Distributed Computing for Neurosciences: the N4U example

Niobe Haitas, Genevieve Romier, Tristan Glatard

Distributed Computing for Neurosciences: the N4U Example

Niobe Haitas (1), Geneviève Romier (2), Baptiste Grenier (3), Giovanni Frisoni (4), David Manset (5), Richard McClatchey (6), Kamran Munir (7), Jérôme Revillard (8), Keith Cover (9), Gabriella Spulber (10), Tristan Glatard (11), Jean- François Mangin (12), Arthur Toga (13), Alan Evans (14)

(1) *niobe.haitas@idgrilles.fr, Institut des Grilles et du Cloud, IDGC*
(2) *genevieve.romier@idgrilles.fr, Institut des Grilles et du Cloud, IDGC*
(3) *bgrenier@maatg.com, maat France, maatG*
(4) *gfrisoni@fatebenefratelli.it, Provincia Lombardo Veneta Ordine Ospedaliero di San Giovanni di Dio- Fatebenefratelli, FBF*
(5) *dmanset@maatg.com, maat France, maatG*
(6) *Richard.Mcclatchey@uwe.ac.uk, University of the West of England, UWE*
(7) *Kamran.Munir@cern.ch, University of the West of England, UWE*
(8) *jrevillard@maatg.com, maat France, maatG*
(9) *Keith@kscover.ca, VU University Medical Center Amsterdam, VUmc*
(10) *Gabriela.Spulber@ki.se, Karolinska Institutet, KI*
(11) *tristan.glatard@creatis.insa-lyon.fr, Université de Lyon, CREATIS, CNRS, Inserm, INSA-Lyon, Université Lyon 1*
(12) *jfmangin@gmail.com, Commissariat à l'Energie Atomique et aux Energies Alternatives, CEA\CATI*
(13) *toga@loni.ucla.edu, UCLA Laboratory of Neuroimaging*
(14) *alan@bic.mni.mcgill.ca, McGill, CBRAIN*

Overview

Imaging data of people

Ser

S63 with brain diseases are becoming increasingly available and accessible, leading to the development of computational infrastructures to facilitate scientists' access to image databases and e-Science services. The N4U (neuGRID 4 you) project offers a platform for neuroscientists where they can find core services and resources for brain image analysis, and access datasets, processing pipelines, computational resources, services and support. The N4U platform, available at http://neugrid4you.eu, is based on the previously developed neuGRID technology[1] for imaging neuroscientists of Alzheimer's disease (AD) but designed to be adaptable to other user communities. N4U offers tools and associated support and training to process imaging datasets on grid resources. This abstract summarizes some issues related to the processing of neuroimaging data, presents the N4U infrastructure, and the resulting tools, applications and datasets that it offers.

Scientific Issues, Computing, Storage and Visualization Needs

Neurodegenerative diseases (NDD) such as Alzheimer's, grey matter, white matter and psychiatric diseases affect the brain and are responsible for a large share of disability in the population. For their early diagnosis, preventive and treatment therapies, effective disease-modifying drugs require accurate disease markers. There are no tracking markers to diagnose the disease but diagnosis is traditionally made on clinical grounds and disease progression is monitored through cognitive testing. This lack of objective diagnosis means that the disease cannot be diagnosed at the mildly symptomatic stage and there is huge heterogeneity of diagnostic accuracy within EU countries. Furthermore, clinical trials of drugs to delay the progression of the disease require hundreds of patients followed for many months, resulting in non-optimised drug discovery of efficient drugs. The lack of open data in brain MRI research makes the effort of addressing AD even more challenging (Barkhof 2012).

