NeuroBase : Management of Distributed and Heterogeneous Information Sources in Neuroimaging

C. Barillot*[1], R. Valabregue[1], J-P. Matsumoto[6], F. Aubry[2], H. Benali[3], Y. Cointepas[3], O. Dameron[4], M. Dojat[5], E. Duchesnay[3], B Gibaud[1,4], S. Kinkingnéhun[3], D. Papadopoulos[3], M. Pélégrini-Issac[3], E. Simon[6]

[1]IRISA, Equipe VisAGeS, CNRS-INRIA, Rennes, France
[2] INSERM U455, Toulouse, France
[3]IFR 49 / CHR La Pitié-Salpêtrière / CEA-SHFJ, Paris / Orsay, France
[4]Laboratoire IDM, Univ. Rennes I, Rennes, France
[5]Unité INSERM U594 / TIMC, Equipe SIC, CNRS-IMAG, Grenoble, France
[6]Projet SMIS, INRIA, Rocquencourt, France

Abstract: The Neurobase project has for objective to establish the conditions allowing the federation through Internet of information sources in neuroimaging, where sources are distributed in different experimental sites, hospitals or research centers in cognitive neurosciences and contain data and image processing methods. This project consists in the definition of a computer architecture allowing the access and the sharing of such heterogeneous and autonomous sources in neuroimaging. This will enable for example, within these information sources, the search for similar data, the search for images containing singularities, the construction of a specific chain of distributed image processing tools or the transverse search of data for excavating possible regularities (similarly to a data mining approach).

Keywords: *web-based processing, data warehouse, mediator/wrappers, 3D medical imaging, interoperability, semantic Web, image indexation*

---

* Principal Investigator and contact author: Christian.Barillot@irisa.fr

Introduction

Neurobase (www.irisa.fr/visages/neurobase) is a cooperative project, supported by a grant from the French ministry of Research, which aims to establish the conditions allowing, through Internet, the federation of distributed information sources in neuroimaging, these sources being located in various experimentation centers, clinical departments in neurology, or research centers in cognitive neurosciences.

One objective of neuroscientists is the construction of functional cerebral maps under normal and pathological conditions. Researches are currently performed to find correlations between anatomical structures, essentially sulci and gyri, where neuronal activations take place, and cerebral functions, as assessed by recordings obtained by the means of various neuroimaging modalities, such as PET (Positron Emission Tomography), fMRI (functional magnetic resonance imaging), EEG (electro-encephalography) and MEG (magneto-encephalography). Then, a central problem inherent to the formation of such maps is to put together recordings obtained from different modalities and from different subjects. This mapping can be greatly facilitate by the use of MR anatomical brain scans with high spatial resolution that allows a proper visualization of fine anatomical structures (sulci and gyri). Recent improvements in image processing techniques, such as segmentation, registration, delineation of the cortical ribbon, modeling of anatomical structures and multi-modality fusion, make possible this ambitious goal of neuroscientists. Nevertheless sophisticated techniques are often located in specific laboratories and are not yet available to the entire community.

Two of the major concerns of researchers and clinicians involved in neuroimaging experiments are on one hand, to manage internally the huge quantity of produced data ($\approx$ 1 Gb per subject) and, on the other hand, to be able to confront their data and the methods they develop, with those existing in other centers or moreover with those described in publications. Furthermore, and this is more particularly true for medium size centers (with limited staff capabilities), or even small ones (it is mostly the case in clinical centers), the researchers or the clinicians face great difficulties to set up large-scale experiments which are self-important, mostly because of lack of human resources and capacities of recruiting subjects. Besides, the statistical validity of the results is sometimes insufficient (the rate of "false negative" is probably not negligible). Finally, another concern is to better insure the profitability of the existing and expensive equipments (for the purchase as well as for the use), by facilitating their access to a wider users community. For all these reasons, we believe that the pooling of neuroimaging data and the sharing of data processing methods, through a network between collaborative centers, will allow widening the scientific achievement of the conducted experimental studies. This will also allow increasing the possible panel of people involved in neuroimaging studies, while protecting the excellence of the supplied work.

