

Intégration des connaissances en neurosciences dans un environnement multi-centrique

Alban Gaignard, Johan Montagnat

I3S, UMR 6070 CNRS, Université de Nice Sophia Antipolis
alban.gaignard@i3s.unice.fr
johan@i3s.unice.fr

Résumé : Dans le cadre d'études de cohortes de patients, les neurosciences computationnelles sont confrontées à la manipulation de volumes de données toujours plus importants et généralement distribués, ainsi qu'à la complexité de leurs traitements. De plus, il est en général difficile de remonter à la source des événements ayant conduit à l'échec d'une expérimentation. Les méthodes d'ingénierie des connaissances peuvent être appliquées dans le domaine des neurosciences pour enrichir la représentation des données et des chaînes de traitement par des connaissances spécifiques au domaine. Elles peuvent également être appliquées pour générer automatiquement l'information de provenance ainsi que de nouvelles connaissances, résultant de l'exploitation d'un environnement sémantique de flots de calculs. Ce papier décrit un cas d'utilisation de cet environnement ainsi que des moyens permettant l'acquisition, l'exploitation et l'extension de connaissances en neurosciences.

1 Ingénierie des connaissances en neurosciences

Introduction

Les neurosciences computationnelles s'appuient très largement sur la disponibilité de bases de données étendues et distribuées d'images cérébrales et de méta-données les complétant. La complexité des processus de traitement due généralement à l'hétérogénéité tant des données et que des outils, ainsi que le volume des données, ont amené les neuroscientifiques à planifier leurs expérimentations sous la forme de chaînes de traitements et à réaliser leurs expériences *in silico* sur des infrastructures de stockage et de calcul distribué à grande échelle.

Dans ce contexte, le projet NeuroLOG [Montagnat *et al.*] vise à proposer une plateforme collaborative de partage et de traitement cohérent de données en neurosciences. Dans ce but, deux ontologies sont développées. Une *ontologie de domaine* [Temal *et al.*] qui formalise les connaissances des experts sur les données médicales et sur les outils de traitement. Elle facilite notamment le partage des données et des traitements dans la plateforme. Une *ontologie de services*, plus technique, décrit la mise en oeuvre des traitements sous forme de services.

Nous souhaitons nous appuyer sur les ontologies développées, et les travaux existants autour des services web sémantiques, afin d'assister les neuroscientifiques dans leur travail quotidien. L'information sémantique capturée par la plateforme peut ainsi être utilisée (i) lors de la description de nouvelles expériences pour valider leur cohérence (adéquation des données et des outils de traitement utilisés), et (ii) pour faciliter la recherche de ressources neuro-informatiques selon des critères avancés. A ces fins, la plateforme doit acquérir, pérenniser mais aussi étendre les connaissances liées à son exploitation. Des connaissances sont ainsi générées au cours de la réalisation de nouvelles expériences, par exemple en produisant de nouvelles informations à partir de règles sémantiques spécifiques portant sur les traitements réalisés et identifiées par les experts du domaine.

Etat de l'art

Le support des méthodologies et technologies liées à l'ingénierie des connaissances est très prometteur pour la communauté des sciences biomédicales. En effet de nombreux projets partagent des objectifs similaires en termes d'assistance lors d'expérimentations *in silico* ou encore de provenance afin d'appréhender leur complexité.

Basé sur l'ontologie *my*Grid [Wolstencroft *et al.*] de services et de données en bioinformatique, FETA [Lord *et al.*] est un environnement léger de découverte de services impliquant spécialistes de l'ingénierie des connaissances, développeurs de services et utilisateurs scientifiques. Un annuaire de service UDDI est couplé à des descripteurs XML pour mettre en correspondance la description des services et les éléments de l'ontologie de domaine. Dû aux contraintes de légèreté de l'environnement, seules des requêtes RDF prédéfinies sont proposées, mais les auteurs soulignent que les raisonnements en logique de description sont nécessaires dans un contexte réel d'utilisation bioinformatique.

