



HAL
open science

Construction collaborative d'une Ressource Termino-Ontologique (RTO) pour le droit des collectivités territoriales

Ourdia Ressad-Bouidghaghen, Sylvie Szulman, Haïfa Zargayouna, Eve Paul

► To cite this version:

Ourdia Ressad-Bouidghaghen, Sylvie Szulman, Haïfa Zargayouna, Eve Paul. Construction collaborative d'une Ressource Termino-Ontologique (RTO) pour le droit des collectivités territoriales. IC - 24èmes Journées francophones d'Ingénierie des Connaissances, Jul 2013, Lille, France. hal-00860104

HAL Id: hal-00860104

<https://hal.science/hal-00860104>

Submitted on 12 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Construction collaborative d'une Ressource Termino-Ontologique (RTO) pour le droit des collectivités territoriales

Ourdia Ressad-Bouidghaghen¹, Sylvie Szulman¹, Haïfa
Zargayouna¹, Eve Paul²

¹ Université Paris 13, Sorbonne Paris Cité, LIPN, CNRS, (UMR 7030), F-93430, Villetaneuse,
France.

prénom.nom@lipn.univ-paris13.fr

² Groupe Victoires-Editions 38, rue Croix-des-Petits-Champs, 75038 PARIS cedex 01
e.paul@victoires-editions.fr

Résumé :

La construction collaborative de ressources sémantiques est devenue un domaine de recherche en plein essor et une solution pratique notamment dans les grands projets de construction à grande échelle. En effet, contrairement à une construction centralisée, une construction collaborative permet de tracer les différents points de vue et de mettre en place une modélisation consensuelle. De plus, la construction collaborative permet de faire communiquer des acteurs aux profils différents dans une approche distribuée.

Nous présentons dans ce papier un début de mise en place d'une construction collaborative d'une Ressource Termino-ontologique (RTO) dans le cadre du projet LégiLocal. Ce projet vise à proposer des fonctionnalités d'accès à l'information juridique locale pour les citoyens, les élus et les personnels de mairie. Dans ce projet, une RTO est nécessaire pour permettre un accès sémantique ciblé qui prend en compte les aspects de réglementation locale en rapport avec le droit. La construction de cette RTO nécessite la collaboration de plusieurs acteurs du projet, avec des profils différents : des ingénieurs de la connaissance, des terminologues et des juristes (spécialistes en droit des collectivités). Nous présentons les deux scénarios de collaboration que nous mettons en place qui permettent à la fois une couverture maximale du domaine et l'obtention d'un consensus et la traçabilité des décisions.¹

Mots-clés : Construction de Ressource Termino-Ontologique, Construction Collaborative, Droit des Collectivités Territoriales.

1. Ce travail a été en partie financé par le projet FUI LégiLocal.

1 Introduction

La construction de ressources sémantiques a longtemps reposé sur une pratique centralisée (Gomez-Perez *et al.*, 2003). Dès les années 2000, l'approche collaborative a été désignée comme une démarche clé pour la construction de connaissances consensuelles autour d'un domaine (Holsapple & Joshi, 2002), notamment dans la construction d'une ressource à grande échelle qui nécessite plusieurs contributeurs. Des méthodes et des formalismes ont été proposés qui tiennent compte de la richesse des points de vue et qui permettent de converger vers un consensus.

La construction collaborative peut concerner différentes étapes du cycle de vie de construction d'une ressource : spécification, mise en place, évaluation, maintenance et évolution. Pour chacune des étapes, différents scénarios peuvent être mis en place. Mais quel que soit le scénario choisi, la cohérence globale de la ressource, la traçabilité des différents choix et la prise en compte du rôle de chaque contributeur sont des points cruciaux de la réussite du projet.