We can therefore envisage that the generalization of this approach will be one of the keys of the future discoveries in the field of clinical and cognitive neuroimaging, by the means of the existing new communication technologies and computation paradigms (e.g. high speed networks, computers grid …). It requires that the users can diffuse, exchange or reach neuroimaging information with appropriate access means, in order to be able to retrieve information almost as easily as if it were stored locally.

In this context, the Neurobase project consists in specifying how to connect and access distributed, heterogeneous and autonomous information sources in neuroimaging by defining a data-processing architecture allowing the access and the sharing of experimentation results and data processing methods within a same site or between different sites. Such a system would allow for example within these information sources:

- The search of experimentation conducted according to a specific protocol (for example allow to retrieve fine description of these experiments, to be able to examine these results or even to retrieve the related images),

- The search for similar results (for example for the study of anatomo-functional networks),

- The search for images containing singularities (e.g. spatio-temporal singularities),

- The search for image processing and data analysis methods, and the customization of specific chains of image processing tools doing a specific treatment (segmentation, registration, activation detection, etc ...),

- The transverse searches to highlight possible regularities, similarly to a "data mining" type approach (for instance to excavate possible similarities of the corpus of the protocols, of images, of results or of spatio-temporal invariants).

Background

Recent neuroimaging techniques, especially PET and then, non invasive fMRI, have literally provoked an explosion of researches in neurosciences. Early in the 90's has appeared the imperative necessity for innovative techniques for data and knowledge sharing and reuse [1, 2]. This led to the starting of the North American ambitious "Human Brain Mapping" project. An objective recently added to this project is the development of data analysis and data processing softwares to operate on various data repository system for data mining and knowledge discovery purposes. In parallel the development of web applications has stimulated the interest of researchers for distributed data bases and information sharing.

Four research topics are particularly relevant for our project:

1. Digital atlases and probabilistic models of brain anatomy
   Among the various research efforts for digital atlases construction, we can cite the work of Graf von Keyserlingk et al. on the labeling of post-mortem brain to quantify individual anatomical variability of cortical regions [3], of Toga et al. with the construction of an anatomical and functional rat atlas [4], of Bloom et al. with the " Brain Browser" to associate symbolic data and graphical data about the nervous system [5], and of Van Essen et al., on surface atlas of the primate visual system [6]. Some atlases are developed to support interpretation of functional data [7], image processing instantiation in a specific context [8] or training [9].
   For probabilistic atlases, significant steps have been performed after 1995 by the International Consortium for Brain Mapping. About 300 MRI brain scans plus post mortem data of 30 subjects were merged in a common referential. A detailed labeling of anatomical structures allowed the construction of a probabilistic atlas. Several image processing tools were added to allow segmentation and mapping of brain images to this brain reference [10].

2. Conception of image processing tools
   The BRAID[1] project at Hopkins University is relevant here. It explores the anatomy-function relationship based on activation-response experiments and deficit-lesion analysis. The proposed system integrates mechanisms for complex queries, combining selection with multiple criteria, images quantification, and statistical tests to calculate correlations between deficits and lesions. Group studies rely on matching all brains to a target (reference) by the means of linear or non-linear (deformable elastic model) matching methods, each with its own pro and con. Several participants of this project have a well-known experience on the conception of such robust image processing tools.

3. Multi-center data bases
   Several laboratories belonging to the Illinois University participate to the constitution of a commonly shared database devoted to neuronal patterns recordings[2]. This work, oriented to animal recordings, is close to our project. The database is used for instance, to find temporal series specific to neurons populations under various stimuli conditions. A common data model has been developed to organize the experimental data. An atlas is available to enter, search and analyze heterogeneous data in a common referential. Ontology sharing and data schemata updating facilities have been explored [11]. In the similar way, the BIRN[3] project is devoted for the sharing of heterogenous neurological information (from animal models to the human brain morphometry and function).

