

## Projet ANR-11-BS02-007

# Typex

Programme Blanc 2011

A IDENTIFICATION	1
B LIVRABLES ET JALONS	2
C RAPPORT D'AVANCEMENT	2
D VALORISATION ET IMPACT DU PROJET DEPUIS LE DÉBUT	5
E ANNEXES ÉVENTUELLES	9

## A IDENTIFICATION

Acronyme du projet	Typex
Titre du projet	Typeful certified XML: integrating language, logic, and dataoriented best practices
Coordinateur du projet (société/organisme)	Giuseppe Castagna (CNRS - Université Paris Diderot)
Date de début du projet	6/1/2012
Date de fin du projet (conventions)	5/1/2015
Labels et correspondants des pôles de compétitivité (pôle, nom et courriel du corresp.)	Pôle de compétitivité Systematic Laurent Séguin <a href="mailto:l.seguin@systematic-paris-region.org">l.seguin@systematic-paris-region.org</a>
Site web du projet, le cas échéant	<a href="http://typex.lri.fr">http://typex.lri.fr</a>

Rédacteur de ce rapport	
Civilité, prénom, nom	M. Giuseppe Castagna
Téléphone	+33157279340
Courriel	<a href="mailto:Giuseppe.Castagna@univ-paris-diderot.fr">Giuseppe.Castagna@univ-paris-diderot.fr</a>
Date de rédaction	20 juillet 2013

## B LIVRABLES ET JALONS

Consulter la page <http://typex.lri.fr/deliverables.html> pour plus de détails

N°	Intitulé	Nature*	Date de fourniture			Partenaires (souligner le responsable)
			Prévue initialement	Replanifiée	Livrée	
D.2.a	Constraint generation for inferring types in polymorphic function application with semantic subtyping	Rapport	M12		<b>M12</b>	LRI PPS
D.2.b	Type synthesis for the logical solver	Rapport	M12		<b>M12</b>	LRI, PPS, WAM

N°	Intitulé	Nature*	Date de fourniture			Partenaires (souligner le responsable)
			Prévue initialement	Replanifiée	Livrée	
D.2.c	Resolution of sets of constraints by means of an extended logical solver	Rapport	M21		Part1 <b>M18</b> Part 2 <b>M18</b>	<u>WAM</u>
D.2.d	Design and implementation of the extension of CDuce with polymorphic functions	Rapport	M30		Part 1 <b>M18</b>	<u>LRI</u> , PPS
D.3.a	SXSI with backward axes	Prototype	M21		<b>M18</b>	<u>LRI</u>
D.3.b	Coq libraries for XPath and alternating tree automata manipulation	Prototype	M15	<b>M22</b>		<u>LRI</u> , PPS, <u>WAM</u>
D.3.c	Formal proof of SXSI's core	Rapport	M24			PPS, <u>LRI</u>
D.3.d	Automated proofs for the solver		M30			<u>WAM</u>
D.4.a	XQuery equipped with static type system	Rapport	M24		Part1 <b>M18</b> Part 2 <b>M18</b>	<u>LRI</u> PPS
D.4.b	XQuery equipped with semantic annotations	Rapport	M30		Part 1 <b>M9</b>	<u>LRI</u> , PPS, <u>WAM</u>
D.4.c	XQuery engine with static type system and logical assertions	Prototype	M36			<u>LRI</u> , PPS <u>WAM</u>
D.5.a	Proposal for a standard XML transformation language with pattern matching, poly-morphism, assertions, precise type check-ing, efficient certified implementation	Rapport	M36			<u>LRI</u> , PPS, <u>WAM</u>
D.6.a	Formalization of NoSQL Languages	Rapport	<b>new</b>		<b>M12</b>	<u>LRI</u> PPS
M.2.i	Polymorphic types for languages and solvers.	Jalon	M15		<b>M12</b>	
M.2.ii	Polymorphic functions application	Jalon	M24		<b>M18</b>	
M.2.iii	Polymorphic CDuce	Jalon	M36			
M.3.i	Handling of backward axes within SXSI	Jalon	M21			
M.3.ii	SXSI's core functionality is verified.	Jalon	M24			
M.3.iii	SXSI with backward axes is verified	Jalon	M36			
M.3.iv	The algorithm of the solver is verified	Jalon	M30			
M.4.i	A functional Core for XQuery 3.0.	Jalon	M12		<b>M14</b>	
M.4.ii	Type system for XQuery with higher-order functions and pattern matching.	Jalon	M24		<b>M14</b>	
M.4.iii	Statically verified XQuery assertions.	Jalon	M21			
M.5.i	Fundamentals of XQuery 4.0	Jalon	M36			

