NeuroLOG: a community-driven middleware design

Johan Montagnat, Alban Gaignard, Diane Lingrand, Javier Rojas Balderrama, Philippe Collet, Philippe Lahire

NeuroLOG: a community-driven middleware design

Johan MONTAGNAT [a,1], Alban GAIGNARD [a], Diane LINGRAND [a],
Javier ROJAS BALDERRAMA [a], Philippe COLLET [a] and Philippe LAHIRE [a]
[a] *UNSA / CNRS, I3S laboratory, 930 route des Colles, 06903 Sophia Antipolis, France.*

Abstract. The NeuroLOG project designs an ambitious neurosciences middleware, gaining from many existing components and learning from past project experiences. It is targeting a focused application area and adopting a user-centric perspective to meet the neuroscientists expectations. It aims at fostering the adoption of HealthGrids in a pre-clinical community. This paper details the project's design study and methodology which were proposed to achieve the integration of heterogeneous site data schemas and the definition of a site-centric policy. The NeuroLOG middleware will bridge HealthGrid and local resources to match user desires to control their resources and provide a transitional model towards HealthGrids.

Keywords. Medical Image Analysis, Grid Computing, Neurosciences

1. Goals

Grid computing has been seriously considered to tackle a broad range of requirements arising from the clinical world [2]. In particular, the coarse grain parallelism of grids suits well the manipulation of large medical data sets geographically fragmented and independent computations over large patient populations [18]. As a concrete example, HealthGrids can provide data mediation interfaces, secured and efficient transfer protocols as well as authentication & authorization schemes enabling data exchanges across health enterprises boundaries [11]. Yet, there are many difficulties limiting the practical usability of grids when considering real clinical scenarios. In particular, stringent security constraints apply [5], reluctance to externalize many medical resources is often encountered and new technologies cannot disrupt the use of well established legacy systems.

The NeuroLOG project[2] described in this paper aims at integrating software technologies and medical resources for supporting the neurosciences community needs. The NeuroLOG middleware builds on many existing software components to face the software design complexity. To really meet end users expectations and to ease technology adoption, it was decided to focus on the neuroscience community. Neuroscientists are pre-clinical end-users showing a high familiarity with Information Technologies. Computational neurosciences have for long demonstrated the power of computing techniques

[1]Corresponding Author: Johan Montagnat, http://www.i3s.unice.fr/~johan, E-mail: johan@i3s.unice.fr.

[2]NeuroLOG: http://neurolog.polytech.unice.fr

to analyze neurological data sets and study the brain functions. Large scale infrastructures have often been deployed in brain imaging centers steered by the computational needs arising from multi-patients statistical studies and biomodeling activities. The neuroscience community is keen on exploring further computational technologies able to tackle the challenging problems arising in this field [12,4]. To further circumscribe the expertize area covered, NeuroLOG specifically targets three specific pathologies: multiple sclerosis, brain tumours and brain strokes.

The NeuroLOG consortium mixes partners with expertise in the areas of software technologies, knowledge representation, and neurological image analysis. The imaging experts input is driving the technical choices and policies applied. Grid technologies is clearly targeted: the EGEE [10] European production grid infrastructure is already exploited for most demanding processings. However, the need for a transitional model is also recognized. Participating centers are mostly exploiting privately their internal resources so far. The grid is considered as an extension which should not require additional effort to exploit. From a user perspective, the interest is not in migrating towards the grid but rather in integrating the grid resources transparently into the locals methodological workflows to ensure continuity with the legacy systems in use.

This paper describes the NeuroLOG project goals and methodology. It targets in particular the area-specific requirements analysis and known limitations of existing systems (section 2) from which admissible policies are proposed (section 3). The software architecture design is then described (section 4).