[1] BRAID ("Brain Image Database"): http://oasis.rad.upenn.edu/sbia/braid/
http://www.nimh.nih.gov/neuroinformatics/herkov.cfm
[2] http://www.nimh.nih.gov/neuroinformatics/mgabriel.cfm
[3] http://www.nbirn.net

A Useful Concept: The Mediation and the Wrapping of Information

Mediators were introduced to allow the virtual integration of heterogeneous distributed information sources in cooperative federated database systems. Mediators differ from standard database management systems in several aspects. Firstly, they do not supply mechanisms for simultaneous information sources updating. They only support queries to information sources in order to preserve their autonomy and the fact that they are locally managed. Secondly, to reinforce interoperability, mediators support various data models from standard structured data, such as relational, object or multi-dimensional models, to semi-structured models, such as XML. The architecture of mediators is also different, based on a "mediator/wrappers" concept [12], in which a mediator offers a central view about all information sources and the associated wrappers dedicated to each source, make this heterogeneity more transparent. A mediator redefines, using the corresponding wrappers, the user query into source dependent queries, then recomposes the various responses and formats the final response to the user. The query redefinition in sub-queries is optimized by the means of a cost based model to obtain the most efficient execution plan. This architecture clearly specifies the respective role of the mediator, which processes the user queries, and the wrappers, which translate the sub-queries into the relevant format for the associated information source. The pragmatic interest of such architecture is to lower the amount of work linked to the introduction of a new information source to the creation of the corresponding wrapper. Several mediators have been already developed (for instance DISCO [13], and Mocha [14]). Since 1998, one of the project participant has been developing a new generation of wrappers, called Le Select, which allows distributed, heterogeneous and autonomous data and methods (image processing methods, applications or web site) sharing via a high level query language. Le Select is the corner stone of the NeuroBase approach.

The NeuroBase Approach

One of the most important aspects in this project is to define the main concepts managed by the different information sources, data, image processing methods, and ontologies, in order to define a common semantic referential model every site can subscribe to (see Figure 1). From that base line, each neuroimaging collaborative partner site can map their own concepts, data, image processing methods and ontologies, to this semantic referential [11, 15]. For this purpose, we rely on a mediator/ wrapper approach [12], as defined above, where both the aspects of (i) modelling anatomical and functional experimental data and protocols, and (ii) description and access of image processing methods which can be applied to these data (segmentation, registration, statistical analysis …) can be expressed. These "descriptors" will supply the elements of indexation relative to the data susceptible to be retrieved by means of numerical information (for example anatomical and functional images, PET, fMRI or MEG functional activation maps, regions/volumes of interest, anatomical delineations, …) or by means of symbolic values (e.g. labels corresponding to anatomical entities, qualitative or quantitative information of region of interest or anatomical delineations like length, volumes, areas, expansion, shrink, etc…).

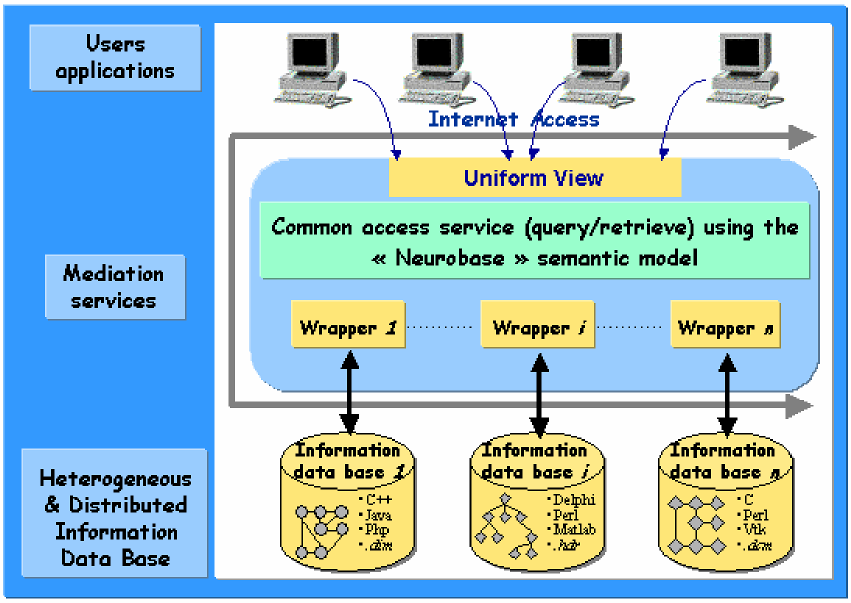

Figure 1: Example of the NeuroBase system architecture for managing distributed information sources in neuroimaging

From this semantic referential, contributors will have to subscribe to it and so forth declare their own views, which means what they will and what they will not share. Similar terms at different local sites could cohabit in the system as soon as the contributors define adequate wrappers to the common semantic referential (which it has to be defined without any ontology ambiguities).