Imaging can detect the changes that take place in the brain of patients with Alzheimer's, multiple sclerosis and psychiatric diseases at the molecular, cellular, tissue and organ levels and therefore, addressing the need for accurate disease markers. Recent advances in neuroimaging demonstrate subtle cognitive alterations that are detectable years prior to the development of objective memory deficit meaning that the disease can be diagnosed much earlier. Modern imaging techniques are essential not only for diagnosing but also for processing and sharing data all over the world. Imaging techniques are computationally intensive, and as such, distributed computing (e.g., grids) are key to addressing the emerging need for user-friendly access to data, analysis pipelines and computational resources as imaging data of people with brain diseases are becoming increasingly available and accessible.

These needs have led to the development of computational infrastructures in Europe and North America offering scientists access to the image databases themselves and e-Science services, including sophisticated image analysis algorithm pipelines, access to powerful computational resources, visualization and statistical tools. As such, imaging data management has been successfully facilitated by projects such as LONI[2], CBRAIN[3] and DECIDE[4]. N4U is meant to be the European counterpart of these platforms.

---

[1] http://neugrid.eu/pagine/home.php

2 http://www.loni.ucla.edu

Infrastructure Development and Use

N4U was developed to provide a facilitated e-Science environment, a user-friendly Virtual Laboratory, where the broader neuroscience community can find a large range of scientific resources, services, datasets, algorithm applications, intuitive computational resources and support (Frisoni et al. 2011). The N4U platform addresses the needs of neuroscientists who wish to experiment and interact with the most popular pipelines for brain image analysis, benchmark their image analysis pipeline, or share large datasets with colleagues.

The N4U architecture is based on the neuGRID one, offering direct or facilitated access to large brain image datasets, popular image analysis pipelines, flexible computational resources, users' training and support. N4U follows the Service Oriented Architecture (SOA) paradigm.

As shown on Figure 1, the N4U Science Gateway consists of two components:

- a backend, composed by all the Web Services being the ground of the platform with the addition of external applications;
- a frontend, offering a user friendly web interface allowing user communities to easily interact with the platform.

The backend and the fronted are decoupled; it is possible to have multiple frontends accessing the same services with heterogeneous and/or specialized interfaces.