L'environnement Taverna [Belhajjame *et al.*] est également un cadre d'application de l'ingénierie des connaissances. L'objectif principal est d'exploiter les méta-données décrivant les chaînes de traitement, mais également leurs exécutions. Les bénéfices de l'annotation des services de traitement sont attendus au niveau de la recherche de services dans le catalogue. L'environnement permet également d'automatiser la génération de données de test afin de prévenir les régressions et de garantir la disponibilité et le bon fonctionnement des services. Ces expérimentations sont pour le moment en cours dans la plateforme Taverna.

Le projet BioMOBY [Gordon & Sensen] est une initiative favorisant l'interopérabilité des centres de données biologiques et les centres d'analyse de ces données. C'est un des premiers projets à exploiter en pratique le standard SAWSDL pour mettre en correspondance des descriptions syntaxiques de services, avec des ontologies de domaine.

[Lebreton *et al.*] proposent l'utilisation d'ontologies de domaine (*my*Grid), de format et de tâches afin d'améliorer l'intégration de services web sous la forme de flots de services, en particulier les activités de découverte, de sélection et de composition de services. La sélection de services est supportée par OWLS-MX ; la planification et l'exécution d'un flot de services est supportée par le moteur de composition SAREK.

Basé sur le moteur de recherche sémantique CORESE [Corby *et al.*], [Lo & Gandon] proposent une solution pour l'intégration d'applications d'entreprise au travers de services web sémantiques. L'annotation des services s'appuie sur l'ontologie OWL-S [Mar-

tin *et al.*]. Le moteur CORESE est intégré à un portail web pour permettre la découverte de services. La composition est envisagée de manière interactive, mais également automatique par la découverte de séquences de services candidats entre deux services non composables directement. La composabilité de services est alors calculée par CORESE et son moteur de règles de production.

La traçabilité des données et des traitements dans les plateformes biomédicales est une préoccupation largement partagée par la communauté des sciences biomédicales. L’initiative OPM [Moreau *et al.*] est née de cet intérêt croissant et vise à standardiser la représentation des informations de provenance dans les environnements de flots de calculs avec pour principal objectif l’amélioration de leur inter-opérabilité. Notre approche se base pour l’instant sur un modèle simple de provenance à des fins de prototypage, mais bénéficierait à terme d’un standard établi.

De nombreux travaux ont été réalisés par la communauté bioinformatique dans le domaine des services web sémantiques, notamment pour faciliter les échanges de données et l’inter-opérabilité d’outils de traitement. Les neurosciences, qui nécessitent le traitement de cohortes de données de plus en plus grandes et le partage de ressources entre centres, peuvent bénéficier aujourd’hui du même type de techniques. Une illustration est donnée dans la section 2. La section 3 présente la plateforme technique NeuroLOG pour les neurosciences. Les sections 4 et 5 exposent notre proposition pour acquérir et pérenniser l’information sémantique mise en jeu dans les expérimentations, mais également comment nous envisageons la production de nouvelles connaissances au travers de l’usage de la plateforme.

2 Motivations et cas d’utilisation

La figure 1 représente un exemple simple de chaîne de traitement qui peut être mis en œuvre dans le cadre d’une expérimentation en neurosciences. Les technologies de l’ingénierie des connaissances sont un support pour valider la cohérence de l’expérience, et pour étendre un entrepôt de données sémantiques par génération automatique d’annotations au cours de l’analyse de données.

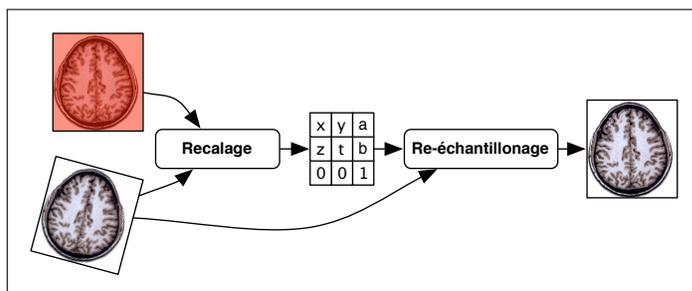


FIGURE 1 – Cas d’utilisation : chaîne de traitement impliquant différentes natures de données et de traitements.