This requires the definition of ontology services and the way people will access to these services with their own views. Le Select (http://www.le-select.com/) already offers some of these capabilities. In this context, different views of the semantic referential could be built according to the way the different subscribers of the Neurobase system will define their own wrappers.

This semantic referential has to be defined by a collaborative community, this requires quite of work since there is no fully defined common ontology yet, from which we can directly derive our semantic referential. We have to build it in a domain which is complex and not well defined. On the other hand, people from the Neurobase project have a strong experience in this type of problem (e.g. through the MIMOSA project [16]). Some realizations are also already available in the neuroimaging domain like the Neuronames terminology [17] or the fMRI data center ontology [18] which can serve as valuable starting points. At the end, the design of the Neurobase referential was driven by the pragmatic requirements of the application addressed by this project and well-mastered by the Neurobase participants (epilepsy, visual cortex exploration, Alzheimer disease …)[4].

Accessible information sources in a neuroimaging community cannot be known exhaustively at a given time (some of them can also not be created yet). Some of these information sources exist but have generally been set up for purely local needs and so are dedicated to only a very local user community. In this context, the Neurobase system could be used to manage the evolution or even the upcoming of new information sources. This evolution is possible thanks to the capability of this system to accept new contributions by just updating wrappers or creating new ones (this somewhat corresponds to changing or adding new views to the semantic referential).

The goal of the system is to allow the sharing of two types of information: on the one hand neuroimaging data, typically results from neuroimaging experiments, on the other hand computer aided methods, typically image processing programs or statistical tools, being applied to the data available in the distributed system. Data processes are modeled by the use of data flows. A data flow specifies the inputs and the types of parameters required for the completion of a given processing method, and the outcomes of this procedure. These descriptors referred to concepts in the semantic referential. Data sources can then be stored in relational databases or just in local files (wrappers will find their own way to the information).

With such a system, we will be able to make various queries using services available on this distributed system. A neuroimaging data oriented query could be `"find studies showing brain activities in the 'X' region defined in the 'Z' atlas"`. The completion of this query assumes that candidate brains must be registered together to the 'Z' template in order to localize the 'Y' region and to decide if there is actual activations in this 'Y' region.

A second typical query mixes retrieving data and image processing methods. Such a query could be `"find subjects where the retinotopic areas have been delineated"`. Then, a study is performed to assess the variability of functional borders and anatomical landmark (calarin sulcus) in the visual cortex for a given type of population. The Neurobase system will aim at building a specific corpus of information coming from different sites on which we could apply, through the data flow framework, specific segmentation or registration algorithms in order to study for instance local anatomical and functional visual areas. The same scheme can also be useful to build large studies for the prospect of validating segmentation or

[4] The definition of this shared ontology goes beyond the scope of this paper; this is described in more details in a second submission focused on this important aspect.

registration procedures with large enough databases (up to now, very few centers over the world can manage such projects).

Outcomes

In a short term, our goal is to elaborate a demonstrator based on some existing modules like Le Select[5], BrainVISA/Anatomist[6] or VIsTAL[7] extendable to modules largely used in neurosciences community such as SPM[8] software. Several functionalities are being implemented such as data wrappers & mediators, medical image processing methods (data access, registration, segmentation, visualization, …) through a data flow model approach (see Figure 2). This figure shows an example of the Neurobase demonstrator architecture with a generic server model ("*server 1i*") based on the LeSelect architecture with a PostGres/SQL[9] data base for managing storing of temporary image data computed from Data Flow process. The *server 2* is a specific extension of the generic server in order to ensure compatibility with the BrainVisa environment while keeping the same service from the Client. The data flow processes for image processing workflow procedures are executed from a Tomcat[10] server. This workflow can be designed from any client web browser. In order to build relevant neuroimaging queries, we are actually implementing data flows being able to process brain segmentation, brain tissue classification (white matter, grey matter, CSF), cortical sulci delineation and rigid and non-rigid registration of MR images. At the present time, security aspects are just addressed by using a SSH tunneling for the transactions between the servers and the clients. Confidentiality of the data including anonymization has to be performed by the publisher. This issue should be addressed in future versions. The computation mapping implemented in the current version (i.e. where data meet the image processing codes) can be in principle performed everywhere in the collaborative network. In practice, because we haven't set up efficient processing servers on each collaborative site, in current experiments, codes are executed were the data are located.