Dans ce papier nous nous intéressons à la construction d'une ontologie et d'une terminologie liées que nous désignons par ressource termino-ontologique (RTO). Cette construction a pour cadre le projet LégiLocal qui vise à proposer des fonctionnalités d'accès à l'information juridique locale pour les citoyens, les élus et les personnels de mairie. La ressource sert à annoter des documents pour pouvoir répondre à des questions relatives à l'activité juridique d'une commune. Dans le cadre du projet de recherche, cette activité juridique a été restreinte aux problèmes communaux induits par l'activité de randonnée pédestre. Ainsi un secrétaire de mairie peut rechercher des documents lui permettant de répondre à la question suivante : « un quad peut-il circuler sur le GR ®² traversant la commune ? ».

La ressource est construite à partir de textes et nécessite la collaboration de plusieurs acteurs du projet, avec des profils différents : des ingénieurs de la connaissance, des terminologues et des juristes (spécialistes en droit des collectivités).

Nous présentons dans la section 2 un état de l'art des méthodes existantes pour la construction collaborative de ressources sémantiques. En section 3, nous décrivons les deux scénarios de collaboration que nous mettons en place pour la construction de cette RTO.

2. sentier de Grande Randonnée.

2 Méthodes pour la mise en place d'une approche collaborative

Les enjeux de la mise en place d'une démarche collaborative de construction de ressource sémantique sont importants. Ils découlent de l'objectif même de la construction collaborative de la ressource sémantique, à savoir assurer la couverture de la ressource en impliquant un grand nombre de collaborateurs et la construction de connaissances consensuelles autour d'un domaine. En effet, cela nécessite d'orchestrer le travail des collaborateurs, de garder le contrôle sur le processus du développement qui devient distribué, d'assurer la qualité de la ressource développée et de résoudre des conflits de conception qui peuvent survenir.

Pour mieux orchestrer les efforts de collaboration, des modes de collaboration sont définis. On distingue principalement le mode de collaboration parallèle et le mode de collaboration mutuelle. Le choix d'un mode de collaboration ou d'un autre est lié principalement à la complexité du domaine (qui peut couvrir plusieurs secteurs ou sous-domaines) et aux compétences et profils des acteurs intervenants. Dans le mode de collaboration parallèle (Sure *et al.*, 2002; Vrandecic *et al.*, 2005; Kotis & Vouros, 2006; Tudorache *et al.*, 2010), les collaborateurs sont affectés à des sous-domaines de la ressource. La mise en place de ce scénario suppose des collaborateurs avec des compétences et des connaissances complémentaires, et nécessite une phase d'intégration dans le produit final à travers des processus d'intégration (alignement et/ou fusion). En effet, des concepts peuvent être communs à plusieurs sous-domaines ou avoir des relations avec des concepts appartenant à d'autres sous-domaines. Dans le mode de collaboration mutuelle (Holsapple & Joshi, 2002; Karapiperis & Apostolou, 2006; Mangeot-Lerebours *et al.*, 2003; Siorpaes *et al.*, 2008), les collaborateurs travaillent sur une même ressource. Chaque proposition de conception est soumise à un processus de validation avant son intégration à la ressource finale.

La mise en place d'une méthode de construction collaborative nécessite dans un premier lieu de définir le rôle de chaque acteur qui peut avoir le statut de contributeur ou de collaborateur selon le processus de contrôle et de décision mis en place. En effet, selon Holsapple & Joshi (2002), la réussite d'une méthode de collaboration dépend des compétences mises en place pour la supervision du processus de collaboration. Comme exemple de ces travaux on peut citer (Holsapple & Joshi, 2002; Karapiperis & Apostolou, 2006), où un médiateur est défini pour conduire le processus itératif de construction de l'ontologie. Les méthodes (Tudorache *et al.*, 2010; Vrandecic *et al.*, 2005) proposent un comité de pilotage pour la gestion des

mises à jour de la ressource. Les travaux qui reposent sur la myriadisation (*crowdsourcing*), tels que (Mangeot-Lerebours *et al.*, 2003; Siorpaes *et al.*, 2008), ne prévoient pas de comité de pilotage ou de supervision.

