NeuroGrid: Using Grid technology to advance neuroscience

Citation for published version (APA):

Published in:

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Takedown policy

NeuroGrid:

Using Grid technology to advance Neuroscience

John Geddes [1], Sharon Lloyd [2], Andrew Simpson [2],Martin Rossor [3], Nick Fox [3], Derek Hill[4]; Joseph V Hajnal [5], Stephen Lawrie [6] , Andrew McIntosh [6], Eve Johnstone[6], Joanna Wardlaw [6], Dave Perry [6], Rob Procter [7], Philip Bath [8,] Ed Bullmore [9]

[1] University of Oxford, Department of Psychiatry, [2] University of Oxford, Computing Laboratory [,3] University College London, Institute of Neurology, [4] University College London, Centre for Medical Image Computing (MedIC), [5] Imaging Sciences Department, Imperial College London, [6] Edinburgh University Department of Psychiatry, [7]School of Informatics, University of Edinburgh, [8]The University of Nottingham, [9]Addenbrookes Hospital, Cambridge.

John.Geddes@pysch.ox.ac.uk, {Sharon.Lloyd, Andrew.Simpson}@comlab.ox.ac.uk, {M.rossor, Nfox}@dementia.ion.ucl.ac.uk, Derek.hill@ucl.ac.uk,   jo.hajnal @ imperial.ac.uk, {S.lawrie, Andrew.mcintosh, E.Johnstone, Joanna.wardlaw, David.perry}@ed.ac.uk , rnp@inf.ed.ac.uk, philip.bath@nottingham.ac.uk, etb23@cam.ac.uk

Abstract

Large –scale clinical studies in neuro-imaging are hampered by several factors including variances in acquisition techniques, quality assurance and access to remote datasets. The Neurogrid project will build on the experience of other UK e-science projects to assemble a grid infrastructure, and apply this to three exemplar areas: stroke, dementia and psychosis, to conduct collaborative neuroscience research.

1. Introduction.

Advances in neuroimaging have already led to breakthroughs in the clinical management of neurological disorders and current developments hold 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. This paper provides an overview of the NeuroGrid project, which aims to tackle some of these problems.

 NeuroGrid is a three-year project, funded by the UK's Medical Research Council. NeuroGrid will bring together experts in both the development of Grid technology solutions through experience of first round UK e-Science programme projects such as e-DiaMoND and IXI, and neuroscientists from Oxford, Edinburgh, Cambridge, Nottingham, London, and Newcastle.

The principal aim of NeuroGrid is to develop a Grid-based collaborative research environment for neuroscientists. This research environment will be developed via the investigation of three exemplars in the clinical fields of stroke, dementia, and psychosis. A Grid-based infrastructure and sophisticated services for quantitative and qualitative image analysis will be developed to support these exemplar applications.We start by considering generic neuroscience challenges that make the development of an appropriate grid infrastructure desirable, before discussing the objectives of the project. We conclude by providing an overview of the anticipated results.

2. Challenges in Neuroscience.

Current neuroimaging research is characterised by small studies carried out in single centres. While neuroimaging research groups regularly share algorithms for image analysis (many groups make their algorithms available for download over the Web) it is very unusual for there to be widespread sharing of data. When data is shared, subtle differences between centres in the way that the images acquired 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. Imaging techniques are increasingly used to detect features that can refine a diagnosis, phenotype subjects, track normal or often subtle pathophysiological changes over time and/or

improve our understanding of the structural correlates of the clinical features. The identification of true disease-related effects is obviously crucial and problems are 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 – both of which pose key challenges. We introduce 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 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 enroll 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. There are, of course, other challenges associated with managing imaging data for clinical trials.

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. Even when using exactly the same scanner at two different locations, variations in the way the scanner is used, and the way the patients are positioned, can have important implications for subsequent analysis. Large studies require the cooperation of multiple sites. In addition, scanner changes are almost unavoidable over the time of longitudinal studies.

Secure long term data storage of large datasets. Data curation is almost invariably performed poorly in current neuro-imaging research, with the data stored on removable media that frequently become unreadable after a few years. Much 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.

Effective use and integration of data. The cost overhead of setting up database structures and the tools for manipulating and mining the 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.

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.

Image data quality and consistency. Problems range from simple labelling errors, incorrect acquisition parameters, patient movement, incorrect positioning or artefact, to wrong scanner or scanner fault. All of these impair or even negate information from images. Some faults are detectable and correctable by operators during scanning, others are more subtle. Algorithms running on a computer that has rapid access to the image data after it is collected could detect many of these problems, and alert the radiographers before the subject leaves the department, enabling a repeat scan if required.

The NeuroGrid project is motivated, in part, by a desire to address all of the above problems.

3. 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 Grid 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.

There are three main elements to NeuroGrid. First, NeuroGrid will create a Grid-based infrastructure to connect neuro-imaging 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 the experiences of the eDiaMoND project team through its development of an architecture to allow the federation of mammography X-ray archives.

Second, NeuroGrid will develop distributed, Grid-based data analysis tools and services, including a neuroimaging toolkit for image analysis, image normalization, anonymisation and real-time acquisition error trapping. These tools will aim to improve diagnostic performance, to enable differences between images from different scanners (either in time or place) to be compensated for and to 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 identical.

Finally, NeuroGrid will deploy the tools and techniques it creates in three clinical exemplar projects in stroke, psychosis, and dementia to explore real world problems and solutions. In particular, NeuroGrid will use the clinical exemplars both to derive detailed requirements and to validate them. Collectively, the exemplars will address generic issues that are fundamental to the overall aims of NeuroGrid and to the field of e-Health more generally. It will, of course, be necessary to carefully manage and integrate the three exemplars to ensure that common requirements are identified. Ones already identified include: data curation and management, data access and security and desktop tools for image presentation, manipulation and annotation.

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. 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 (receiving images from very many sites, not just a few), 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. The technology developed in NeuroGrid will have the potential to serve NHS needs such as those identified in the National Service Frameworks for Mental Health and for Older People and Scottish Executive's CHD and Stroke Strategy. NeuroGrid will foster sharing of resources, expertise and data and so speed scientific advance through knowledge transfer across all areas of clinical neuroscience.

5. Conclusions

Neuroimaging clinical trials are a good testbed for e-science technologies, as they involve distributed data collection, assessment and analysis, and need good data curation and access to high throughput computing infrastructure. The Neurogrid project will build on the experience of other UK e-science projects to assemble a grid infrastructure, and apply this to three exemplar areas: stroke, dementia and psychosis.