## C RAPPORT D'AVANCEMENT

### C.1 OBJECTIFS INITIAUX DU PROJET

Le but du projet est de combiner trois approches différentes (logique et solveur, programmation fonctionnelle, langages de manipulation de données) dans un cadre formellement certifié (notamment en Coq) pour définir un cadre fondationnel, des techniques et des outils, permettant de définir, d'analyser statiquement, d'implémenter efficacement et d'optimiser des transformations de documents semi-structurés, en particulier en format XML. Chacun des trois partenaires possède des compétences dans une de ces trois approches et le but du projet est de mutualiser ces expériences et connaissances pour d'abord améliorer par « cross-fertilisation » chaque domaine spécifique pour ensuite produire une approche unique/unifiée. La première partie du projet a été donc dédiée à la fertilisation mutuelle des approches.

Le projet est structuré en 5 « Task ». Le Task 1 est organisationnel tandis que le Task 5 concerne l'intégration des résultats des trois Task qui le précèdent et il est donc prévu pour la fin du projet. Ce rapport donc concerne les Tasks 2, 3 et 4 plus un task 6 supplémentaire ajouté pour ce rapport.

Le Task 2 est dédié à la définition et l'implémentation du polymorphisme pour une plate-forme XML. En particulier il vise à étudier la génération et la résolution d'ensembles de contraintes pour l'inférence de types polymorphes. Le Task 3 est dédié à l'extension et à la vérification formelle des outils principaux développés par le projet, c'est-à-dire le moteur de requêtes SXSI et le solveur du  $\mu$ -calcul. Le Task 4 est dédié au développement d'analyses statiques pour XQuery et en particulier pour la vérification des types et l'ajout et la vérification d'assertions logiques. Le Task 6 regroupé de nouveaux axes de recherche non prévu par la proposition (qui désormais date de il y a deux ans et demi).

## C.2 TRAVAUX EFFECTUÉS ET RÉSULTATS ATTEINTS SUR LA PÉRIODE CONCERNÉE

Nous énumérons les travaux et résultats par Task (cf. Section C.1 et Document Scientifique de la proposition disponible sur le site Typex : [http://typex.lri.fr/files/typex\\_proposal.pdf](http://typex.lri.fr/files/typex_proposal.pdf))

### **Task 2 : Polymorphism and solvers.**

Le Task 2 est dans un état très avancé par rapport à la planification et a donné des résultats au delà de tout espoir. Rappelons qu'il était prévu de développer ce task en quatre phases, dont les deux premières devaient être développées en parallèle.

La première phase consistait en la définition d'un calcul polymorphe explicitement typé (avec types récursifs, flèches, unions, intersections et négation) et en son étude pour la génération de contraintes de typage permettant l'inférence locale de types pour l'application de fonctions polymorphes. La définition du calcul s'est avérée plus difficile que prévu (elle a demandé la définition d'un calcul explicitement typé pour types intersections, un problème étudié depuis plus de 20 ans et jamais résolu de manière satisfaisante) : il est présenté dans [CNXILP13]. La génération de contraintes est présentée dans [CNX13].

La deuxième phase devait implémenter la synthèse de type polymorphes. Pour cela il était nécessaire de définir et de représenter dans le solveur une algèbre de types d'ordre supérieur. Puisque son achèvement était crucial pour permettre la fertilisation mutuelle des approches cela a été la tâche que nous avons attaquée dès le début : nous avons ajouté un opérateur flèche nativement dans le solveur pour supporter les fonctions ainsi que les variables de types, ce qui a donné l'article en cours de soumission [GGL13] et l'implémentation complète avec résolution logique [GLS13a].