La participation de plusieurs acteurs dans la construction collaborative de la ressource sémantique peut induire des conflits de conception. Plusieurs types de conflits peuvent survenir, aussi bien au niveau terminologique (des choix différents de termes à intégrer ou du niveau de spécialisation d'un terme et de son appartenance à un sous-domaine), mais aussi au niveau conceptuel avec des conflits sémantiques entre définitions de concepts (des acteurs qui utilisent différentes représentations ou interprétations d'un même concept). Pour permettre d'atteindre un consensus lorsqu'il y a un conflit ou une divergence entre les collaborateurs ou lors de la validation des contributions, différentes solutions sont proposées. Ces solutions proposent l'obtention du consensus via une évaluation par les pairs (Tudorache *et al.*, 2010), la mise en place de votes (Mangeot-Lerebours *et al.*, 2003) dans un processus itératif jusqu'à l'obtention d'un consensus. Des méthodes de gestion de projet peuvent être pertinentes, la méthode Delphi³ (Linstone & Turoff, 1975) est expérimentée dans (Holsapple & Joshi, 2002). (Karapiperis & Apostolou, 2006) emploie la Technique du Groupe Nominal (NGT)⁴ (Delbecq & VandeVen, 1971) pour parvenir à un consensus lors de la validation des propositions des contributeurs. D'autres travaux mettent en place des méthodes d'argumentation, la méthode IBIS (Kunz & Rittel, 1970) a été utilisée dans (Vrandečić *et al.*, 2005; Kotis & Vouros, 2006) pour modéliser le processus d'argumentation lors des discussions des propositions de conception des collaborateurs pour atteindre un consensus.

Notre travail s'intéresse à décrire des scénarios de collaboration qui s'inspirent de l'état de l'art adaptés aux acteurs du projet et à son périmètre. Ils mettent en jeu à la fois une collaboration parallèle et mutuelle. Ils permettent à des ingénieurs de la connaissance, des terminologues et des juristes (spécialistes en droit des collectivités) de travailler en concertation à la construction d'une RTO. Dans un premier temps, le résultat des échanges et les décisions prises sont consignés dans un guide qui est mis à jour à chaque itération. Même si l'organisation est faiblement hiérarchisée,

3. La méthode Delphi est une méthode visant à organiser la consultation d'experts sur un sujet précis.

4. La technique du groupe nominal (The nominal group technique-NGT) est une méthode de prise de décision dans un groupe d'experts prenant en compte les points de vues de toutes les personnes dans le groupe.

un responsable est désigné pour la validation finale des décisions. La suite du papier décrit ces scénarios.

3 Scénarios de collaboration dans LégiLocal

Dans le cadre du projet LégiLocal, une ontologie et une terminologie liées sont à construire. L'ontologie doit assurer une couverture conceptuelle du projet, la terminologie doit permettre d'une part de structurer un vocabulaire permettant une annotation sémantique des documents, d'autre part d'aider à la construction de l'ontologie. La terminologie est liée à l'ontologie afin d'assurer un vocabulaire commun.

La construction de ces deux ressources se font à la fois à partir de textes en suivant la méthode Terminae (Aussenac-Gilles *et al.*, 2008) et à partir des experts métiers. Les corpus ont été constitués par l'éditeur juridique.

Deux scénarios différents ont été élaborés. Le domaine de l'application étant vaste, un scénario « global » dans lequel un découpage en modules nécessite un mode de collaboration parallèle pour construire la ressource. Pour chaque module, un scénario de collaboration mutuelle a été mis en place permettant de prendre en compte toutes les compétences (voir figure 1).

3.1 Construction de l'ontologie

Un découpage en modules permet de définir l'ossature de l'ontologie. Il est à la fois guidé par les expériences acquises par l'éditeur juridique et des fonctionnalités sémantiques liées à l'application. Un premier découpage a permis l'identification de trois modules relatifs à des sous-domaines :

- un module Organisation permettant de décrire l'organisation de l'état français (ministères, préfectures,...) et les juridictions comme le conseil d'état, la cour de cassation... ;
- un module Droit permettant de décrire les notions relatives au droit français que l'on restreindra au droit de la voirie et au droit communal dans un premier temps ;
- un module Randonnée permettant de décrire les principales notions propres à la pratique de la randonnée du point de vue de la collectivité territoriale.

