

# NeuroLOG: A federated software architecture for the sharing and reuse of distributed data and processing tools in neuroimaging

Michel Dojat<sup>1</sup>, Mélanie Péligrini-Issac<sup>2,3</sup>, Bénédicte Batrancourt<sup>4,3</sup>, Alban Gaignard<sup>5</sup>, Bernard Gibaud<sup>6</sup>, David Godard<sup>7</sup>, Gilles Kassel<sup>8</sup>, Franck Michel<sup>5,6</sup>, Johan Montagnat<sup>5</sup>, Xavier Pennec<sup>9</sup>

1. INSERM / Université Joseph Fourier, U836, Institut des Neurosciences, Grenoble, France
2. INSERM / UPMC Univ. Paris 06, UMR\_S678, Laboratoire d'Imagerie Fonctionnelle, Paris, France
3. Univ. Paris 11, IFR49, Gif-sur-Yvette, France
4. INSERM / CNRS / UPMC Univ. Paris 06, UMR\_S975 CRICM, Paris, France
5. CNRS / UNS, I3S lab, MODALIS team, Sophia Antipolis, France
6. INSERM / INRIA / CNRS / Univ. Rennes 1, IRISA Unit VISAGES U746, Rennes, France
7. Visioscopie, Nice, France
8. Univ. de Picardie Jules Verne, MIS, EA 4290, Amiens, France
9. INRIA, ASCLEPIOS team, Sophia Antipolis, France

*Abstract—In this work, we propose a federated software architecture to facilitate the integration of neuroimaging data and processing tools available from distributed neuroscience research or clinical centres, while preserving an autonomous data management for each partner centre. This architecture relies upon an application ontology dedicated to the description of neuroimaging MR data, neuropsychological scores, and tools. Semantically annotated processing tools are packaged as Web Services and can be pipelined by means of a workflow manager. In the future, such a framework is aimed at facilitating the sharing of data and tools, and the organization of multicentric neuroimaging studies.*

## Keywords

*Database, Ontology, Collaborative work, Data integration, Mediation*

## I. INTRODUCTION

Sharing a part of the huge amount of neuroimaging data produced in research studies in neuroscience or clinical centres is a challenging topic in our quest for understanding brain function and finding neural substrates and imaging biomarkers for brain diseases. Despite strong efforts to develop relevant federated infrastructures [1], structural and semantic heterogeneity of data sources remains a major obstacle to sharing. Databases already exist that either provide a large amount of unstructured data (e.g. ADNI [2]) or require to adopt a specific data organization schema (such as Xnat [3] data model) not necessarily compatible with local end-user preferences.

Similarly, sharing distributed processing tools, com-

binning them to define a robust chain for a specific application and facilitating its execution on grid platform is inevitable for large multicentric studies. Solutions exist (e.g. SPM batchs [4], BrainVISA platform [5], Nipype initiative [6], Xnat pipeline engine [3]) that automate in part data analysis and enhance processing pipelines diffusion. However data provenance [7] and semantic links between individual processing tools are missing, hindering the composition of tools from various origins.

In this context, we propose a federated architecture for the integration of neuroimaging data and tools available from multiple distributed sites.

## II. METHODS

The NeuroLOG architecture allows the integration of partly structured, heterogeneous and distributed data, and the design and execution of processing workflows in an open environment. A convivial user interface is proposed for browsing data, metadata, and tools, and comprises an integrated image viewer. Protection of private medical data is guaranteed.

An extensive effort has been put to define a common semantics through an application ontology : OntoNeuroLOG [8], which describes datasets (MR images) and entities involved in their generation, such as MR acquisition protocols and sequences, as well as subjects or studies. It also models neuropsychological instruments and behavioral scales in order to take the results (scores) of clinical examinations into account. OntoNeuroLOG relies on a foundational ontology (DOLCE [9]) ensuring a common and consistent modeling framework. From this ontology, a relational federated schema is derived, to which local legacy database schemas are dynamically mapped. This pro-

vides the end-user with a global and federated view of metadata and processing tools while hiding their initial heterogeneity. Each local site hence preserves its autonomous data management. Distributed tools are integrated and pipelined using a workflow manager [10]. Tools are packaged in self-consistent executable units exposed as Web Services, which can be remotely invoked. Most importantly, semantic annotation of processing tools is available through the ontology. A specific data management layer facilitates the access to core grids middleware.