2. The user in front

To help gaining the users interest, the starting point was the analysis of the users practices and their image processing pipelines for each application targeted. There are many commonalities in the data sets and processing chains for the pathologies studied. The data sets manipulated are composed of images completed with clinical information and a rich variety of additional annotations (*e.g.* segmentation contours, related medical records…). The imaging modalities used are mostly Magnetic Resonance (MR) modalities (T1, T2, FLAIR…) with some Computed Tomography (CT) and Positron Emission Tomography (PET) scans. Basic processings are common to all image analysis pipelines. They include intensity corrections, images co-registration, tissues classification and skull stripping. Each pipeline also contains pathology specific processings such as brain structures segmentation, parameters quantification, image interpretation and visualization. In some cases, variations of similar algorithms are needed (*e.g.* both mono-modal and multimodal registration may be needed).

Unsurprisingly, the technical analysis of user requirements led to: data sets organization and federation; data collections selection; data sharing and access control; algorithms exchanges and comparison; image processing pipelines description and efficient execution; and computation results exploration. These items have to be interpreted in the context of the user practices to steer useful software development though. The technical analysis requires permanent feedback from users to reach a satisfying level of quality.

In practice, the structures of manipulated data sets are very heterogeneous (from disk file hierarchies to complex relational databases) and data is spread over many independent databases [3]. Relevant data selection to address brain studies requires advanced search capabilities. Different kinds of data representations are therefore considered:

- files, mostly images, containing the core medical data;
- associated metadata from different origins: medical metadata associated to image files, image processing tools metadata, administrative metadata…; and
- semantic data, enabling rich queries and retrieval capabilities.

The organization of medical data, and especially the organization of metadata, is a very complex problem as soon as realistic use cases are considered. Large scale databases have been deployed in some cases (ADNI [1], DDSM [8]). However, they are usually making simplistic assumption such as the public availability of the whole data or the centralization of data stores. Many initiatives have proposed area-specific data schemas and are relying on the user adoption of a specific structures to build reference databases [4,6]. This policy was unacceptable to the different users within the project: there is too much study-specific information to expect building a common schema without introducing over-complexity. It was rather decided that the system had to cope with site-specific schemas, providing a core common structure and extensibility. Data security is also not required, and therefore hardly addressed, internally to each site as long as data is not transferred outside. However this aspect becomes critical as soon as data sharing facilities are provided as each site expects to control access to its data resources.

The need for exchanging image analysis tools arises from the need to compare and validate results as well as to mutualize software development efforts. Some initiatives such as the ITK medical image processing library [15] or the SPM [22] have contributed a lot to the homogenization of medical imaging softwares and analysis procedures. However they do not account for the local development efforts from many specialized teams. Similarly to common data schema, these approaches make the assumption of all integrated, tightly coupled codes. Studying the software development status among the participating sites revealed the large heterogeneity of technical foundations adopted over the past years. Tools for deploying and exploiting different flavors of algorithms are increasingly needed. These are well known from the software engineering and Service/Component Architecture communities. In addition, image analysis procedures are often not limited to the application of a single algorithm but are best described as processing chains or pipelines.

In the light of these requirements, the EGEE infrastructure and its middleware were studied. EGEE provides a multi-sciences foundational grid computing service. The infrastructure is a federation of computing centers, each operating batch computing resources and storage repositories. Initially motivated by the High Energy Physics embarrassingly parallel problems, EGEE provides a very large scale service (more than 40,000 CPUs in more than 250 computing centers today) shared by several scientific communities or *Virtual Organizations* (VOs). In EGEE, VOs represent the security control units. Based on the VO she belongs to, each user is granted access to a given amount of the grid resources. The EGEE computing resources are accessible through a Workload Management System designed as a two levels batch-system: a *Resource Broker* (RB) queues computing requests and dispatch them to various site batch systems. The storage resources are interfaced through the OGF [19] promoted standard *Storage Resource Manager* (SRM). The EGEE Data Management System provides a virtual file hierarchy expanding over the participating sites resources. EGEE has been exploited in production for more than 3 years: the usability of such an infrastructure for many medical imaging-related applications has been demonstrated [18]. However, it also only provides a low-level middleware layer with respects to the complexity of the neuroscience requirements.

