NeuroLOG: A framework for the sharing and reuse of distributed tools and data in neuroimaging

Submission No:

2088

Authors:

Michel Dojat[1], Mélanie Pélégrini-Issac[2,3], Farooq Ahmad[4], Christian Barillot[4], Bénédicte Batrancourt[5,3], Alban Gaignard[6], Bernard Gibaud[4], Pascal Girard[1,7], David Godard[8], Gilles Kassel[9], Diane Lingrand[6], Grégoire Malandain[7], Franck Michel[4], Johan Montagnat[6], Xavier Pennec[7], Javier Rojas Balderrama[6], Bacem Wali[4]

Institutions:

[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 / INRIA / CNRS / Univ. Rennes 1, IRISA Unit VISAGES U746, Rennes, France, [5]INSERM / CNRS / UPMC Univ. Paris 06, UMR_S975 CRICM, Paris, France, [6]CNRS / UNS, I3S lab, MODALIS team, Sophia Antipolis, France, [7]INRIA, ASCLEPIOS team, Sophia Antipolis, France, [8]Visioscopie, Nice, France, [9]Univ. de Picardie Jules Verne, MIS, EA 4290, Amiens, France

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 its alterations. Despite strong efforts to develop relevant federated infrastructures [4], 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 [1]) or require to adopt a specific data organization schema (such as Xnat [6] data model) not necessarily compatible with local end-user preferences. Similarly, sharing distributed processing tools, combining 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 [10], BrainVISA platform [2], Nipype initiative [9], Xnat pipeline engine [6]) that automate in part data analysis and enhance processing pipelines diffusion. However data provenance [5] 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.

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 [7], 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 to account for clinical examination scores. OntoNeuroLOG relies on a foundational ontology (DOLCE) 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 provides 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 [3]. 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.

Results:

The NeuroLOG platform (see global architecture in Fig. 1) currently includes five participating sites. Fig. 2 shows how datasets involved in selected studies can be retrieved. Fig. 3 shows the user interface for semantic annotation of processing tools (here: brain extraction tool BET from the FSL library [8]). Fig. 4 shows how tools from the semantic repository (left side) can be composed using the workflow manager interface (right side).

Conclusions:

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). This work was funded by French National Agency for Research (NeuroLOG project ANR-06-TLOG-024).

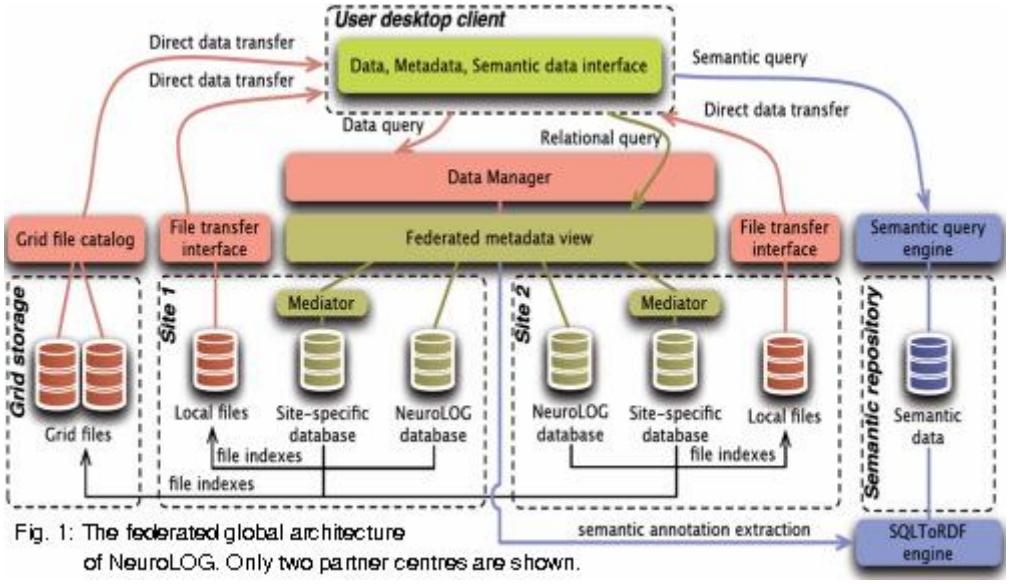

Fig. 1: The federated global architecture of NeuroLOG. Only two partner centres are shown.