### III. RESULTS

The NeuroLOG platform (see global architecture in Figure 1) currently includes five participating sites. Figure 2 shows how datasets involved in selected studies can be retrieved. Figure 3 shows the user interface for semantic annotation of processing tools (here : brain extraction tool BET from the FSL library [11]). Figure 4 shows how tools from the semantic repository (left side) can be composed using the workflow manager interface (right side).

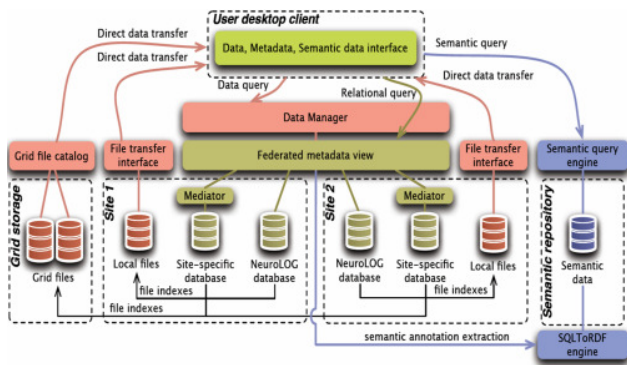


FIG. 1 – The federated global architecture of NeuroLOG. Only two partner centres are shown.

### IV. CONCLUSION

The NeuroLOG platform provides a federated view of distributed neuroimaging resources (associated metadata and semantic data, tools with semantic annotation), while preserving local site autonomy and data organization. The workflow manager allows remote execution of processing pipelines and interface with the EGI grid is being implemented. Further work concerns the semantic validation of processing tools (to check compatibility of inputs/outputs and assist the user in workflow design) as well as the semantic annotation of produced results (so that they can be retrieved from the federated view).

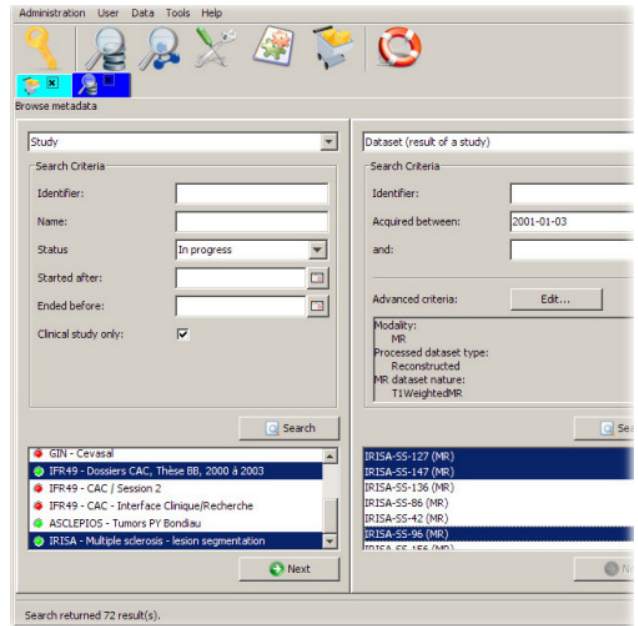


FIG. 2 – Interface for browsing available datasets from selected centres.