At the present time, this demonstrator is being developed and tested in the context of two major applications:

- -Clinical application dealing with epilepsy surgery
- -Cognitive application dealing with the delineation of the visual cortex areas

Secondary applications are being addressed such as in clinical neurosciences (Alzheimer, Multiple Sclerosis …).

---

[5] http://www.le-select.com/
[6] http://brainvisa.info/
[7] http://www.irisa.fr/visages/software-fra.html
[8] http://www.fil.ion.ucl.ac.uk/spm/
[9] http://www.postgresql.org/
[10] http://jakarta.apache.org/tomcat

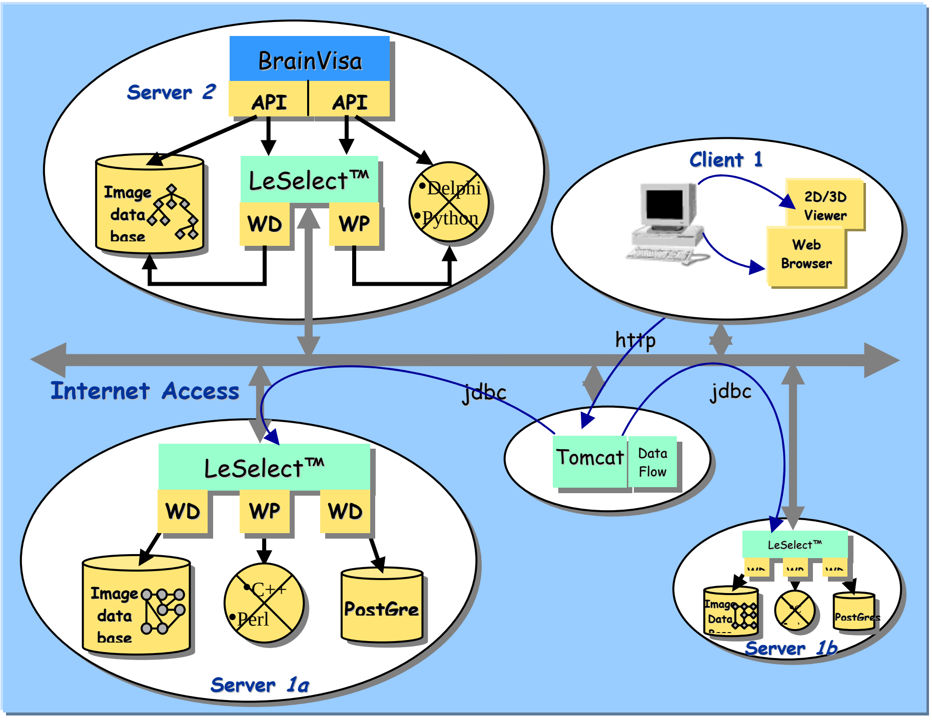

Figure 2: Neurobase demonstrator architecture with a generic server model ("*server 1i*") based on the "LeSelect" architecture with a PostGres/SQL data base for managing storing of temporary image data computed from Data Flow process. The server 2 is a specific extension of the generic one in order to ensure compatibility with the BrainVisa environment while keeping the same service from the Client.

In summary

The NeuroBase project offers a unique opportunity to acquire knowledge and experience about data and knowledge sharing and reuse, data fusion, data mining and knowledge discovery, in the real context of large distributed neuroimaging sources. We are proposing a specific and original system able to manage data and computing procedures shared by a collaborative network of partners motivated to distribute their expertise and experience. The proposed system infrastructure is able to deal with heterogeneous sources of information distributed within this collaborative network, without the requirement of developing and hosting a master (shared) repository. A preliminary demonstrator is currently under evaluation to exhibit the added value of such scheme in the neuroimaging domain.

