

# Compte rendu

## Réunion semestrielle n°7 – Projet ANR MDCO – GEONTO

**Pau le 24 juin 2011 – 9h30 – 16h30**

Présents : N. Abadie (COGIT), N. Aussenac (IRIT), M. Gaio (LIUPPA), S. Mustière (COGIT), N. Pernelle (LRI), C. Reynaud (LRI), B. Safar (LRI), C. Sallaberry (LIUPPA)

### 1. Points sur les livrables à rendre avant la fin du projet

#### T0 + 42

a) Livrable n°10 : réconciliation d'instances pour l'alignement : LRI

Le rapport a été rendu en juillet 2010, le logiciel sera livré en juillet 2011.

b) Livrable n°11 : Restructuration d'une ontologie construite automatiquement : LRI

Le module logiciel correspondant correspond à TaxoMap Framework, déjà livré.

#### Rappel des livrables à T0+46

Livrable n°13 : Intégration et accès aux schémas de bases de données : COGIT

Livrable n°14 : Indexation automatique du contenu des documents : LIUPPA

Ces deux livraisons ne devraient pas poser de problème.

Livrable n°15 : Logiciels expérimentaux, site web et rapport final. Voir discussion sur le rapport final dans ce qui suit.

### 2. Point sur l'ontologie topographique créée dans GeOnto

L'approche générale a été rappelée. Chacun des partenaires a ensuite récapitulé les travaux intermédiaires partiels réalisés durant le projet.

a) IRIT : construction d'ontologies à partir des spécifications textuelles de BD-Topo.

Cette ontologie a été livrée en juillet 2010. Une toute dernière version a été communiquée au LRI courant Juin 2011 intégrant des corrections demandées.

L'approche n'a pas été appliquée sur BD-Carto dont l'intérêt semblait moindre et du fait de la nécessité de traitements complémentaires ad-hoc.

Marion Laignelet a travaillé sur les définitions des spécifications de BD-Topo afin d'en extraire des relations (390 fragments ont été extraits). Son travail a été livré en juillet 2010.

90 fragments ont été extraits des 390 afin de les aligner, via TaxoMap, avec l'ontologie construite automatiquement et venir l'enrichir. Ce travail fait l'objet d'un stage de Master en cours. Cet enrichissement sera terminé fin août 2011.

b) LIUPPA :

L'application d'indexation développée par le LIUPPA nécessitait d'utiliser une ontologie. Le LIUPPA a travaillé sur la construction manuelle d'une ontologie de domaine utilisée dans ses travaux d'indexation. Cette ontologie a été construite en collaboration avec le COGIT. Elle s'appuie sur l'ontologie initiale du COGIT au sein de laquelle ont été intégrés à la main les termes extraits de récits de voyage plongés dans complétés par des extraits du thésaurus RAMEAU. Cette ontologie a été utilisée pour le typage des entités spatiales.

Le LIUPPA fait état également d'une autre ontologies locales (LSO) issues des spécifications de BD. ~~LSO appelée aussi~~ Toponymes.owl, mise au point par le COGIT

manuellement suite à des discussions avec l'IRIT, intègre une partie des spécifications de la BD-Topo. Cette ontologie est utilisée pour pointer la bonne table dans la bonne base de données de l'IGN et géolocaliser les entités spatiales. Ceci permet de faire le lien entre l'ontologie du domaine et la base de données. Toponymes.owl LSO ne correspond qu'à une partie de BD-Topo.

Des expérimentations ont été réalisées pour tester l'apport d'une ontologie en indexation par rapport à l'utilisation d'un lexique ou d'aucune ressource externe. Les chiffres présentés sont très bons et confirment l'idée que l'utilisation d'une ontologie géographique améliore grandement la chaîne d'indexation.

Dans ses travaux, le LIUPPA a également produit une liste de termes extraits des récits de voyage (11 livres) et des « mini-ontologies » au nombre de 118 correspondant à ces termes plongés dans RAMEAU (livrées en octobre 2010).

#### c) LRI :

Les travaux réalisés tout au long du projet sont les suivants :

Mise au point de l'outil d'alignement TaxoMap et identification de mappings à raffiner

- Alignement de Topo-Cogit et de Carto-Cogit validé par le Cogit (T0+12)
- Alignement de Carto-Cogit et de NewTopo-Cogit construite à partir de TopoCarto-Cogit, validé. (T0+12)
- Alignement de Topo-IRIT (T0+18), poursuivis en T0+24

Elaboration et test de TaxoMap Framework pour la tâche de raffinement de mapping

- Alignement de Topo-Cogit et Carto-Cogit avec raffinement des mappings en T0+24

Méthodologie d'enrichissement

- Identification de patterns d'enrichissement à partir des mappings obtenus par alignement avec Topo-IRIT en T0+30
- Identification de patterns de validation de contexte à partir des mappings obtenus par alignement des sous-graphes extraits de RAMEAU fournis par le LIUPPA avec Topo-Cogit en T0+36,
- Alignement de Yago avec TopoCarto\_Cogit. En test. Pour un enrichissement de termes anglais. T0+42

Alignement des instances de points remarquables de l'IGN avec 2 autres sources. T0+30

#### d) COGIT :

Un travail a permis de comparer les résultats d'appariement obtenus lorsqu'on dispose uniquement des schémas de bases de données et lorsqu'on dispose en plus d'une ontologie.

L'alignement a également été fait avec des schémas belges.