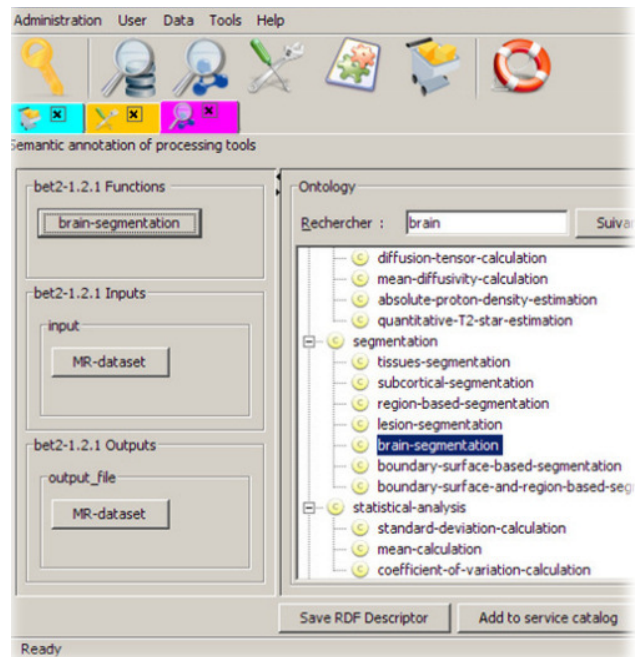


FIG. 3 – Interface for browsing available processing tools in each centre based on their corresponding semantic annotations.

### ACKNOWLEDGEMENTS

This work was funded by the French national research agency (ANR), NeuroLOG project<sup>1</sup>, under contract number ANR-06-TLOG-024 (2007-2010).

<sup>1</sup>NeuroLOG : <http://neurolog.polytech.unice.fr>

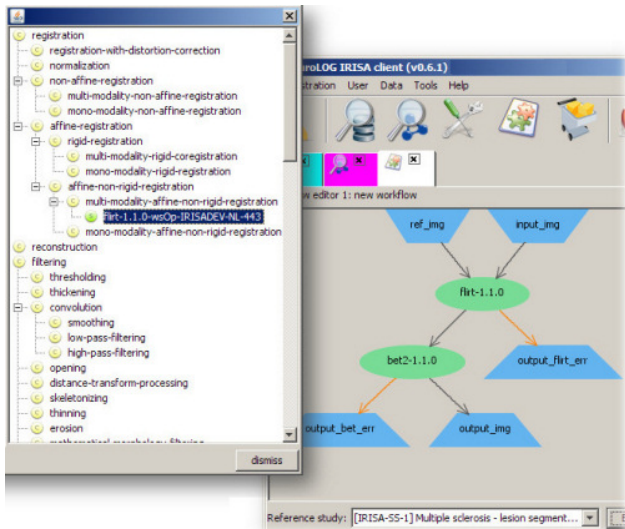


FIG. 4 – Interface for designing a processing chain from semantically annotated tools shared between centres.

with MOTEUR, *International Journal of High Performance Computing Applications (IJHPCA)*, Vol. 22, pp. 347–360, 2008.

- [11] Smith, S.M., et al., Advances in functional and structural MR image analysis and implementation as FSL, *NeuroImage*, Vol. 23, pp. 208–219, 2004.

## REFERENCES

- [1] Keator, D.B., et al., A national human neuroimaging collaboratory enabled by the Biomedical Informatics Research Network (BIRN), *IEEE Transactions on Information Technology in Biomedicine*, Vol. 12, pp. 162–172, 2008.
- [2] ADNI, Alzheimer’s Disease Neuroimaging Initiative <http://adni.loni.ucla.edu/>.
- [3] Marcus, D.S., et al., The Extensible Neuroimaging Archive Toolkit (XNAT) : An informatics platform for managing, exploring, and sharing neuroimaging data, *Neuroinformatics* Vol. 5, pp. 11–34, 2007.
- [4] <http://www.fil.ion.ucl.ac.uk/spm/>.
- [5] Cointepas, Y., et al., The BrainVISA project : a shared software development infrastructure for biomedical imaging research, In *Proceedings of 16th HBM Meeting*, 2010
- [6] Nipype : Neuroimaging in Python, Pipelines and Interfaces, <http://nipy.sourceforge.net/nipype/>.
- [7] Mackenzie-Graham, A.J., et al., Provenance in neuroimaging, *NeuroImage*, Vol. 42, pp. 178–195, 2008.
- [8] Michel, F., et al., Grid-wide neuroimaging data federation in the context of the NeuroLOG project, In *Proceedings of HealthGrid’10 (HG’10)*, 2010.
- [9] Masolo, C., et al., The WonderWeb Library of Foundational Ontologies and the DOLCE ontology, *WonderWeb Deliverable D18, Final Report*, vr. 1.0, 2003.
- [10] Glatard, T., et al., Flexible and efficient workflow deployment of data-intensive applications on grids