Trois modules supplémentaires sont identifiés pour permettre de répondre aux questions des utilisateurs. Ces modules sont relatifs aux types de documents à annoter ainsi qu'aux connaissances territoriales et tem-

poelles, utiles pour identifier des documents produits par des communes voisines ou pour identifier l'antériorité des décisions juridiques :

- un module Document décrivant les différents types de documents juridiques (arrêtés, circulaires, ...).
- un module Territorial qui est construit à l'aide d'ontologies déjà existantes comme celle de l'Insee⁵.
- un module Temporel décrivant les éléments de datation nécessaires à la mise en relation des informations juridiques.

Le découpage de l'ontologie permet une construction en parallèle des différents modules. Il nécessite en amont une identification du périmètre de chaque module qui se traduit par la constitution d'un corpus qui guide le processus de construction. La figure 1(a) montre le scénario adopté pour les trois premiers modules. Le scénario propose trois étapes : 1) une identification des périmètres, 2) une construction de modules et 3) une intégration. Chaque module est relatif à un corpus et est la responsabilité d'un acteur (ici un ontologue de l'éditeur juridique) qui en assure la cohérence et qui supervise le processus de construction. La phase d'intégration réunit les différents acteurs qui doivent valider en commun les choix de modélisation et s'assurer de la cohérence globale de la RTO.

Le scénario global débute par une parallélisation des tâches de construction de chaque module, puis est suivi (une fois les modules construits) par une étape d'intégration afin d'obtenir une ontologie cohérente.

La phase d'intégration permet une synchronisation des différents modules et pose la question épineuse de prise en compte des points de vue. En effet, un terme peut avoir plusieurs sens suivant le sous-domaine ou point de vue. Ainsi le terme de *chemin rural* a plusieurs sens. Il est représenté par un objet physique dans le module randonnée et est modélisé par un concept CHEMIN_RURAL sous OBJET PHYSIQUE. Mais la définition suivante a été trouvée : « *Le chemin rural est avant tout une voie privée communale affectée à la desserte des riverains et ouverte au public fg.* Le sens du terme *chemin rural* exprimé par cette définition peut être restreint à une VOIE PRIVÉE COMMUNALE OUVERTE AU PUBLIC et est modélisé par le concept CHEMIN_RURAL comme une voie privée communale ayant un régime juridique public dans le module Droit.

Ce problème classique pour l'intégration d'ontologies se pose de manière accrue dans le domaine juridique car certains termes de la langue générale ont une signification précise et différente dans le domaine juridique. Une première phase d'alignement lexical permet de détecter les

5. <http://rdf.insee.fr/geo/>

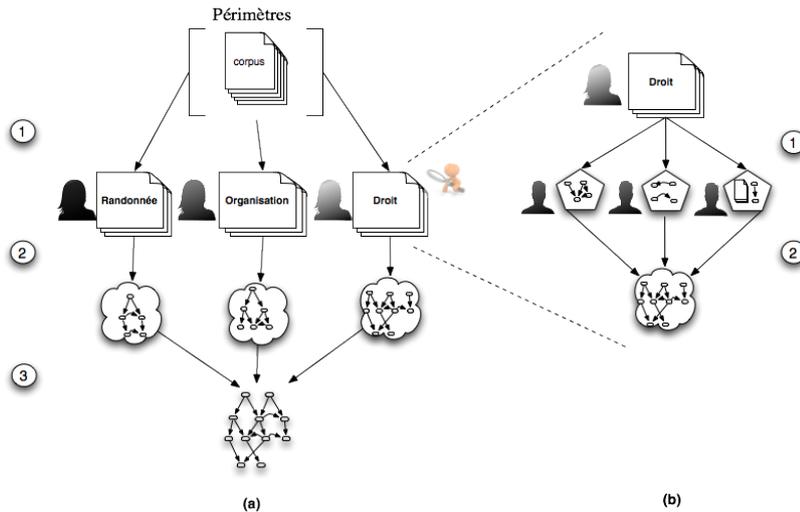


FIGURE 1 – Scénarios de collaboration pour la construction de la RTO dans LégiLocal

éléments communs dans les différents modules.