Cette chaîne de traitement représente un processus de recalage d’images qui consiste à superposer deux ensembles de données dans le même système de coordonnées. Ce processus est composé de deux étapes. Dans un premier temps, le recalage lui-même

consiste à calculer, à partir des deux entrées, une IRM de cerveau et un atlas (IRM de cerveau considérée comme référence), une matrice de transformation géométrique. Dans un second temps, l'IRM de cerveau à recalcer est effectivement re-échantillonnée par application de la matrice de transformation.

2.1 Assistance lors de la conception et l'exécution de chaînes de traitement

Considérons par exemple un service de recalage. D'un point de vue syntaxique, ou technique, ce service peut être mis en œuvre à partir d'un fichier binaire exécutable consommant en entrée deux fichiers de données et produisant en sortie un fichier texte contenant la matrice de recalage. Mais d'un point de vue sémantique, le service peut mettre en œuvre un type particulier d'algorithme caractérisant la méthode d'estimation des paramètres de transformation. Le bon fonctionnement du service peut être contraint par un format particulier pour chacune de ses entrées, ou encore une modalité d'acquisition spécifique, par exemple une image ultrasonore ou acquise par résonance magnétique.

Supposons que les données à traiter et le service mis en jeu dans le processus de traitement aient été décrits au préalable avec les termes du vocabulaire de l'*ontologie de domaine*. Nous envisageons d'étendre l'environnement de conception et d'exécution de chaînes de traitement afin de réaliser, en s'appuyant sur l'*ontologie de domaine*, différentes validations garantissant une meilleure cohérence de la chaîne de traitement. Par exemple, nous souhaitons valider que la modalité d'acquisition des données à traiter est conforme à celle attendue par le service de recalage, ou de valider que le service de re-échantillonnage peut être connecté comme successeur du service de recalage via un type sémantique compatible comme par exemple une matrice de transformation affine.

D'un point de vue utilisateur, dès lors qu'un neuroscientifique a planifié son expérience sous forme de flot de calcul, et dès lors que chaque service de calcul a été décrit sémantiquement, un raisonneur ou un moteur d'inférence peut-être invoqué pour proposer une liste de services candidats compatibles pour l'étape suivante du flot. Dans la figure 1 lorsque l'utilisateur sélectionne le service de recalage au cours de la conception du flot de calcul, le système peut être interrogé pour proposer les services successeurs candidats capables de consommer la matrice produite, en exploitant les relations de subsumption entre les ports de sortie et d'entrée des services.

On peut également envisager, lors de l'exécution, que le moteur de flots de calcul vérifie l'adéquation des données poussées en entrée de chacun des services, en s'appuyant sur les méta-données décrivant, et ce service, et ces données.

2.2 Enrichissement de la base de connaissances

Couplé à un moteur d'inférences et à un entrepôt de données sémantiques, nous envisageons une extension de l'environnement de flots de calcul permettant la production et la déduction de nouveaux faits. Par exemple, dans un environnement clinique, des procédures qualité peuvent nécessiter une certaine traçabilité des données produites au cours de l'analyse. Dans notre processus de recalage (figure 1) il est intéressant de pouvoir reconstruire l'information de provenance liée au résultat du recalage. Cela permettrait de lier l'atlas qui a servi de référence pour le ré-alignement spatial aux données

recalées, ainsi que l’image source non ré-alignée. De plus nous souhaiterions pouvoir propager l’effet d’un service (et plus généralement d’une partie d’un chaîne de traitement) sur les données produites. Dans ce sens, nous voudrions automatiser la génération d’annotations signifiant par exemple “cette donnée a été recalée avec cette donnée source et cet atlas”, car dans certains cas, des outils de traitement peuvent nécessiter que les données soient positionnées dans le même repère spatial.

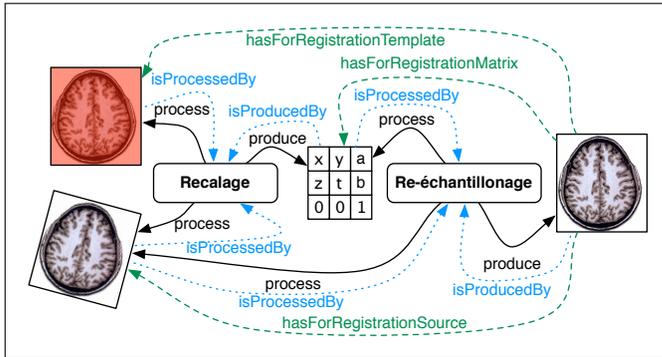


FIGURE 2 – Génération d’annotations sémantiques à l’exécution du flux de calcul.

