analysis algorithms have recently been developed that offer great promise in terms of quantification and localization of signal differences. These methods have important applications in longitudinal imaging studies designed to identify change within individuals (e.g., in cohorts at risk of dementia or schizophrenia) or in clinical trials with imaging outcome measures (e.g., of treatments for Alzheimer's disease). Current practice relies on these algorithms being locally implemented, which can lead to a lack of standardization. Further, changes in staff may mean that the software becomes unmaintainable and outdated. This, of course, leads to a lack of scalability and an inability to compare methods or cross validate results from different groups.

In qualitative assessment, many large randomised controlled trials or observational studies use imaging to phenotype patients (e.g., haemorrhagic versus ischaemic stroke), to assess severity, and may use follow-up scans to

assess disease progression or response to treatment. Such studies may enrol thousands of patients from hundreds of centres using a large variety of imaging equipment, reflecting the reality of imaging provision in health services. A reliable system is required for managing the scans (collection, storage and dissemination), the results from raters and the study metadata.

2. NeuroGrid project objectives

The principal objective of the NeuroGrid consortium is to enhance collaboration within and between clinical researchers in different domains, and between clinical researchers and e-scientists. Sharing data, experience and expertise will facilitate the archiving, curation, retrieval and analysis of imaging data from multiple sites and enable large-scale clinical studies. To achieve this, NeuroGrid will build upon technologies and tools developed within the UK e-Science Programme to integrate image acquisition, storage and analysis, and to support collaborative working within and between neuro-imaging centres.

The project has three main strands of activity.

First, NeuroGrid will create a Grid-based infrastructure to connect neuroimaging centres, thereby providing rapid and reliable flow of data, and facilitating easy but secure data sharing through interoperable databases and sophisticated access control management and authentication mechanisms. This aspect of the project will leverage experience gained within the eDiaMoND project¹.

Second, NeuroGrid will develop distributed data analysis tools and services, including a neuroimaging toolkit for image analysis, image normalization, anonymisation and real-time acquisition error trapping. The intention is that these tools will: improve diagnostic performance; enable differences between images from different scanners (either in time or place) to be compensated for; and allow quality and consistency verification before the patient leaves the imaging suite, thereby permitting rescanning if required. An advantage of a Grid-based approach to imaging studies is that, by aggregation of large amounts of image data, it is possible to *learn* scanner variability from the data itself. By non-rigidly registering all the

scans together, scans can be warped to make them look virtually identical² [2, 3].

Finally, NeuroGrid will deploy the tools and techniques it creates in support of three clinical exemplar projects to explore real world problems. The clinical exemplars are associated with stroke, psychosis, and dementia. In particular, NeuroGrid will use the clinical exemplars both to derive detailed requirements and to validate them. Collectively, the exemplars will enable the project team to address several generic issues, including data curation and management, data access and security and desktop tools for image presentation, manipulation and annotation.

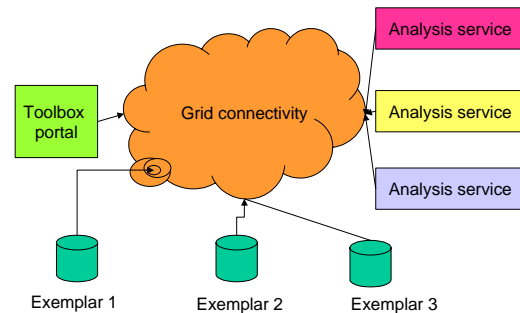


Fig 1. NeuroGrid Workpackages

3. Challenges

There are, of course, a number of challenges associated with managing imaging data that will have to overcome within this project. We consider some of these challenges below.

Image data quality and consistency. Problems in this respect include simple labelling errors, incorrect acquisition parameters, patient movement, incorrect positioning or artefact, and wrong scanner or scanner fault. All of these issues can impair or even negate information from images. Some faults are detectable and correctable by operators during scanning, while others are altogether more subtle. NeuroGrid will devise a method of immediate quality and consistency verification before the patient leaves the imaging suite to enable rescanning if required. The validity of the solution developed

¹ <http://www.ediamond.ox.ac.uk/>

² <http://www.itk.org/>

within the project will be assessed within the dementia exemplar.

Multiple and ever changing technologies. Image data differences arise from the use of different scanners from different manufacturers at different sites and potentially, in the case of MRI, different field strengths in cross-sectional or longitudinal studies. Large studies require multiple sites, and scanner changes are almost unavoidable over the time of longitudinal studies. The variance found in imaging data increases, perhaps by a factor of 2 or 3 when data from multiple scanners are combined. While trained observers can adjust for differences arising from different scanners to interpret images, computer-based analysis systems cannot readily do this. Within the project, the differences between images from different scanners (either in time or place) will be modelled, and compensated for via Grid-based normalization. The validity of the solution developed within the project will be assessed within the psychosis and dementia exemplars.