VOs are representing very coarse grain control units. Increasing efforts are made to refine VOs and authorization policies but few middleware components can exploit VO subgroups and user roles currently. The Data Management System only supports file hierarchies and it hardly provides any tool to structure neurological data. The computing infrastructure is homogeneous and application codes are assumed to be transportable as binary executables on the Linux worker nodes of the grid. The aim of EGEE is to provide a sustained production quality service to its users. Yet, users are still reluctant to archive precious original data on grid storage: several issues of files migration (upon maintenance operations or storage system failures) and long term archiving are not solved yet.

The NeuroLOG middleware aims at complementing the EGEE services in light of the neuroscience needs and at integrating both site-specific and grid-wide resources.

3. Security policies

The NeuroLOG platform is a federation of administratively independent neuroscience institutes. Users belonging to the different institutions have both collaborative interests and competiting activities. They are also tided by local ethical committee rules. Sites have invested in the local storage and computing infrastructure, although they are not dedicated computing centers. The resources at each site are governed by a local policy. Grid resources in the other hand are administrated in external computing centers, usually non specialized in medical data handling, with different policies.

Multi-sites data federation is the highest priority requirement to foster collaborative work. In a widely-distributed environment, with long range communications over the Internet, data access control and protection is critical to assemble distributed system-wide data sets. To achieve data sharing in practice, the data security requirements and usages have to be taken into account: data sets are often assembled for groups of neuroscientists for a particular experiment. The primary storage entrusted by the users for their original data is the local resources. The grid is seen as more experimental and volatile, although this may change as the users become more confident and used to this externalized storage.

The NeuroLOG Security Policy (NSP) described in this section aims at fulfilling the project security requirements, especially regarding sensitive medical data access control. It accounts for two *a priori* antagonist roles: to make data exchanges among users from different sites possible; and to ensure that each site solely controls the access to the data it owns. The proposed solution intends to be as lightweight and easy to deploy as possible. Both local site data and on-grid data are considered. To summarize, the main aspects of the proposed policy are:

1. The NSP is administrated locally on each site by site administrators. There is no global administrator of the distributed platform.
2. All users are securely identified and registered into the system. A few particular users have administration privileges at each site.
3. Data access is controlled at the level of user groups. A group is created and owned by one site but users from other sites may be registered into it. Site administrators may decide which data on their site is accessible to which group.
4. Accesses to local data is traced individually in a non repudiable manner on each site.

This policy ensures that each site controls its data: local data access is under the responsibility of the site administrators. Complementary to the NSP, all data exported from a site is anonymized and encrypted prior to transmission for protection.

3.1. Users, administrators and groups

To implement the NSP, all system users are authenticated through non-repudiable nominative X509 certificates. Each user is registered into one site (and thus known by the system) by site administrators. Normal users and administrators certificates identify their owner name and institution. Several credentials are managed in the NeuroLOG platform to deal with the different services composing the system: login/password identification, CPS (Health Professional Smartcard), SQL92 identifiers and certificates. The system ensures the proper mapping of a single user certificate to all these credentials to interoperate with the services.

On each site, a group of users (usually limited to one administrator with a deputy) gets the administrator privileges allowing its members to (i) register or unregister other users into the local site; (ii) change the administrators privilege recipient; (iii) create groups and populate groups with user identities; and (iv) grant group access to individual data files. Site administrators are warrant of the local site data control. Access to data is controlled at the group level: as many groups as needed may be created and data files are individually controlled by group. A group is a unique name identifying a list of users. Note that a group may contain users from different sites. Two particular groups are automatically created on each site upon system installation:

- A site-specific group. All members registered to a site will belong to this group. By default all data registered to a site will be readable by members of the site group. No members of other sites can be registered into the site-specific group.
- An administrator group containing the administrator users. All data registered to a site will be readable and writable to the administrator group.

Other groups are created and populated without restrictions by site administrators.

A group is locally administrated by the site administrators it belongs to but users from different sites may be registered into a same group. Conversely, site administrators can grant access to their local files for groups owned by external sites. Thus, users belonging to different sites can share data from multiple sites upon joint authorization by the group administrator and the site administrator the file belongs to. Each site controls the access to its files and the administrator is the warrant of the application of the site access control policy.