La figure 2 illustre des relations sémantiques établies entre services de traitement et données. Les flèches pleines sont des relations créées à la volée par le moteur de flux de calcul. Elles représentent des connaissances sur la production de données. Nous souhaitons nous appuyer sur les *ontologies de domaine* et de *services* pour inférer de nouvelles connaissances. Les flèches en pointillés fins représentent des relations de provenance génériques *process/produce*, alors que les flèches en pointillés longs représentent des connaissances spécifiques au domaine, et notamment les informations sur le processus de recalage. Il faut également noter que ces informations ne sont pas uniquement dépendantes d’une invocation de service unitaire mais de l’enchaînement des deux étapes de recalage et re-échantillonnage.

3 Architecture de la plateforme

Les spécialistes du traitement de l’information et des neurosciences impliqués dans le projet NeuroLOG ont pour objectif de développer une plateforme collaborative afin de fédérer des centres de recherche en neurosciences et ainsi faciliter la mise en œuvre d’études multi-centriques. Nos objectifs sont guidés par les besoins cliniques et bénéficient du support de l’ingénierie des connaissances pour l’acquisition, la réutilisation et l’extension des connaissances liées à la pratique des neurosciences. La figure 3 illustre la plateforme NeuroLOG déployée sur deux sites partenaires, “Site A” et “Site B”. La plateforme n’impose pas de modification des bases de données spécifiques à chacun des sites. Cependant ils doivent fournir la mise en correspondance de leurs schémas spécifiques avec le schéma fédéré au travers du composant de médiation/adaptation DataFederator. La conception de l’ontologie NeuroLOG est une activité centrale dans le projet et permet de dériver le schéma relationnel utilisé comme schéma de référence global.

Ce schéma fédéré est utilisé dans la couche de médiation pour transformer et pousser les requêtes vers les bases spécifiques de chacun des sites partenaires et proposer ainsi une vue unifiée sur les résultats collectés.

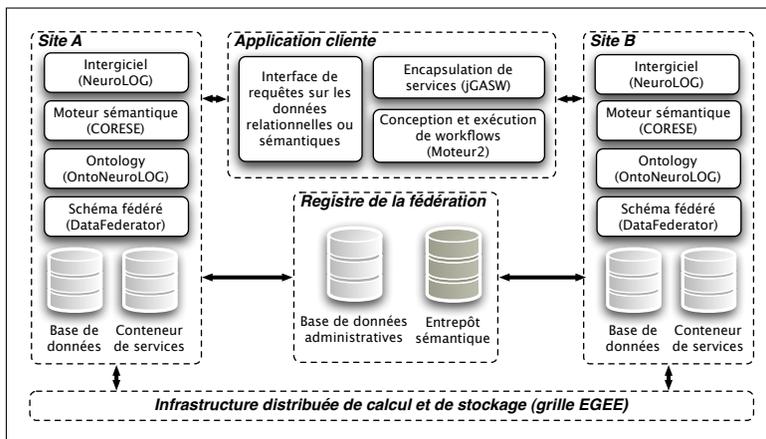


FIGURE 3 – Architecture de la plateforme pour la capitalisation de connaissances dans des études multi-centriques.

Outre le gestionnaire de données, la plateforme intègre un moteur de flots de calcul dédié à la distribution des calculs à grande échelle (infrastructures de grilles). L'utilisation de la plateforme ne nécessite pas d'adapter les outils de traitement de données mais propose *jGASW* (Java Grid Application Service Wrapper), un service d'encapsulation et d'instrumentation d'outils sous forme de services web prenant en charge l'invocation sur grille de calcul et les transferts de données nécessaires. Les procédures d'analyse de données sont conduites par le moteur de flots de calcul *Moteur2*.