Secure long-term data storage of large datasets. It is understood that data curation is almost invariably performed at a less than optimum level in current neuroimaging research: typically, the data is stored on removable media that frequently become unreadable after a few years. Vastly improved curation is needed to allow future re-analysis as new techniques become available, or meta-analysis with imaging data from other trials of similar treatments or of observational studies of similar diseases. Imaging studies are expensive and difficult to do and rarely achieve an adequate sample size to perform subgroup analyses – the ability to combine datasets from different scanners and store data for subsequent pooling into much larger combined datasets will not only lead to better science, but is an inevitable consequence of the wider use of imaging in clinical neuroscience. This issue will be addressed by building on the Grid services developed within the IXI and e-DiaMoND projects. The validity of the solutions developed within the project will be assessed within all three exemplars.

Effective use and integration of data. The cost overhead of setting up database structures and the tools for manipulating and mining neuroimaging data are presently only afforded by larger scale studies. Provision of Grid-enabled databases and meta analysis tools could

make the power of these methods much more widely available for imaging studies. The project team will build on solutions developed within the e-DiaMoND project to facilitate incorporation of these capabilities in NeuroGrid. The validity of the solutions developed will be assessed within the stroke exemplar.

Efficient observer rating for large imaging studies. Although this is difficult to achieve, it is required in multi-centre trials and observational studies to improve consistency of diagnosis; define observer reliability; improve training; development of automated image interpretation algorithms; provide complementary image analysis tools; and facilitate rapid evaluation of emerging diseases. The validity of the solution developed will be assessed within the stroke exemplar.

Achieving much larger sample sizes in imaging studies. Large studies are necessary to identify modest but clinically important associations and treatment effects. The management of imaging data, which may be crucial to such trials, is difficult. Very large data stores, appropriate meta-data, and rapid accessibility for observer image rating or computational image analysis are essential to future trial conduct. The Grid infrastructure developed in the project will address these issues. The validity of the solution developed will be assessed within all three exemplars.

Workflow and data provenance. As data processing becomes more complex, keeping track of what has been done with which algorithm and which software version becomes both more difficult and more critical if errors are to be avoided and results to be replicated by others. Building on the IXI project we will make provenance tools available so that processing history can be tracked and processing pipelines can be modified at low cost in time and wasted calculations. The suitability of the tools developed will be assessed within the dementia and psychosis exemplars.

Translating subtle research findings into clinical practice. The initial benefits of a grid approach in neuroimaging will be in research studies. However, we expect that these research studies will demonstrate that a Grid infrastructure underpinning image acquisition, analysis and distributed working has substantial potential in healthcare delivery as well. For example, such an infrastructure could provide hospitals throughout the country with the

capability for early detection of Alzheimer's dementia and accurate early diagnosis of stroke. A grid infrastructure, with appropriate image analysis, curation and handling tools, is likely to be the best way to achieve standardised and equitable diagnosis for the majority of patients. NeuroGrid is an initial and important step towards this goal, and during the project we will assess the obstacles to implementing a clinical version of NeuroGrid.

4. Anticipated results

We expect the solutions developed by the NeuroGrid consortium to deliver a range of benefits for the clinical neurosciences research community. Collectively, these have the potential to streamline data acquisition, aid data analysis and improve the power and applicability of studies. By enabling the rapid and secure flow of data, distributed data analysis with image analysis tools and data sharing through interoperable databases, NeuroGrid will improve the collaborative capability of neuroscience research groups and encourage the re-use of data.

By enabling current algorithms and existing data management procedures to be more accessible and interoperable, NeuroGrid will lower barriers to entry, reduce time wasted re-engineering well established algorithms and avoid established research groups having to adopt alien practices. The clinical exemplars will focus on outcomes within separate, but complementary areas.

The dementia exemplar requires real-time transfer and processing of images to return information to the scanner before the examination is over so that data quality can be assured and, if necessary, there can be an intervention to deal with problems. It will use both secure data transfer and Grid services. From the augmented studies, there will be scientific outcomes to demonstrate added value resulting from the Grid; specifically, from the direct benefits of the Grid imaging toolkit to reduce variance resulting from multiple sites and also from cooperative working and aggregation of data.

The stroke exemplar will establish and test mechanisms for interpretation and curation of image data, which are essential to the infrastructure of many multi-centre trials in

common brain disorders. This includes testing issues of security, confidentiality, accessibility, linking to metadata, 'hooks' to computational image analysis methods, resolving issues of scale, and ensuring commonality of key data requirements for future meta-analyses.

The psychosis exemplar will test the capabilities of NeuroGrid to deal with retrospective data, assimilate material into databases and use of the toolkit for normalisation and analysis.

Although the primary aim is not to produce new research data, NeuroGrid will produce valuable new scientific findings from the application of cutting edge analysis techniques to combined existing datasets. It is anticipated that NeuroGrid will foster sharing of resources, expertise and data and so speed scientific advance through knowledge transfer across all areas of clinical neuroscience.

References

1. The Governing Council of the Organization for Human Brain Mapping (OHBM), Neuroimaging Databases. <http://www.csbmb.princeton.edu/ncc/publications/2001/OHBM2001.pdf>
2. Hill DLG, Burns M, Pennec X, Parkin M, Hajnal J, Stefanescu R, Rueckert D, Montagnant J. Interoperable medical image registration grid service. HealthGrid presentation, p36 January 2004
3. Burns M, Rueckert D. An OGSA based environment for Image Analysis: Grid enabling the Insight Segmentation and Registration Toolkit (ITK). GlobusWorld 2004, poster 18