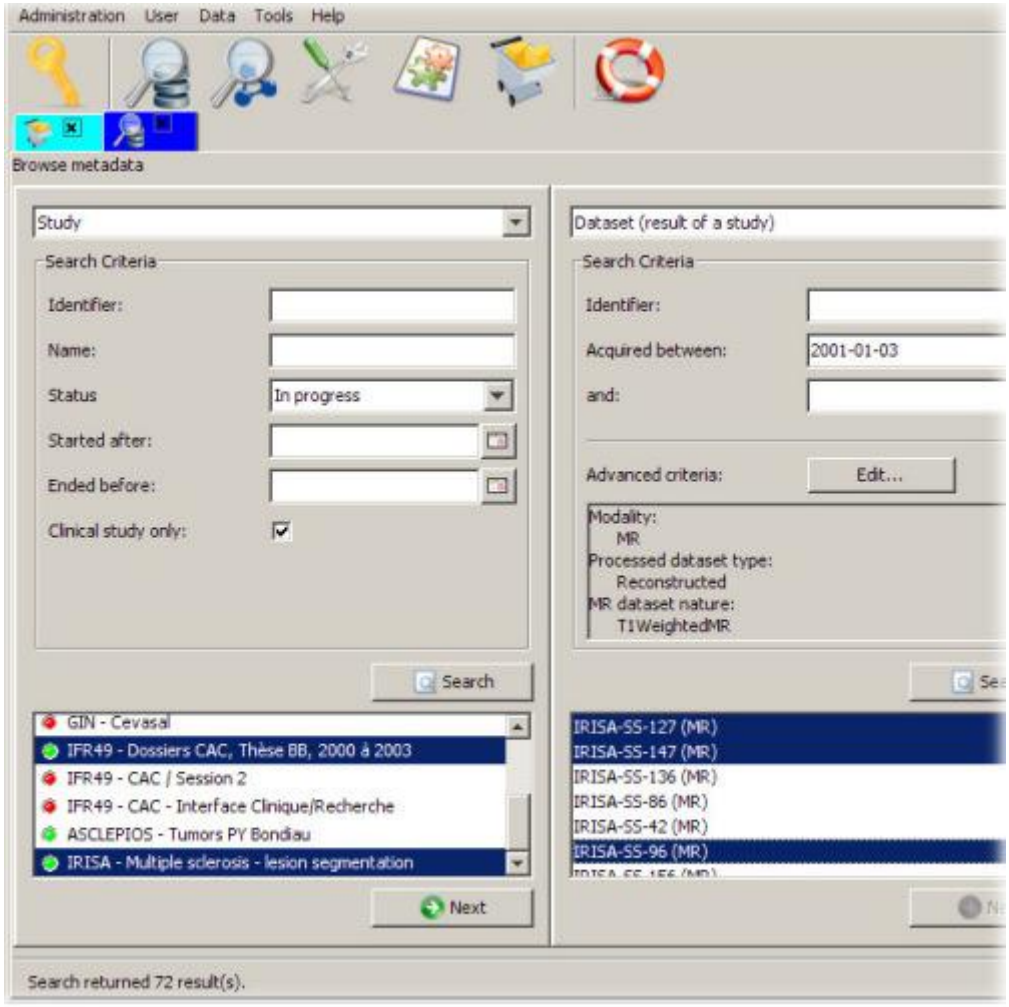

Fig. 2: Interface for browsing available datasets from selected centres.