Les données sémantiques sont stockées dans un entrepôt centralisé et accessibles par toute la fédération. Cet entrepôt est peuplé avec les *ontologies de domaine* et de *services* et toutes les annotations produites au cours de l'usage de la plateforme, que ce soit lors du déploiement et de l'annotation des services, ou bien lors de la génération automatique d'annotations lors de l'analyse des données. Cet entrepôt est couplé au moteur sémantique de requêtes et d'inférences *CORESE* permettant la recherche d'informations sémantiques et la production d'annotations dans l'entrepôt.

4 Représentation et acquisition des connaissances

4.1 Ontologies du projet NeuroLOG

L'*ontologie de domaine* [Temal et al.] vise à formaliser et à capitaliser les connaissances du domaine des neurosciences. Elle décrit principalement les données manipulées dans la plateforme et leurs traitements. Les concepts centraux sont les *Dataset*, *DatasetProcessing*, *Subject* et *Study*. Un effort particulier a été apporté à l'obtention

d'une description et classification précise (i) des données à traiter (*Dataset*) notamment en termes de modalité d'acquisition, d'expression (format, mode inscription dans un ou plusieurs fichiers sur disque, en mémoire), et (ii) des traitements réalisés sur ces données, par exemple la restauration de données bruitées, le recalage, ou bien la segmentation.

Une *ontologie de services*, plus technique et destinée à décrire les services web est en cours de développement. Elle vise à représenter la description du service, en terme de fonctionnalités et de paramètres, mais également à décrire son invocation afin de construire l'information de provenance.

4.2 Acquisition de connaissances sur les outils de traitement

L'annotation d'un service consiste à mettre en correspondance des concepts techniques (éléments de description du service) avec des concepts spécifiques au domaine d'application au travers d'une ontologie. Nous considérons cette activité comme un processus manuel et sensible car il nécessite une connaissance précise du domaine d'application (ontologie des neurosciences), et de la description technique du service (*ontologie des services*). En effet, l'expert responsable de l'annotation doit arbitrer parmi des choix de concepts justes mais ayant des niveaux d'abstraction différents, ce qui impactera la précision des requêtes sémantiques et donc les possibilités de raisonnement.

Afin de réaliser cette tâche d'annotation manuelle, nous proposons un prototype permettant de traduire un descripteur de service *jGASW* en une instance RDF mettant en jeu des concepts de l'*ontologie de domaine* et de *services* NeuroLOG. L'*ontologie de services* a été utilisée pour représenter les services de traitement de donnée *jGASW*. Ce prototype a été développé en Java et s'appuie sur l'interface de programmation JENA API. Le descripteur *jGASW* pouvant facilement être mis en correspondance avec les concepts de *WebService*, *Input* et *Output*, la génération de l'annotation RDF du service *jGASW* est entièrement automatisée. Une interface graphique représentant le service permet à l'utilisateur de glisser et déposer des éléments de la taxonomie *de domaine* afin de typer sémantiquement la fonctionnalité du service, ses entrées ainsi que ses sorties. Le service annoté peut par la suite être enregistré dans l'entrepôt de données sémantiques et chargé dans le moteur d'inférences.

```
jgasw:register-1.0.0
  a nlogService:web-service;
  nlogService:performs nlogDomain:affine-registration;
  nlogService:hasInput jgasw:inputImage, jgasw:inputTemplate;
  nlogService:hasOutput jgasw:registrationMatrix.
jgasw:inputImage
  a nlogService:Input, nlogDomain:T1-weighted-MR-dataset, nlogDomain:floating-
  dataset.
jgasw:inputTemplate
  a nlogService:Input, nlogDomain:reference-dataset.
jgasw:registrationMatrix
  a nlogService:Output, nlogDomain:registration-dataset.
```

FIGURE 4 – Service annoté mêlant annotations techniques et spécifiques au domaine.

La figure 4 montre une syntaxe N3 de la description sémantique d'un service de recalage annoté. Deux types sémantiques sont affectés au service : le type *nlogService:web-service* de l'ontologie de services et le type *nlogDomain:affine-registration* représentant un certain type d'algorithme de recalage dans l'ontologie de domaine. Ce service a pour paramètres d'entrée *jgasw:inputImage* et *jgasw:inputTemplate* et pour paramètre de sortie *jgasw:registrationMatrix*.