Côté synthèse de types, un travail initial à partir du cas monomorphe sans fonctions ni variables a été effectué avec succès. Le solveur logique du projet WAM recherche pour une formule donnée un arbre fini pour lequel la formule est satisfaite. La synthèse de types dans le cas monomorphe consiste à calculer tous les arbres qui satisfont la formule. Une des difficultés majeure consiste à représenter de manière succincte le type synthétisé vers l'avant à partir d'une formule qui a été obtenue à partir de types et d'expressions de chemins (XPath) combinés logiquement. En particulier, la récursion et les axes arrières ont besoin d'être réécrits dans la direction avant. Cela correspond à la représentation usuelle des types sous forme d'automates d'arbres. Le rapport [JGL13] décrit un mécanisme de synthèse de types monomorphe qui ouvre la voie pour le cas polymorphe d'un point de vue purement logique. La technique utilisée consiste à extraire de l'ensemble obtenu par calcul de point fixe du solveur un automate à requêtes (avec des états identifiés pour le contexte et le résultat) extrait à partir d'une formule (contenant des expressions XPath avec possiblement d'autres contraintes logiques). Le rapport décrit l'algorithme de synthèse ainsi que les preuves d'équivalence entre les formules et les automates.

La troisième phase devait porter sur la résolution des systèmes de contraintes générés dans la première phase. Initialement il était prévu d'utiliser le solveur logique, mais dans l'attente de résultats sur la synthèse de types nous avons développé des algorithmes spécifiques aussi bien pour l'inférence locale que pour la reconstruction de types (cette dernière donne des résultats plus précis que la reconstruction de types pour ML) qui ont été présentés dans [CNX13] ce qui remplit déjà les conditions du livrable D.2.c (prévu pour le mois 21).

La quatrième phase (dont le commencement était prévu à M24, c'est-à-dire janvier 2014) est déjà dans un état très avancé : elle devait porter sur le design et l'implantation d'une extension polymorphe de CDuce. L'article [CNXILP13] présente une technique d'implantation qui, grâce à une extension intelligente de la technique des fermetures, permet d'exécuter l'application de fonctions polymorphes avec la même efficacité que celle des fonctions monomorphes (ce qui est normal pour des langages comme ML mais extraordinaire dans un langage comme CDuce qui permet le test dynamique des types et qui donc ne permet pas de *type erasure*). Nous avons aussi implémenté des prototypes pour le Core fonctionnel de CDuce, avec inférence locale de types (et donc avec la résolution des contraintes). Ainsi pour terminer cette phase (et le Task) il reste à définir la syntaxe de CDuce polymorphe et y adapter l'implémentation ce qui sera la tâche du Post-Doc que PPS doit recruter.

### **Task 3 : Logically-driven implementation and certification of XML standards.**

Pour ce qui concerne la Task 3, la track 1 est achevée. Une implantation d'un moteur XPath basé sur les automates d'arbres (TAToo [N13], livrable D.3.a) est maintenant opérationnelle. Ce travail est une réécriture et extension d'une implantation préliminaire partielle (sans axes arrières) faite par un stagiaire de Master 1 (Lucca Hirschi encadré par Kim Nguyễn). Son rapport de stage [Hir12], contient en particulier la preuve « papier » de correction de la compilation d'XPath vers les automates. Même si l'implantation du livrable D.3.a n'est pas achevée (il faut maintenant y intégrer toutes les optimisations bas-niveau se trouvant dans SXSI et étendre le fragment supporté à toutes les fonctionnalités d'XPath, ce qui représente un gros travail d'ingénierie), le prototype est suffisamment avancé pour (i) confirmer que l'approche de l'évaluation d'XPath (incluant les axes arrières) est viable et (ii) servir de code « à certifier » en Coq (il contient déjà la compilation d'XPath vers automates et les optimisations de haut niveau que nous souhaitons formaliser). Nous notons que malgré la quantité importante de travail reliant automates d'arbres, logique et XPath, [N13] est à notre connaissance la première implantation OpenSource d'un moteur XPath basé sur des automates d'arbres).