References

1. Mazziotta J.C., Toga A.W., Evans A.C., Fox P., and Lancaster J.L., A Probabilistic Atlas of the Human Brain: Theory and Rationale for its development, Neuroimage, 2:89-10, 1995.
2. Roland P.E., Zilles K., Brain atlases - a new research tool, Trends in neurosciences, 17(11):458-467, 1994.
3. D. Graf von Keyserlingk, K. Niemann, and J. Wasel, “A quantitative approach to spatial variation of human cerebral sulci,” Acta Anatomica, 131:127-131, 1988.
4. W. Toga, “A Three-Dimensional atlas of structure/function relationships,” Journal of chemical neuroanatomy, 4:313-318, 1991.
5. F. E. Bloom, “The multidimensional database and neuroinformatics requirements for molecular and cellular neuroscience,” Neuroimage, 4:S12-3, 1996.
6. D. C. Van Essen, H. A. Drury, S. Joshi, and M. I. Miller, “Functional and structural mapping of human cerebral cortex: solutions are in the surfaces,” Proc Natl Acad Sci U S A, 95:788-795, 1998.
7. R. J. Seitz, C. Bohm, T. Greitz, P. E. Roland, and L. Erikson, “Accuracy and precision of the computerized brain atlas programme for the localization and quantification in positron emmision tomography,” J. Cerebr Blood Flow Metab, 10:443-457, 1990.
8. E. D. Lehmann, D. J. Hawkes, D. L. Hill, C. F. Bird, G. P. Robinson, A. C. Colchester, and M. N. Maisey, “Computer-aided interpretation of SPECT images of the brain using an MRI-derived 3D neuro-anatomical atlas,” Medical Informatics, 16:151-66, 1991.
9. K. H. Höhne, M. Bomans, M. Riemer, R. Schubert, U. Tiede, and W. Lierse, “A volume based anatomical atlas,” IEEE Comp. Graphics and Application, 12:72-78, 1992.
10. Tiede, T. Schiemann, and K. H. Höhne, “Visualizing the Visible Human,” IEEE computer Graphics and Applications, 16:7-9, 1996.
11. Kahng J. and McLeod D. “Dynamic Classificational Ontologies: Mediation of information sharing in Cooperative Federated Database Systems.” In: Cooperative Information Systems: Trends and Directions, Michael P. Papazoglou and Gunter Schlageter eds, pp.179-203. Academic Press, London, U.K., 1998.
12. Wiederhold G. and Genesereth M., “The Conceptual Basis for Mediation Services”, IEEE Expert, 12(5):38-47, 1997.
13. A. Tomasic et al. “The distributed Information Search Component (Disco) and the World Wide Web”, Proceedings of the ACM SIGMOD Int. Conf., pp. 546-548, Tucson, Arizona, May 1997.
14. M. Rodriguez-Martinez, N. Roussopoulos, “MOCHA: a Self-extensible Middleware Substrate for Distributed Data Sources”, in Proc. of the ACM SIGMOD: Int Conf on Management of Data, Houston, May 2000.
15. Rosse C., Mejino J.L., Modayur B.R., Jakobovits R., Hinshaw K.P., Brinkley J.F., “Motivation and organizational principles for anatomical knowledge representation,” Journal of the American Medical Informatics Association, 5(1):17-40, 1998.
16. B. Gibaud, H. Garfagni, F. Aubry, A. T. Pokropek, V. Chameroy, Y. Bizais, and R. D. Paola, “Standardization in the Field of Medical Image Management: The Contribution of the MIMOSA Model,” IEEE Transactions on Medical Imaging, 17:62-73, 1998.
17. Bowden, D. M. and Martin, R. F., “NeuroNames Brain Hierarchy.” NeuroImage 2:63-83., 1995
18. Van Horn J.D., Grafton S.T., Rockmore D., Gazzaniga M.S., “Sharing neuroimaging studies of human cognition”, Nature Neuroscience, 7:473:481, 2004