4.3 Pérennisation des connaissances acquises

Afin d'être exploitées, soit au cours de la conception d'une chaîne de traitement, soit directement au cours de son exécution, les données sémantiques doivent être stockées de manière pérenne dans un entrepôt. D'un point de vue pratique, cela consiste à créer des individus (ou instances) des classes techniques, ou spécifiques au domaine, et de les lier par des propriétés (ou relations) définies dans les ontologies. Ces instances sont décrites avec le formalisme RDF qui peut-être considéré comme un graphe dirigé avec des transitions labellisées. Chaque transition représente des liens nommés entre les ressources, elles-mêmes identifiées par des URIs.

Un premier prototype d'entrepôt de données a été implémenté avec l'interface de programmation JENA API et son support de la persistance afin de stocker et rapatrier de tels graphes au moyen d'une base de donnée relationnelle MySQL.

5 Exploitation des données sémantiques

5.1 Interrogation sémantique de l'entrepôt

L'interrogation sémantique de l'entrepôt permet d'utiliser le vocabulaire spécifique à la communauté des neurosciences et les relations liant les données sémantiques afin de rechercher les données ou outils de traitements adéquats. Nous nous appuyons sur le moteur sémantique de recherche d'information et d'inférences *CORESE* pour interroger l'entrepôt à partir du langage de requêtes SPARQL (standard du W3C permettant de rechercher des ressources RDF en spécifiant un patron de graphe).

La figure 5 illustre une requête SPARQL ayant pour objectif de chercher des couples de services potentiellement enchaînables dans un flot de calcul au travers de types d'entrée et de sortie sémantiquement compatibles. Cette contrainte de compatibilité est vérifiée ligne 7 et 8. La requête signifie que nous cherchons deux services *A* et *B* pour lesquels le port de sortie de *A* est de même type (ou sous-type) que le port d'entrée du service *B*. Une bonne maîtrise des ontologies sous-jacentes est nécessaire afin d'écrire ce type de requête. C'est pourquoi nous envisageons de proposer des requêtes pré-assemblées à l'utilisateur final via une interface graphique adaptée.

5.2 Enrichissement de l'entrepôt de données sémantiques

Bien que ce travail soit à un stade préliminaire, nous proposons un prototype permettant de reconstruire l'information de provenance des données analysées dans une chaîne de traitement. Ce prototype simule l'exécution du processus de traitement de données présenté dans la figure 1. Chaque étape du processus génère de nouvelles annotations

qui sont chargées dans le moteur d'inférences afin de déduire de nouveaux faits (ressources et propriétés RDF) via l'interface de programmation *CORESE Java API*.

```

1 select distinct ?serviceA ?serviceB where {
2   ?serviceA rdf:type nlogService:web-service .
3   ?serviceB rdf:type nlogService:web-service .
4   ?serviceA nlogService:hasOutput ?serviceA_output .
5   ?serviceB nlogService:hasInput ?serviceB_input .
6   ?serviceB_input rdf:type ?serviceB_inputType .
7   ?serviceA_output rdf:type ?serviceA_outputType .
8   ?serviceA_output rdf:type ?serviceB_inputType .
9   FILTER (?serviceA != ?serviceB)

```

FIGURE 5 – Recherche de couples de services compatibles.

Autant que possible, nous essayons de découpler le traitement sémantique des données de leur traitement effectif. Le mécanisme de chaînage avant de *CORESE* permet de déduire de nouveaux faits de manière déclarative (à partir de règles de production). Nous proposons deux règles permettant de reconstruire l'information de provenance associée aux données traitées.

```

<cos:rule>
  <cos:if>
    PREFIX neurolog: <&neurolog;>
    { ?port neurolog:produce ?data .
      ?service_invoc neurolog:involve ?data . }
  </cos:if>
  <cos:then>
    { ?data neurolog:isProducedBy ?port .
      ?data neurolog:isInvolvedIn ?service_invoc . }
  </cos:then>
</cos:rule>

```

FIGURE 6 – Génération automatique de l'information de provenance au travers d'une règle de production *CORESE*.