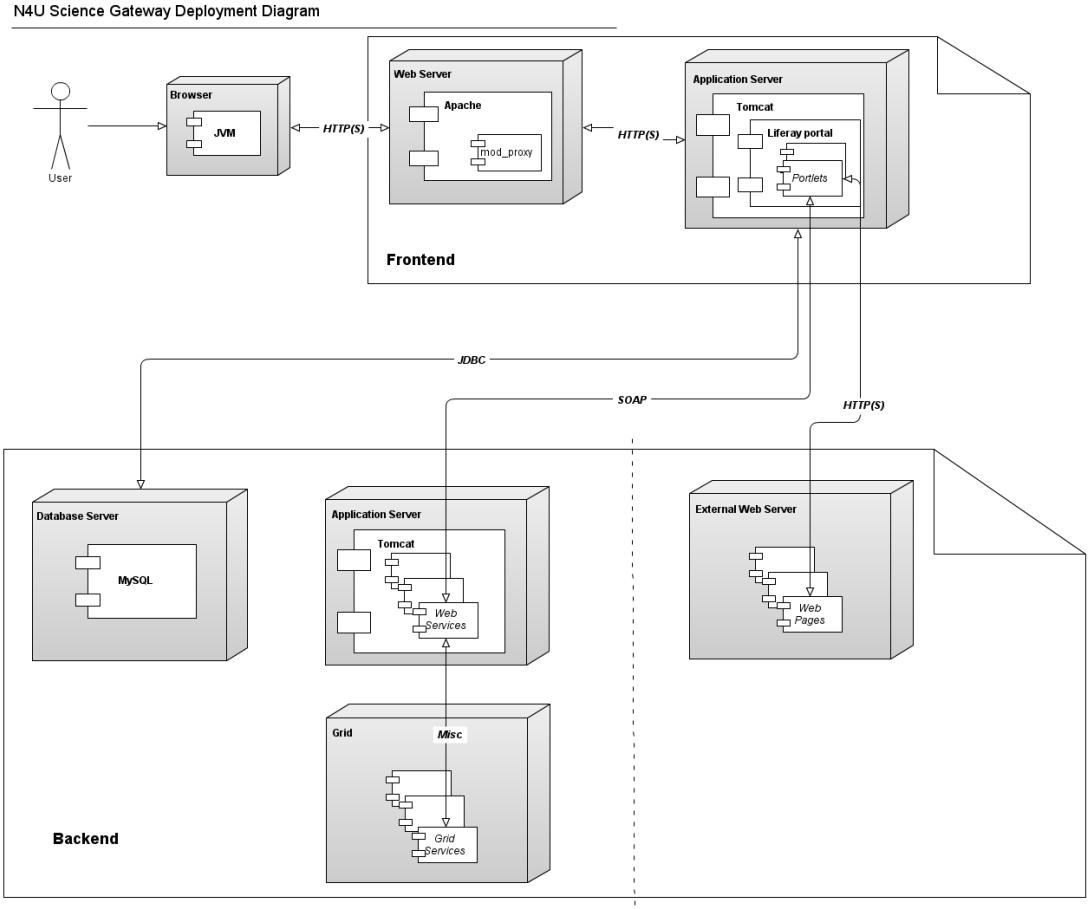

Figure 1: N4U platform overview

Results and Discussion

3 http://cbrain.mcgill.ca
4 http://www.eu-decide.eu

Although the N4U Science Gateway is not yet finalized, a number of tools, datasets and applications are already available and accessible through the N4U platform. They are briefly described below:

The Desktop Fusion offers users access to different tools and facilities (LONI client, an XTerm, a graphical text editor, a file browser). Users only need a Java virtual machine and don't need to install anything on their computer. It also allows users to upload a shared directory of their local computer inside a folder of the Pandora Gateway, and easily upload content on the grid. Figure 2 shows a screenshot of the main tools available through desktop fusion.

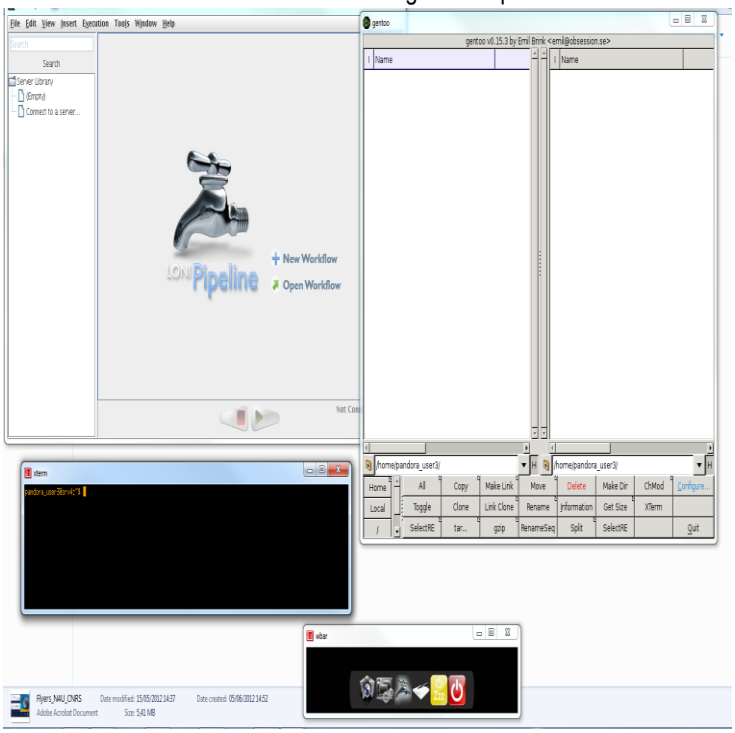

Figure 2: screenshot of the tools available in Desktop Fusion

The GridBrowser allows users to browse the Logical File Catalog (LFC) and to download or upload files from their computer to the grid. This portlet does not require any special plugin but is a very user-friendly way of interacting with the storage capabilities of the Grid. Users can easily manage their data like with a traditional file manager.