La mécanisation de cette formalisation en Coq (prototype D.3.b) est presque achevée. Elle a pris un léger retard (prévue pour M15 et replanifiée M21) car les mêmes personnes travaillent aussi sur le Task 2 qui, par contre, est dans un état très avancé. Une autre raison de ce retard est lié à une mauvaise estimation de la dépendance entre Track 1 et Track 2 de cette tâche. Nous pensions initialement qu'après un travail préliminaire pour définir le modèle d'automates, les deux tracks (implantation du moteur de requête et sa preuve formelle) pourraient avancer en parallèle. Il se trouve que l'implantation et les tests effectués nous ont contraints plusieurs fois à changer le modèle d'automate de manière significative (pour des raisons d'expressivités ou de performance). Nous avons donc choisi de finaliser en premier lieu la track 1 (implantation complète du modèle d'automate) et de nous concentrer ensuite sur la track 2. Le retard relatif (M15 replanifié en M21) de la track 2 est donc à contraster avec l'avance prise sur la track 1 (Deliverable 3.a, Milestone 3.I prévus pour M21 sont livrés en M18).

Concernant les autres aspects de la track 2, et dans l'objectif d'améliorer la formalisation des standards du Web, nous avons ajouté un support pour les valeurs d'attributs au solveur. Nous avons alors tenté de modéliser formellement les feuilles de style CSS au moyen de la logique [GLQ12]. Ce travail, nous a permis d'introduire une notion de validité (le solveur du mu-calcul étant utilisé ici comme un prouveur de théorème) et donc de certification d'un certain nombre de propriétés visuelles des feuilles de style utilisées dans les pages Web. Il s'est avéré possible de trouver une traduction logique à des mécanismes complexes comme les priorités et l'héritage des valeurs. Par ailleurs, nous avons étendu le support des structures de base du solveur aux graphes et aux requêtes pour les graphes et nous l'avons équipé de capacités de raisonnement avec un support pour les langages SPARQL et RDFS [EGL12, CEG12, CEG13] ainsi que les axiomes de la logique SHI. Enfin, toujours dans le cadre de la track 2, Hyeonseung Im fournit un travail [INP13], où il formalise une algèbre de type récursifs riche (intégralement réalisée en Coq), qui peut servir de base à une formalisation d'objets récursifs utilisés dans une formalisation d'XML : automates d'arbres, schéma ou formules du mu-calcul.

#### **Task 4 : Reasoning about XML data and programs.**

La track 1, « Type aware query engine for efficient XML programming » est le sujet du stage post-doctoral proposé à Hyeonseung Im (post-doctorant Typex au LRI). La track prévoyait initialement 2 étapes : (i) définition d'un « cœur » fonctionnel pour XQuery 3.0 (permettant de se livrer à des développements formels sans pâtir de la complexité syntaxique de la spécification complète du W3C) et (ii) définition d'un système de types pour ce cœur fonctionnel. Le travail de Hyeonseung Im (en cours de soumissions), part du constat que CDuce, qui est un lambda-calcul typé avec entre autres l'ordre supérieur, est « presque » un bon cœur fonctionnel pour XQuery. La différence (de taille) entre les deux langages consiste en leurs différentes primitives de navigations (XPath pour XQuery, et motifs pour CDuce). Cette différence implique une différence au niveau du modèle de données : XPath nécessite de pouvoir naviguer arbitrairement dans un arbre (en particulier vers les nœuds parents) ce qui n'est traditionnellement pas possible dans une représentation purement fonctionnelle des documents XML (utilisée par CDuce). Utilisant d'une technique standard de programmation fonctionnelle (les « *zipper*s »), nous avons rajouté au modèle de données de CDuce la possibilité de naviguer « vers le haut ». Loin d'être une simple application d'une technique connue, il nous a fallu étendre le système de types de CDuce et intégrer les *zipper*s au niveau des types. La complexité théorique de l'approche apporte en contrepartie beaucoup d'avantages pratiques au langage : possibilité de contraindre la forme de l'arbre ou les navigations autorisées au niveau des types, extension du pattern-matching avec navigation à la XPath... Le langage obtenu est suffisamment riche pour en faire une cible de compilation typée pour XQuery 3.0, ce qui était le but de la track 1. Les résultats sont formalisés dans [CINB13] (en cours de soumission) et qui constitue le livrable D.4.a.