La figure 6 représente une règle permettant de générer des relations dirigées depuis les données produites vers le port de sortie d'un traitement et son invocation. Le bloc `<cos:if>` réalise une requête SPARQL sur la base de connaissances afin de chercher un sous-graphe liant une invocation de service, une donnée produite et le port de sortie par lequel la donnée a été effectivement produite. Pour chacun des sous-graphes résultats, le bloc `<cos:then>` réalise la création des relations *isProducedBy* entre les variables *data* et *port*, et *isInvolvedIn* entre les variables *data* et *service_invocation*. La génération de l'information de provenance pour les données d'entrée est réalisée de façon similaire.

6 Conclusion

Ce papier décrit un environnement multi-centrique dédié aux neurosciences dans lequel l'ingénierie des connaissances est appliquée pour supporter les neuroscientifiques et pérenniser les connaissances mises en jeu depuis les phases de conception des expérimentations jusqu'aux phases de réalisation.

Adossé au moteur sémantique *CORESE*, un prototype est en cours de développement. Il valide la faisabilité de l'interrogation et de l'extension d'un entrepôt de données sémantiques permettant la sélection d'outils de traitement, l'assistance dans leur composition, ainsi que la construction d'information de provenance au cours de l'exécution de chaînes de traitement. Une intégration plus fine de ce travail dans la plateforme permettra de mesurer la plus-value de ce type d'approche auprès de la communauté des neuroscientifiques. Cette intégration permettra aussi de raffiner la conceptualisation des services de traitement en neurosciences ainsi que leurs assemblages sous la forme de chaînes de traitement.

Références

- BELHAJJAME K., WOLSTENCROFT K., CORCHO O., OINN T., TANO F., WILLIAM A. & GOBLE C. A. (2008). Metadata management in the taverna workflow system. In *International Conference on Computational Science (4)* : IEEE Computer Society.
- CORBY O., DIENG-KUNTZ R. & FARON-ZUCKER C. (2004). Querying the semantic web with corese search engine. In *ECAI*, p. 705–709.
- GORDON P. M. K. & SENSEN C. W. (2008). Creating bioinformatics semantic web services from existing web services : A real-world application of sawsdl. In *ICWS*, p. 608–614.
- LEBRETON N., BLANCHET C., CHABALIER J. & DAMERON O. (2009). Utilisation d'ontologies de tâches et de domaine pour la composition semi-automatique de services web bioinformatiques. In *Journées Ouvertes Biologie, Informatique et Mathématiques (JOBIM 2009)*. in press.
- LO M. & GANDON F. (2005). *Integrating dynamic resources in corporate semantic web : an approach to enterprise application integration using semantic web services*. Rapport interne 5663, INRIA.
- LORD P., ALPER P., WROE C. & GOBLE C. (2005). Feta : A light-weight architecture for user oriented semantic service discovery. In *European Semantic Web Conference*, p. 17–31 : Springer Berlin / Heidelberg.
- MARTIN D., BURSTEIN M., MCDERMOTT D., MCILRAITH S., PAOLUCCI M., SYCARA K., MCGUINNESS D. L., SIRIN E. & SRINIVASAN N. (2007). Bringing semantics to web services with owl-s. *World Wide Web*, **10**(3), 243–277.
- MONTAGNAT J., GAINARD A., LINGRAND D., ROJAS BALDERRAMA J., COLLET P. & LAHIRE P. (2008). NeuroLOG : a community-driven middleware design. In *HealthGrid*, p. 49–58, Chicago : IOS Press.
- MOREAU L., CLIFFORD B., FREIRE J., GIL Y., GROTH P., FUTRELLE J., KWASNIKOWSKA N., MILES S., MISSIER P., MYERS J., SIMMHAN Y., STEPHAN E. & DEN BUSSCHE J. V. (2009). The open provenance model — core specification (v1.1). *Future Generation Computer Systems*.
- TEMAL L., DOJAT M., KASSEL G. & GIBAUD B. (2008). Towards an ontology for sharing medical images and regions of interest in neuroimaging. *J. of Biomedical Informatics*, **41**(5), 766–778.
- WOLSTENCROFT K., ALPER P., HULL D., WROE C., LORD P. W., STEVENS R. D. & GOBLE C. A. (2007). The mygrid ontology : bioinformatics service discovery. *IJBRA*, **3**(3), 303–325.