Les choix d'intégration doivent être effectués manuellement. Dans certains cas l'identification de points de vues différents permet d'harmoniser la modélisation et d'adopter une conception unique. Dans d'autres cas plus difficiles, il s'agit de garder les différents points de vue. Il faut donc s'assurer lors de l'intégration de la cohérence des définitions des concepts.

Cette intégration n'a pas encore été discutée. Ainsi pour l'exemple ci-dessus les choix sont les suivants :

- Si l'intégration est une fusion, un seul concept CHEMIN_RURAL doit exister. Il faudra choisir entre une hiérarchie multiple (le concept CHEMIN_RURAL a deux pères) ou définir deux concepts CHEMIN_RURAL_PHYSIQUE et CHEMIN_RURAL_JURIDIQUE éventuellement liés par des relations conceptuelles APOUROBJETPHYSIQUE et APOUROBJETJURIDIQUE
- Si l'intégration est un alignement, il faudra définir des liens entre les deux concepts CHEMIN_RURAL.

Ce travail étant en cours, certains modules sont plus avancés que d'autres. Ainsi le module Organisation est construit à partir de ressources appartenant à un partenaire. Il a été demandé aux responsables des modules d'ex-

pliciter les liens entre modules détectés lors de la construction. Ainsi lors de l'élaboration du module Droit, il a été explicité qu'une notion juridique comme la responsabilité est actée par un document juridique. Il y a dans ce cas un lien évident entre le module Droit et le module Document. Ces liens explicites permettent de faciliter l'intégration des éléments qui n'auraient pas été détectés par un alignement lexical.

3.2 Construction de modules termino-ontologiques

La construction de chaque module suit elle même un scénario de collaboration. Si pour la construction de la termino-ontologie, le scénario de collaboration permet de paralléliser le processus de construction pour une couverture maximale, le but pour la construction des modules est de s'assurer de l'obtention d'un consensus et de la traçabilité des décisions. Les différents choix portent sur les termes du domaine ainsi la conception adoptée pour le module. Le responsable du module s'assure de la cohérence au sein des modules et arbitre les décisions en cas de conflit. Le scénario de construction est en cours d'expérimentation et porte actuellement sur le volet terminologique. La figure 1(b) schématise ce scénario pour le module Droit. Le scénario met l'accent sur deux étapes : 1) une étape de construction par chaque contributeur et 2) une étape de mise en commun.

Les premières expériences ont permis de détecter ces situations classiques : (i) convergence : accord sur les candidats termes sélectionnés, (ii) divergence : désaccord sur la forme canonique d'un terme, ces cas de désaccord sont aussi liés à la taille des candidats termes. En effet, certains terminologues favorisent la forme longue et d'autres la forme courte. Le responsable du module initie le dialogue et tient à jour un guide qui recense les cas détectés et décrit les choix adoptés. Ces choix sont liés à l'application et nécessitent d'avoir une vision globale de la termino-ontologie. Pour illustrer le propos, soit l'expression *avant-projet de budget*. Dans la liste des termes, il existe le terme *avant-projet* et le terme *budget*. Une terminologue a proposé le terme *avant-projet de budget*. L'autre terminologue a estimé que ce n'est pas un terme mais l'association de deux termes car dans de nombreuses procédures juridiques, un avant-projet est nécessaire et constitutif d'une étape obligatoire pour la validité de la procédure. Il s'agit donc de décider s'il faut lister tous les termes possibles ou organiser la terminologie pour permettre la combinaison de ce terme et d'autres termes.

L'organisation de la terminologie doit prendre en compte l'usage pour

lequel elle est créée. Pour LégiLocal, la terminologie est utile pour l'annotation de documents. Le choix des termes pertinents est influencé par l'exploitation des annotations par le moteur de recherche. La possibilité de combiner des termes simples comme *avant-projet* et *budget*, *vente* permet de retrouver de nouvelles formes, ce qui est potentiellement plus exhaustif et permet également de détecter des néologismes, comme un *avant-projet d'une nouvelle procédure juridique*. Le fait d'annoter avec des termes composés permet de diminuer le nombre de documents retournés et d'être plus précis dans la réponse à une requête.