Dans [GLV12], nous présentons une technique nouvelle et un outil pour le contrôle de type statique des programmes XQuery se basant sur une approche purement logique. L'outil recherche les erreurs dans le programme en analysant à la fois le code source du programme et les schémas d'entrée et de sortie qui décrivent respectivement les ensembles de documents admissibles en entrée et en sortie. L'originalité de nos résultats réside dans l'utilisation conjointe de l'inférence de type arrière et d'une logique avec programmes inverses pour représenter des fragments de types d'arbre. Cela nous a permis de concevoir et de réaliser un contrôleur de type pour XQuery qui est capable de générer un schéma d'entrée lorsque celui-ci est absent. Ce mécanisme d'inférence arrière s'applique pour le fragment cœur de XQuery et est complémentaire avec l'inférence avant et le mécanisme de typage décrit précédemment.

En ce qui concerne la track 2, « XQuery annotations and static verifications » nous avons travaillé en particulier en effectuant une analyse des view-updates de XQuery en transformant ces dernières en formules SMT modulo en passant par une transformation des vues et des mises à jour sous forme d'ensemble d'expressions XPath qui sont ensuite traduites en formules pour le solveur [JGL12]. Nous avons aussi étudié l'évolution du schéma statique sur les programmes XQuery et proposé une méthode automatique qui vise à aider les développeurs à réaligner un programme XQuery avec le nouveau schéma et qui s'appuie sur le solveur logique du projet WAM. Le premier travail permet de supporter directement des contraintes d'intégrité exprimable avec les solveurs SMT Modulo. Le second travail

[OGL12] est une application des techniques de raisonnement pour l'écriture ou réécriture de code XQuery. Ces travaux s'inscrivent pleinement dans la deuxième étape de la track 2. La première étape, assertion statique pour XQuery basées sur les langages d'assertion pour SQL3 est le point de rencontre de la track 1 (fournissant un moteur XQuery typé avec autant de vérifications statiques que peuvent en fournir les types) et le deuxième pas de la track 2, utilisant des technologies allant au delà des types (satisfaisabilité modulo théories) pour enrichir le langage d'assertions.

#### **Task 6 : Nouvelles directions non prévues dans la proposition.**

Depuis la rédaction du projet nous avons pu constater que les requêtes de données semi-structurées ou faiblement structurées ne se limitaient pas à XML/XQuery mais que depuis on a pu constater l'émergence de nouveaux formats de données (notamment JSON et RDF) et de langages de requêtes (typiquement la famille de langages communément appelé NoSQL) ainsi qu'une utilisation de plus en plus diffuse de JavaScript. C'est pourquoi nous avons décidé d'adapter et étendre *les mêmes techniques* étudiées et développées par TYPEX à ce cadre qui en est très proche. L'adaptation des techniques du sous-typage sémantique aux langages NoSQL a été assez aisée ce qui a permis d'obtenir très rapidement des résultats qui ont fait l'objet d'une publication internationale [BCNS13], travaux que nous continuons à poursuivre en collaboration avec IBM Research (New York). Pour ce qui concerne JavaScript une formalisation de JavaScript en Coq est en cours, de la part des partenaires LRI (Arthur Charguéraud) et WAM (Alan Schmitt) en collaboration avec Philippa Gardner et Sergio Maffei (Imperial College).

### **C.3 DIFFICULTÉS RENCONTRÉES ET SOLUTIONS**

Le projet a rencontré plusieurs difficultés importantes (aucune n'est d'ordre scientifique) :