BrainBrowser[5] is an externally hosted web-based, 3D visualization tool for neuroimaging. It allows for real time manipulation and analysis of 3D neuroimaging data whether these are pre-calculated maps, or models and data provided by the user in a supported format (such as Minc, Nifti, object files, plain text). Figure 3 shows a screenshot of the BrainBrowser used to visualise cortical thickness from the N4U platform.

[5] https://brainbrowser.cbrain.mcgill.ca

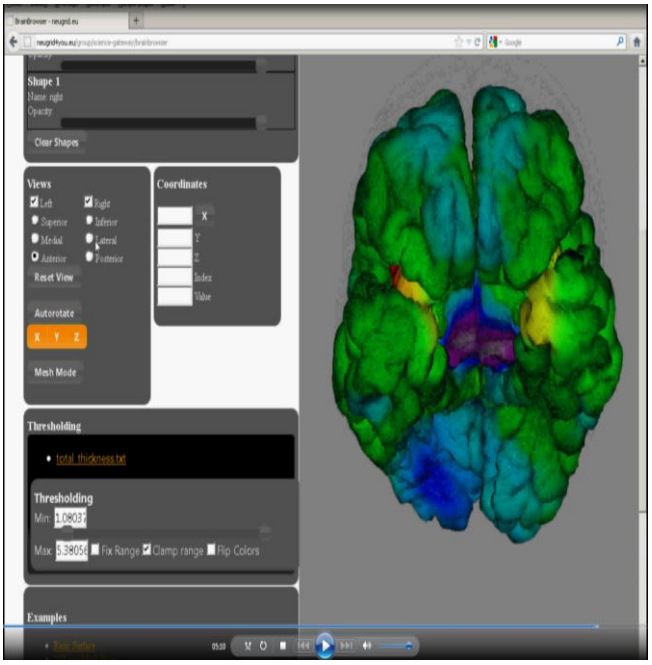

Figure 3: screenshot of the BrainBrowser, started from the N4U gateway

The LONI pipeline is a graphical workflow management tool where users can find predefined modules to build their own processing pipelines from modules situated in the server library.

Expresslane will also soon be available to help users analyse their resources. ExpressLane is a command line submission framework allowing users to easily submit tasks to a computing facility for grids, such as running a given script on multiple scans or datasets and submitting jobs to the DCI only by generating a list of datasets - without knowledge of the underlying job submission system.

Two datasets are available, from the European Union AddNeuroMed program and the US-based Alzheimer Disease Neuroimaging Initiative (ADNI). These initiatives provide more than 10,000 multi-centric MRIs of patients with AD. In Cover et al. (2011) we can find an example of the use of the ADNI dataset.

Concerning applications, FSL[6] is available through the LONI pipeline to compute partial volume maps such as CSF (celebral spinal fluid), Grey Matter and White Matter from MRI scans. On Figure 4 we can see an example of a result obtained from an FSL pipeline executed with LONI within N4U. Freesurfer[7] is also being integrated.