To exploit grid storage resources, VO groups are created in the EGEE VO Management System (VOMS). The use of VOMS groups for access control is only supported by a limited number of storage resources today (SRM v2 compatible resources) and the NeuroLOG middleware therefore limits the use of grid storage to compliant resources.

3.2. Operational set up

To implement the NSP, the administration services of the NeuroLOG sites have to cooperate. A NeuroLOG registry facilitates this operation. The registry main role is to register all sites participating to the platform deployed. The registry is contacted by NeuroLOG services to register a new site, discover the participating sites and create new groups.

The registry is a centralized point of failure. Thus the system should depend as little as possible on it. With the proposed solution, it is needed only upon sites and groups creation which are believed to be rare events. In a long term it would make sense to replicate this service to ensure better fault tolerance.

Upon a new site deployment, the registry is contacted. It registers a unique site name, the site administration service host IP, the site administrators certificate and email addresses. At any time the registry can be queried by one of the sites to discover the other sites registered thus ensuring dynamic extension of the system. Furthermore, the registry allocates a single prefix to each site, thus ensuring the uniqueness of file identifiers generated on each site.

4. Software architecture

The NeuroLOG middleware is decomposed into 7 main components diagrammed in figure 1. The N*euro*LOG *registry* described above is needed to implement the NSP. *Certification authorities* (CAs) are used to deliver sites and users' certificates. Certification authorities might be external to the NeuroLOG middleware (*e.g.* the certificates delivered by the EGEE CAs are recognized) but the system is also able to create its own CA to make is possible to deploy a completely independent infrastructure. An *administration* component is responsible for users, groups and authentication services. A *data management* component is in charge of federated local and grid files management. A *metadata management* component similarly handles distributed metadata. It is also involved in administrative operation *e.g.* to index data files and to record security traces. A *semantic data management* component handles knowledge data. It populates the knowledge database from relational metadata and exploits a domain ontology to manipulate it. Finally a *processing tools manager* interfaces local and grid computing resources.

The system components are deployed on different sites as illustrated in figure 1. The global registry belongs to a single root site and the semantic repository is similarly centralized. All other services are distributed over the sites. The system is inherently distributed and the different system components communicate through various protocols depending on the tools used to implement it. Users access the system through clients that can be mobile: typically site services are deployed on a centralized, fixed host while users connect from their laptop. There are therefore 3 layers communicating in the system:

- *Client layer*: the system is accessed through a rich client application providing data management, visualization and processing tools access functionality. For convenience, a lightweight web-based client with a limited data query interface is also developed. The client is multi-platforms as Linux, MacOS X and MS Windows hosts are used. A large variety of specialized medical image visualization tools will be accessible from the client through external calls if the client Operating System supports them.
- *Server layer*: the core of the system, and the service provider for each collaborating site. In order to access to resources held by other partners, each NeuroLOG server communicates with its peers.
- *Resource layer*: within a site, it is possible to manage local computing resources or local storage resources.