1. Suite à de graves problèmes de santé, à partir de juillet 2012 le coordinateur du projet n'a pas pu se déplacer ni travailler à plein régime pendant 8 mois. Ceci a eu comme conséquence d'annuler la réunion plénière prévue pour le mois 15 du projet (nous comptons la ré-organiser avant la fin de l'année) ainsi que deux réunions du steering committee (une a déjà été récupérée à avril 2013). L'autre conséquence a été de réduire les échanges et collaborations initialement prévus. C'est pourquoi nous comptons demander une rallonge d'au moins 8 mois pour le projet.
2. Suite au recrutement à un poste de Maître de Conférences, Sergueï Lenglet a démissionné de son poste de post-doc après 8 mois, sur les 18 prévus initialement. Il n'a donc pas pu achever le travail pour lequel il avait été recruté : « produce deliverable 2.a, ensure the success of milestone 2.II, contribute to the production and success of deliverable 2.b and milestone 2.I, and, according to the time frame and contingencies, ensure or take part in the production of deliverables 2.d and 5.a » L'embauche n'aura assuré que la production du deliverable 2.a ce qui, en considérant le temps d'apprentissage des techniques et du système peut être déjà considéré comme un succès important. Toutefois ceci signifie que des 18 mois de post-doc ils n'en resteront que 10 qui devront être couverts par un nouveau recrutement. Or, compte tenu du temps d'apprentissage ceci est très probablement insuffisant pour le travail qui reste à faire. Trois considérations en suivent :
  - a) Il est fort probable qu'un certain nombre de milestones prendront quelques mois de retard.
  - b) Il faudra recalibrer les objectifs du poste pour les adapter à la durée restante du poste.
  - c) Quoique ayant duré 8 mois l'embauche de S. Lenglet a utilisé 33.8% des ressources prévues pour l'embauche du post-doc ce qui, en termes de temps, correspond aux ressources afférentes à 6 mois de post-doc (ceci s'explique par l'absence d'ancienneté de S. Lenglet). Avec les ressources restantes on peut donc envisager d'embaucher un post-doc pour 12 mois avec un niveau de qualification moindre que celui demandé lors de la proposition, mais équivalent à celui de S. Lenglet. Une période de 12 mois serait plus adaptée aux objectifs du projet. Il faudrait donc que l'ANR nous autorise à utiliser les fonds restants pour embaucher un post-doc sur 12 mois plutôt que 10.
3. La situation du point précédent est aggravée par l'abandon de Matthias Kende (doctorant qui faisait partie intégrante de la proposition) pour aller travailler dans le privé (il a été recruté ingénieur chez Google).
4. Bien que les partenaires aient proposé de nombreux stages de recherche, les équipes LRI et PPS et WAM n'ont pas réussi lors de la première année à attirer des étudiants de M2 (seulement des étudiants en M1 ou L3) sur les sujets du projet. Nous expliquons cela entre autres par la situation très compétitive de l'offre de stages en région parisienne et grenobloise.
5. L'équipe WAM a eu des difficultés à recruter un postdoctorant suite à des défaillances de candidats. Ceci est une raison supplémentaire qui motive notre demande de rallonger d'au moins 8 mois la durée du projet.

## C.4 FAITS ET RÉSULTATS MARQUANTS

### Résultats marquants :

1. Définition d'un calcul explicitement typé pour les types intersection [CNXILP13,Xu13]
2. Définition d'un modèle d'exécution efficace pour une extension polymorphe de CDuce [CNXILP13]
3. Définition et développement d'une technique de sous-typage polymorphe à base de logique [GGL13]
4. Inférence locale et reconstruction pour un système de types ensemblistes (avec union, intersection et négation) et récursifs : génération et résolution de systèmes de contraintes [CNX13]
5. Définition d'un modèle formel pour langages NoSQL [BCNS13]
6. Définition d'une sémantique formelle et d'un système de types pour XQuery 3.0 [CINB13]
7. Intégration de patterns "navigationnels" (à la XPath) dans CDuce [CINB13]
8. Définition et développement de la première technique d'analyse statique des CSS [GGL13]

### Faits marquants :

1. Pierre Génèves du partenaire WAM vient de recevoir la médaille de Bronze du CNRS (voir <http://www.cnrs.fr/fr/recherche/prix/medaillesbronze.htm>).
2. L'article *Set-Theoretic Foundation of Parametric Polymorphism and Subtyping* par Giuseppe Castagna et Zhiwu Xu du partenaire PPS a reçu la nomination par ACM SIGPLAN aux *CACM Research Highlights* (voir: [SIGPLAN CACM Research Highlights Nominated Papers](#)). Il s'agit de l'article qui est à la base du **Task 2** de TYPEX et en particulier des livrables D.2.a, D.2.b, D.2.c et D.2.d.