Suite à tous les travaux réalisés dans le projet, le COGIT souhaiterait aujourd'hui obtenir une ontologie structurée différemment. L'idée est de repartir de l'ontologie construite à la main en collaboration avec le LIUPPA (Fusion2), et de l'enrichir à partir des mappings générés par le LRI obtenus après application de TaxoMap sur Fusion2 et sur les graphes (mini-ontologies) produits par le LIUPPA provenant des récits de voyage.

Après discussion, le planning de travail suivant a été établi :

- Les concepts provenant des graphes du LIUPPA, proposés automatiquement pour l'enrichissement à l'aide de TaxoMap, seront introduits avec Protégé dans Fusion2, s'ils n'y figurent pas déjà : à faire par le LRI pour début juillet. La version de l'ontologie obtenue sera alors Fusion3.

- Fusion3 sera transmise à l'IRIT qui l'enrichira avec les fragments extraits des définitions de BD-TOPO, issus du travail de M. Laignelet. Le résultat, Fusion4, sera livré par l'IRIT fin août 2011.

- Les mappings de type « grotte marine » is-a « grotte » seront communiqués par le LRI au COGIT. Les termes correspondant aux concepts considérés comme plus spécifiques seront listés. Le COGIT introduira des propriétés au bon niveau de l'ontologie pour prendre en compte ces concepts plus spécifiques.

- Le LRI aligne en septembre Fusion4 avec Topo-IRIT, issue de l'application des traitements automatiques de langage naturel sur les spécifications de BD-TOPO.

### **3. Rapport final :**

Il a été décidé de demander des précisions sur les pièces justificatives à fournir avec les relevés de dépenses fin octobre.

Contenu du rapport final : les membres présents s'accordent pour faire un rapport récapitulatif de l'ensemble des travaux réalisés et joignant en annexe toutes les publications.

Le plan suivi prendra appui sur le plan du papier de la revue RIG. Ce plan sera adapté pour que soient ajoutés :

- Les travaux en réconciliation d'instances
- Les travaux sur la comparaison d'ontologies
- Les travaux sur la mise au point d'un moteur d'accès aux schémas de bases de données.
- Une description de l'ontologie produite et de comment cette ontologie a pu être produite.

Calendrier :

- Plan diffusé à l'ensemble des membres du projet : effectué par Chantal Reynaud dès que possible
- Chaque partenaire propose un plan détaillé des parties le concernant pour le 14/07/2011
- 1<sup>ère</sup> version du rapport final demandée pour fin septembre.

### **4. Questions diverses :**

- Fiche ANR demandée pour figurer dans les cahiers de l'ANR.

Les résultats majeurs du projet sont demandés. Les membres présents s'accordent pour citer : les outils mis au point, les applications intégrant une ontologie topographique, l'ontologie elle-même qui sera livrée en octobre 2011.

- Publications :

De nouvelles publications qui ne sont pas le site web de Geonto ont été acceptées :

1 chapitre de livre suite à une publication à EGC 2011 : COGIT

LGC 2011 (Lexique Grammaire Colloque) : LIUPPA

Atelier RTE (Raisonnement Temps Espace) associé à la plate-forme de l'AFIA : Publication et conférence invitée (Christian Sallaberry) – LIUPPA

C. Sallaberry,

*Conférence invitée sur le thème : Indexation et recherche d'information géographique dans des contenus textuels à connotation « territoriale »*,

RTE'11 – Atelier Raisonnement Temps Espace, Réseau AFIA, Chambéry, France, 16 mai 2011

TALN 2011, poster : IRIT  
ICC International Cartographic Conférence : COGIT

Tous les auteurs doivent me transmettre les références de toutes ces publications pour qu'elles apparaissent sur le site Web du projet.

- Publication à venir :

Prévoir une publication, ou plusieurs, en anglais à partir du rapport final. L'une pourrait mettre l'accent sur l'aspect interopérabilité, une autre sur l'aspect applicatif. Penser à intégrer l'ensemble des membres du projet.

Quelques noms de revues ont été cités : IJHCS, Transactions of KDE, Applied Ontologies (questions théoriques mais intérêt également pour le contenu des ontologies)

Noms de conférence : FOIS : demandeur d'applications utilisant les ontologies

- Mise à disposition de l'ontologie produite :

Il est possible d'utiliser un site du LIRMM pour mettre à disposition l'ontologie (= site de publication des ontologies de DataLift)

Elle sera aussi publiée sur le site du projet GeOnto et du COGIT.

Le travail réalisé par N. Abadie (COGIT) portant sur l'intégration de schémas de bases de données géographiques sera mis en OpenSource dans OpenSource Géoxygène dès qu'elle aura soutenu.

Rq : Seules les ontologies seront publiées, pas les données, même si un travail sur la construction d'un moteur d'accès aux données géographiques a été réalisé dans le cadre du projet.

- Lien avec des partenaires étrangers :

La coopération avec les anglais n'a pu être possible mais l'IGN a néanmoins des contacts avec le groupe visé.

Les espagnols sont demandeurs de contacts. Une 1<sup>ère</sup> tentative a été faite lors de l'organisation de l'atelier OntoGeo à Toulouse en nov. 2010.

- Points à signaler dans le rapport final :

L'embauche de F. Hamdi sur un poste de post-doctorant au COGIT dès qu'il aura soutenu sa thèse

Le contact établi avec Oscar Corcho de Madrid à OntoGeo

Deux contacts pris avec des entreprises ayant donné lieu à des réponses communes à des appels à projet.

L'organisation de l'atelier OntoGeo ayant eu un franc succès.

- Perspectives :

Rendre l'ontologie conforme aux critères « Open Linked Data ». Un tutoriel sera organisé en septembre par F. Gandon et F. Scharffe à Montpellier sur ce thème.

Fin de la réunion à 16h30.