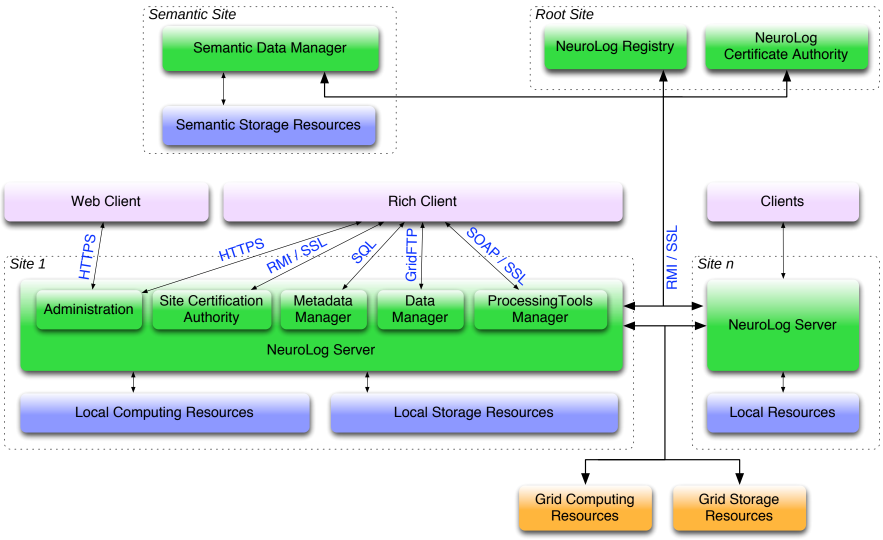

Figure 1. Platform deployment

4.1. Software components

To fulfill the desired functionality, the NeuroLOG middleware integrates as much as possible existing software components. Only complementary services are developed. The components reused are:

- *Data management*. A joint site and grid files identification schema is used. File security is achieved through access control and encryption of outbound data to ensure its protection during transmission and remote storage. Site files are transferred with GridFTP. Grid files are transferred using the EGEE Data Management System client.
- *Metadata management*. Metadata is described through a relational schema. MySQL servers are deployed over sites. The Data Federator [9] tool is used for merging metadata from different sites and expressed using different metadata schemas.
- *Semantic data*. The knowledge database is built from metadata using the METAmorphoses tool [17]. A domain ontology is developed in the context of the project. The semantic repository is queried using the CORESE semantic search engine [7].
- *Computing tools*. Computing tools are bundled locally in OSGi packages [21]. They are deployed as Web Services or Grid Services. The description of processing pipelines is done using the Scufl data flow language [20]. The pipelines are executed on the EGEE grid using the MOTEUR workflow manager [14].

The Clients, Administration and Registry components are specific to the platform and are specifically designed. The Java language has been selected for new NeuroLOG components implementation for its portability, rich library functionality and remote method invocations capability.

4.2. Protocols

The various software components communicate through different protocols as illustrated in figure 1. Due to the sensitivity of the data, all communications have to be securely encrypted. To avoid abuses of the system, all communicating participants authenticate to each other. The Secure Socket Layer provides both abilities by assigning signed certificates to all regular actors and using key-based encrypted communications.

The communication between NeuroLOG specific components is performed through Java RMI. RMI provides a fully integrated, transparent and rich invocation protocol for these pure Java components. The SSL layer is used for the RMI sockets thus ensuring authenticated and secure invocations. For external software components, different protocols are used. When possible, GridFTP is used for file transfers. GridFTP requires range ports opening which hardly comply with stringent firewall restrictions on some sites though. If immutable firewall policy apply, the servers have to be configured in single channel mode, at the cost of performance. For database communications, JDBC drivers are used. The sockets are similarly encrypted. Finally, user codes are invoked as Web Services or Grid Services through the SOAP protocol. Many of the user codes embarked are legacy binaries or regular Web Services without WSRF extensions. Therefore, the middleware provides the WS wrappers and the security extensions needed to adapt to legacy tools. A minimal web client communicating with the servers through HTTPS is also planned.

4.3. Detailed architecture

Administration. At the root level, the global registry component registers sites and group names in its database. The middleware deals with the multiple credential involved in the systems (NeuroLOG certificates, grid certificates, databases login and possibly CPS) transparently. It is possible to separate the NeuroLOG certificates from the EGEE certificates to ensure complete sites independence if desired since the EGEE CAs are managed externally: the middleware can generate a root CA and per-site sub-CAs. The administration subsystem also manages security logs. System logins and file accesses are recorded using time-stamped non-repudiable events.

Data management. The data management relies on the distributed metadata management layer to identify and discover files distributed over the different sites. It bridges the local and grid storage resources. When a user creates a new data segment, she decides whether to store it locally or on the grid. In both cases a file identifier is recorded in the user's site metadata for further retrieval. Similarly, encryption keys are associated to site controlled files and stored in the site metadata. Several medical image file formats are recognized into the system such as DICOM, Nifti and partner specific formats. Image files can be converted from one format to another. A key data management component is the file access controller compliant with the group-based NSP. Through the distributed metadata layer, it is able to localize remote data. It also controls the access to the local data based on the requester identity. In order to securely store and transport data, it is always encrypted prior to outbound communication. This ensures that data, even when stored on remote sites (*e.g.* on the grid) remains private. Only users with access right to the associate encryption key can make use of it. In addition, DICOM image files headers are anonymized to avoid sensitive nominative information to be transported out of the site with the image.