Le corpus utilisé pour créer le module Randonnée a donné naissance à un millier de termes décrits dans des fiches. Ces termes sont des termes simples et composés. Des relations terminologiques doivent être encore définies. La modélisation ontologique est en cours et demandera certainement beaucoup plus de temps.

4 Conclusion

Nous avons présenté dans ce papier, un début de mise en place d'une construction collaborative d'une RTO. La collaboration est nécessaire car le domaine à couvrir est vaste et différents contributeurs aux compétences complémentaires doivent y travailler. Les premières réunions ont conduit à la mise en place de deux scénarios différents. Nous travaillons également sur l'implémentation de *plugins* pour ajouter à l'outil Terminae un outillage de la collaboration. Le travail se poursuit et nous permettra à terme d'évaluer la méthode de construction ainsi que la ressource construite dans le cadre du projet.

Références

- AUSSENAC-GILLES N., DESPRES S. & SZULMAN S. (2008). The Terminae Method and Platform for Ontology Engineering from Texts. In *Bridging the Gap between Text and Knowledge - Selected Contributions to Ontology Learning and Population from Text*, p. 199–223. IOS press.
- DELBEQ A. & VANDEVEN A. (1971). A group process model for problem identification and program planning. *Journal Of Applied Behavioral Science* VII, 7(4), 466–492.
- GOMEZ-PEREZ A., FERNANDEZ-LOPEZ M. & CORCHO O. (2003). *Ontological Engineering*. Springer-Verlag.

- HOLSAPPLE C. & JOSHI K. (2002). A collaborative approach to ontology design. *Communications of the ACM - Ontology : different ways of representing the same concept*, **45**(2), 42–47.
- KARAPIPERIS S. & APOSTOLOU D. (2006). Consensus building in collaborative ontology engineering processes. *Journal of Universal Knowledge Management*, **1**(3), 199–216.
- KOTIS K. & VOURO S A. (2006). Human-centered ontology engineering : The HCOME methodology. *Knowledge and Information Systems*, **10**(1), 109–131.
- KUNZ W. & RITTEL H. (1970). *Issues as elements of information systems*. Working paper n.131, Institut für Grundlagen Der Planung I.A, Universität Stuttgart, <http://www.cc.gatech.edu/ellendo/rittel/rittel-issues.pdf>.
- LINSTONE H. A. & TUROFF M. (1975). *The Delphi Method : Techniques and Applications*. Linstone & Turoff.
- MANGEOT-LEREBOURS M., SERASSET G. & LAFOURCADE M. (2003). Construction collaborative d'une base lexicale multilingue- le projet Papillon. *TAL*, **44**(2), 151–176.
- SIORPAES K., HEPP M., KLOTZ A. & HACKL M. (2008). *myOntology : Tapping the Wisdom of Crowds for Building Ontologies*. Technical report, STI-Semantic Technology Institute Innsbruck, University of Innsbruck, Austria, <http://www.cc.gatech.edu/ellendo/rittel/rittel-issues.pdf>.
- SURE Y., ERDMANN M., ANGELE J., STAAB S., STUDER R. & WENKE D. (2002). OntoEdit : Collaborative Ontology Development for the Semantic Web. In *ISWC'02 Proceedings of the First Int'l Semantic Web Conference on The Semantic Web*, p. 221–235.
- TUDORACHE T., FALCONER S., NYULAS C., STOREY M., USTUN T. & MUSEN M. (2010). Supporting the collaborative authoring of ICD-11 with web-protege. In *Proceedings of the 2010 AMIA- American Medical Informatics Association Annual Symposium*, p. 802–806.
- VRANDICIC D., VR D., PINTO S., SURE Y. & TEMPICH C. (2005). The DILIGENT knowledge process. *Journal of Knowledge Management*, **9**(5), 85–96.