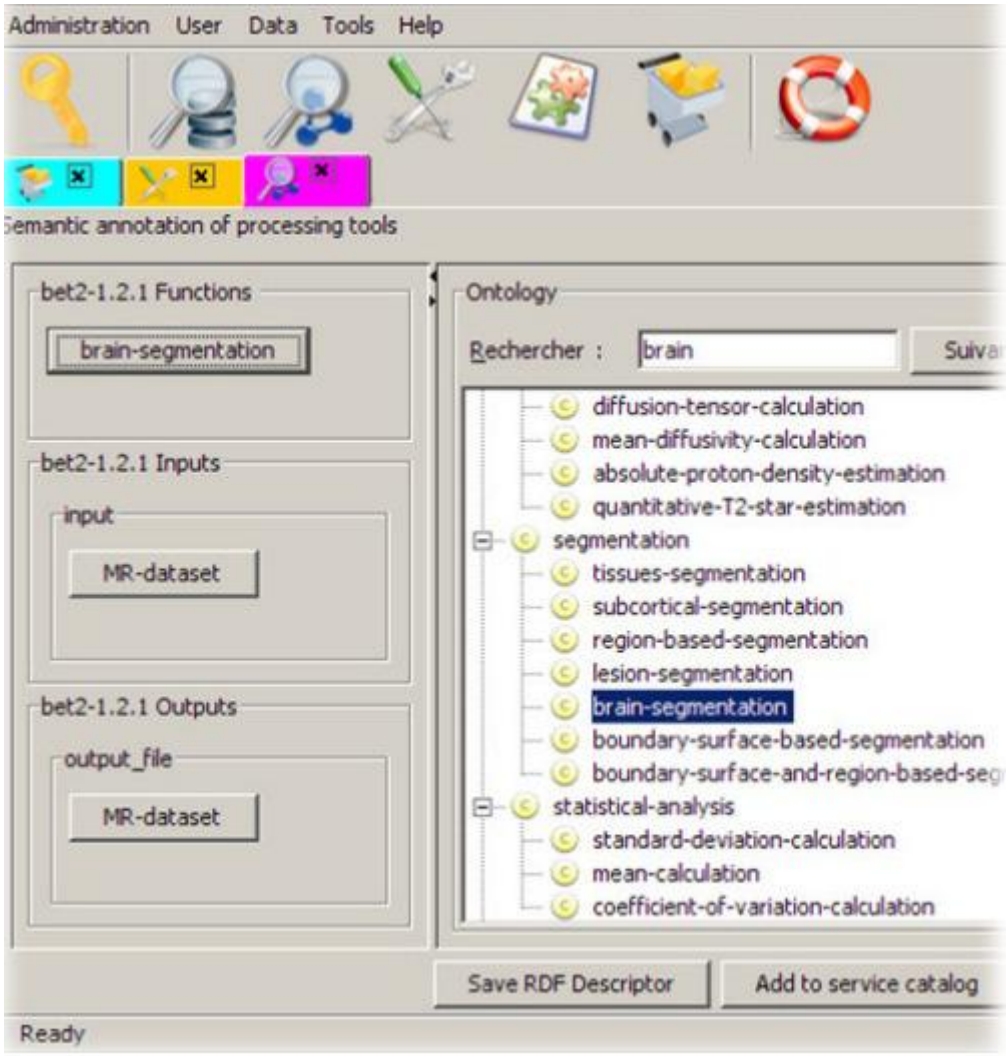

Fig. 3: Interface for browsing available processing tools in each centre based on their corresponding semantic annotations.