Metadata. The metadata is stored in relational databases and distributed through Data Federator (DF). DF provides a mediation interface to adapt to site-specific data schemas and a query engine working with distributed databases. The application metadata is composed into two databases: site-specific metadata is structured using a site-specific schema; and other metadata is stored using a NeuroLOG common schema. The NeuroLOG middleware can directly read and write metadata stored in the (known) common schema repository. In particular, image resulting from computations can be automatically reintegrated into this database. Metadata stored in the site-specific repository can only be imported by a site-specific tool. Yet, the DF mediator enables querying this metadata.
Semantic data. An important effort is made within the NeuroLOG project to define a complex relational data schema matching the needs expressed by the neuroscientists. A related ontology is developed to perform reasoning on this data in OWL Lite. In addition, image processing tools are also considered as knowledge entities that can be searched: a tools specific ontology is planned [16]. The semantic annotations are extracted from relational databases through DF and stored in an RDF repository. Due to the current security limitations of semantic query tools, only non-sensitive data can be stored in the knowledge repository: the metadata is filtered.
Processing tools and workflows. Processing tools or workflows can be published for external use. Image processing tools are packaged in OSGi bundles which enable versioning and dependencies control. A description of the tools is published through the site metadata and ultimately in the semantic repository to enable tools retrieval. Tools can be deployed as web services, either locally or through a grid-interfaced submission web service [13] to enable execution either on a local site resource or to the EGEE grid through the same WS invocation interface. Resulting WSs can be chained in neurological image processing pipelines using the Scufl language.

5. Conclusions

The NeuroLOG project targets an ambitious middleware development geared towards neurosciences. It is addressing the complexity of the software design through the reuse of many external components (including the EGEE middleware stack) and a sound study of the software architecture. But the main challenge is to meet the neuroscientists expectations and to foster the adoption of HealthGrids in a community close to the clinics. Although technically accessible, experience shows that this goal is difficult to achieve due to human factors such as security protective measures, data ownership or legacy systems exploitation. The approach adopted in the project design phase was a close interaction with the system end-users and the integration of their site-specific constraints. It resulted in heterogeneous site data schemas integration techniques and the definition of a site-centric policy. The NeuroLOG middleware will thus bridge HealthGrids and local resources to match user desires to control their resources and provide a transitional model towards HealthGrids.

Acknowledgements. This project is funded by the French Research Agency (ANR) under contract number ANR-06-TLOG-024. We thanks the project partners for their useful input, either on the usage or on the technical side.

References

[1] ADNI (Alzheimer's Disease Neuroimaging Initiative). http://www.loni.ucla.edu/ADNI.

[2] G. Aloiso, S. Benkner, H. Bilofsky, Ignacio Blanquer Espert, Vincent Breton, M. Cannataro, I. Chouvarda, B. Claerhout, K. Dean, S. Fiore, Kinda Hassan, G. Heeren, Vicente Hernández García, J. Herveg, M. Hofmann, C. Jones, V. Koutkias, S. Lloyd, G. Lonsdale, V. López, N. Maglaveras, Lydia Maigne, A. Malousi, F. Martin-Sanchez, R. McLatchey, E. Medico, Serge Miguet, M. Mirto, Johan Montagnat, G. De Moor, K. Nozaki, W. De Neve, I. Oliviera, Xavier Pennec, J. Sanchez, T. Solomonides, M. Taillet, P. Veltri, C. De Wagster, and R. Ziegler. HealthGrid White Paper, September 2005.

[3] C. Barillot, H. Benali, M. Dojat, A. Gaignard, B. Gibaud, S. Kinkingnéhun, J-P. Matsumoto, M. Pélégrini-Issac, E. Simon, and L. Temal. Federating Distributed and Heterogeneous Information Sources in Neuroimaging: The NeuroBase Project. In *HealthGrid'06*, pages 3–13, Valencia, Spain, June 2006.