[6] http://www.fmrib.ox.ac.uk/fsl/fsl/whatsnew.html

[7] http://surfer.nmr.mgh.harvard.edu/

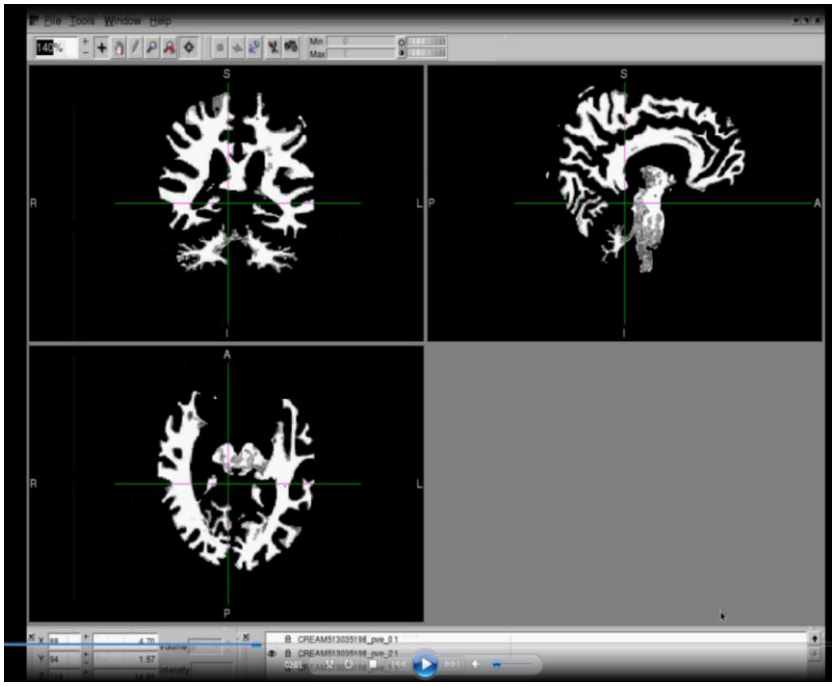

Figure 4: result of an FSL pipeline executed within N4U

Perspectives

Given the data deluge in imaging for neurosciences, users need their data to be more accessible, understandable, usable and shareable. Sharing and interoperability need to be developed not only for data but also for anything that connects to the data production and processing including computing tools, applications, methods, software, metadata, workflows across different platforms and even communication (Haitas 2012). In the coming months, the following new applications will be deployed to ease the usage of the N4U platform and provide more tools to the users' communities:

- Provenance Service building the N4U "data atlas" data model (Anjum 2011)
- Persistency Service to populate the Information Services database by interfacing with datasets and data sources integrated into N4U as well with the algorithms, pipelines and toolkits integrated
- Analysis Services providing a customisable environment for users to conduct their neuroscience analyses using the Provenance, Persistency and DCI Pipeline service
- The Analysis Services Graphical User Interfaces (GUI) allowing users to query data, specify and execute pipelines and visualize data by interacting with the Analysis services
- The Virtual Imaging Platform (VIP[8]) is an openly-accessible web platform for multi-modality image simulation. The user-friendly web interface will be integrated with the N4U infrastructure, allowing users to easily launch predetermined workflows.

References

Anjum A. et al. Provenance Management for Neuroimaging Workflows in neuGrid. International Conference on P2P, Parallel, Grid, Cloud and Internet Computing. IEEE Computer Society, CS Digital Library. DOI Bookmark: http://doi.ieeecomputersociety.org/10.1109/3PGCIC.2011.20 (2011)

Barkhof F. Making better use of our brain MRI research data. European Radiology 22:7, 1395-1396. DOI: 10.1007/s00330-012-2408-3 (2012)

Cover K.S. et al. Assessing the reproducibility of the SienaX and Siena brain atrophy measures using the ADNI back-to-back MP-RAGE MRI scans. Psychiatry Research: Neuroimaging. 193:3, 182-190, http://dx.doi.org/10.1016/j.pscychresns.2011.02.012 (2011)

Frisoni G.B. et al. Virtual Imaging Laboratories for marker discovery in neurodegenerative diseases. Advance online publication Nature Reviews Neurology. Doi 10.1038/nrneurol (2011)

[8] http://vip.creatis.insa-lyon.fr

Haitas N. How e-Science can help to solve pressing societal challenges: fostering a global effort to develop a worldwide e-infrastructure for computational neuroscientists to fight Alzheimer's disease: joint conclusions. Workshop report: https://neugrid4you.eu/c/document_library/get_file?uuid=27266c34-adf9-4ef2-b85e-9cd8051fcc67&groupId=10626