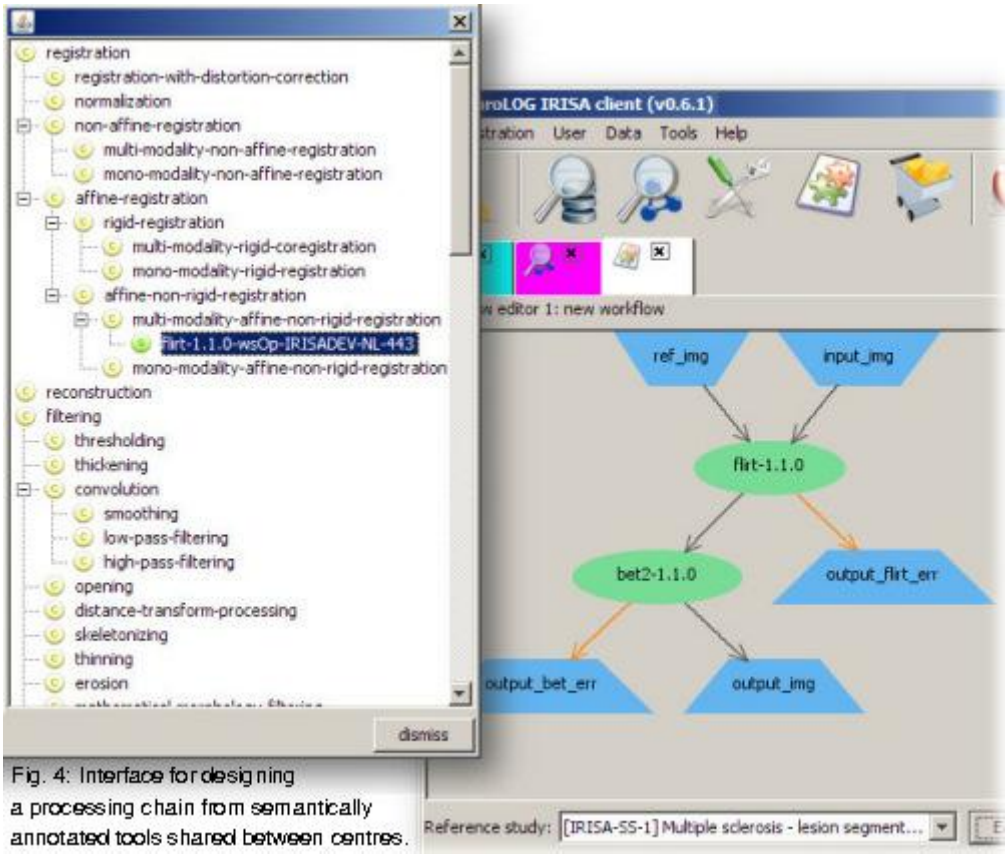

Fig. 4: Interface for designing a processing chain from semantically annotated tools shared between centres.

Informatics

Databasing and Data Sharing

Abstract Information

References

[1]. ADNI, Alzheimer's Disease Neuroimaging Initiative http://adni.loni.ucla.edu/.
[2]. Cointepas, Y., et al. (2010), 'The BrainVISA project: a shared software development infrastructure for biomedical imaging research', In Proceedings of 16th HBM Meeting.
[3]. Glatard, T., et al. (2008), 'Flexible and efficient workflow deployement of data-intensive applications on grids with MOTEUR', International Journal of High Performance Computing Applications (IJHPCA), vol. 22, pp. 347-360.
[4]. Keator, D.B., et al. (2008) '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.
[5]. Mackenzie-Graham, A.J., et al. (2008), 'Provenance in neuroimaging', NeuroImage, vol. 42, pp. 178-195.
[6]. Marcus, D.S., et al. (2007), 'The Extensible Neuroimaging Archive Toolkit (XNAT): An informatics platform for managing, exploring, and sharing neuroimaging data', Neuroinformatics vol. 5, pp. 11-34.
[7]. Michel, F., et al. (2010) 'Grid-wide neuroimaging data federation in the context of the NeuroLOG project', Proceedings of HealthGrid'10 (HG'10).
[8]. Smith, S.M., et al. (2004), 'Advances in functional and structural MR image analysis and implementation as FSL', NeuroImage, vol. 23, pp. 208-219.
[9]. Nipype: Neuroimaging in Python, Pipelines and Interfaces, http://nipy.sourceforge.net/nipype/.
[10]. SPM: http://www.fil.ion.ucl.ac.uk/spm/.