[4] BIRN (Biomedical Imaging Research Network). http://www.nbirn.net.

[5] I. Blanquer, V. Hernandez, D. Segrelles, and E. Torres. TRENCADIS – Secure Architecture to Share and Manage DICOM Objects in a Ontological Framework Based on OGSA. In *HealthGrid'07*, Geneva, Switzerland, April 2007.

[6] CaBIG (National Cancer Institute). http://cabig.cancer.gov.

[7] CORESE (COnceptual REsource Search Engine). http://www-sop.inria.fr/teams/edelweiss/wiki/wakka.php?wiki=Corese.

[8] DDSM (Digital Database for Screening Mammography). http://marathon.csee.usf.edu/Mammography/Database.html.

[9] Data Federator. http://www.businessobjects.com/products/dataintegration/datafederator.asp.

[10] EGEE European IST project of the FP6 (Enabling Grids for E-sciencE). http://www.eu-egee.org.

[11] S.G. Erberich, J.C Silverstein, A. Chervenak, R. Schuler, M.D. Nelson, and Carl Kesselman. Globus MEDICUS - Federation of DICOM Medical Imaging Devices into Healthcare Grids. *Studies in Health Technology and Informatics*, 126:269–278, 2007.

[12] J. Geddes, C. Mackay, S. Lloyd, A. Simpson, D. Power, D. Russell, M. Katzarova, M. Rossor, N. Fox, J. Fletcher, Derek Hill, K. McLeish, J.V. Hajnal, S. Lawrie, D. Job, A. McIntosh, J. Wardlaw, P. Sandercock, J. Palmer, D. Perry, R. Procter, J. Ure, P. Bath, and G. Watson. The Challenges of Developing a Collaborative Data and Compute Grid for Neurosciences. In *19th IEEE International Symposium on Computer-Based Medical Systems, 2006. CBMS 2006*, pages 81–86, Salt Lake City, Utah, June 2006. IEEE Computer Society.

[13] Tristan Glatard, Johan Montagnat, David Emsellem, and Diane Lingrand. A Service-Oriented Architecture enabling dynamic services grouping for optimizing distributed workflows execution. *Future Generation Computer Systems*, 2008.

[14] Tristan Glatard, Johan Montagnat, Diane Lingrand, and Xavier Pennec. Flexible and efficient workflow deployement of data-intensive applications on grids with MOTEUR. *International Journal of High Performance Computing and Applications (IJHPCA)*, 2008.

[15] ITK (Insight ToolKit). http://www.itk.org.

[16] Pascal Lando, Anne Lapujade, Gilles Kassel, and Frédéric Fürst. Towards a general ontology of computer programs. In *International Conference on Software and Data Technologies (ICSOFT'07)*, pages 25–27, Barcelona, Spain, July 2007.

[17] METAmorphoses. http://metamorphoses.sourceforge.net.

[18] Johan Montagnat, Fabrice Bellet, Hugues Benoit-Cattin, Vincent Breton, Lionel Brunie, Hector Duque, Yannick Legré, Isabelle Magnin, Lydia Maigne, Serge Miguet, Jean-Marc Pierson, Ludwig Seitz, and T. Tweed. Medical images simulation, storage, and processing on the european datagrid testbed. *Journal of Grid Computing (JGC)*, 2(4):387–400, December 2004.

[19] OGF (Open Grid Forum). http://www.ogf.org.

[20] Tom Oinn, Matthew Addis, Justin Ferris, Darren Marvin, Martin Senger, Mark Greenwood, Tim Carver, Kevin Glover, Matthew R. Pocock, Anil Wipat, and Peter Li. Taverna: A tool for the composition and enactment of bioinformatics workflows. *Bioinformatics journal*, 17(20):3045–3054, 2004.

[21] OSGi Alliance. http://www.osgi.org.

[22] SPM (Statistical Parametric Mapping). http://www.fil.ion.ucl.ac.uk/spm.