## C.5 TRAVAUX SPÉCIFIQUES AUX ENTREPRISES (LE CAS ÉCHÉANT)

Néant

## C.6 RÉUNIONS DU CONSORTIUM (PROJETS COLLABORATIFS)

Date	Lieu	Partenaires présents	Thème de la réunion
9-10/1/2012	Paris	Tous	Voir programme détaillé <a href="http://typex.lri.fr/meetings.html">http://typex.lri.fr/meetings.html</a>
7/9/2012	Paris	Tous	Réunion de travail du Task 3 Track 2.
21/4/2013	Grenoble	Tous	Réunion steering committee
22/5/2013	Paris	Tous	Réunion de travail du Task 3 Track 2.

En plus un calendrier de séances de travail régulières a été mis en place pour les tâches qui requièrent la collaboration de plusieurs sites (par exemple des réunions hebdomadaires entre des membres du PPS et du LRI)

## C.7 COMMENTAIRES LIBRES

### Commentaires du coordinateur

L'avancement du projet est très satisfaisant et les objectifs de « cross-fertilization » prévus à ce stade du projet pleinement atteints. La production scientifique est légèrement plus importante que ce que nous avons prévu (cette quantité est à rapporter au fait que le projet inclue que 9 permanents et 2 post-docs). Si l'on se limite à une simple lecture syntaxique des publications (voir section D1 ci-dessous) on pourrait avoir l'impression que le projet avance en deux blocs séparés : LRI et PPS d'une part et WAM d'autre part. Une telle impression serait complètement erronée, car même si elle ne s'est pas encore traduite par

des publications communes avec l'équipe WAM l'influence réciproque a été très forte. Nous pouvons à ce titre citer 3 exemples.

1. Le livrable D.2.b produit par le partenaire WAM consiste en la mise en place sur le solveur WAM de la théorie des types développée par le partenaire PPS.
2. Le livrable D.4.a produit conjointement par LRI et PPS, est basé sur une technique (l'utilisation de zippers pour modéliser les axes « arrières » de XPath) introduite et développée par l'équipe WAM pour leur solveur.
3. Le prototype D.3.b (Task 3 Track 1) est actuellement développé par 3 chercheurs (Kim Nguyễn, Alan Schmitt et Matthieu Sozeau) qui appartiennent à 3 partenaires distincts du projet.

En conclusion l'objectif explicitement déclaré de fertilisation mutuelle prévu pour la première partie du projet est à considérer comme pleinement atteint.

### Commentaires des autres partenaires

...

### Question(s) posée(s) à l'ANR

Nous réitérons les questions déjà posées lors du rapport à 6 mois et pour lesquelles nous n'avons pas encore reçu de réponse :

1. Pouvons-nous utiliser les ressources restant prévues pour le recrutement d'un Post-doc à PPS pour embaucher un post-doc sur 12 mois plutôt que sur 10 ? Voir section B.3 pour explication détaillée.
2. Plus qu'une question, une demande très insistante : que tout formulaire à remplir tel que celui-ci soit aussi disponible dans un format ouvert. L'idéal serait de pouvoir utiliser un template latex mais on serait contents même avec un formulaire en format odt (Open Document Text : Word permet de sauver les documents en format odt depuis au moins la version Office 2007). Les partenaires ont perdu une quantité irraisonnable de temps pour synchroniser les différentes versions de Word. Ainsi, sauf avis de contraire de votre part à partir de ce document nous allons transmettre tous formulaires en format odt.

## D VALORISATION ET IMPACT DU PROJET DEPUIS LE DÉBUT

### D.1 PUBLICATIONS ET COMMUNICATIONS

**NB** : CES PUBLICATIONS SONT DISPONIBLES SUR LE SITE DE TYPEX [HTTP://TYPEX.LRI.FR/PUBLICATIONS.HTML](http://typex.lri.fr/publications.html).

Liste des publications multipartenaires (résultant d'un travail mené en commun)		
International	Revue à comité de lecture	1. 2.
	Ouvrages ou chapitres d'ouvrage	1. 2.
	Communications (conférence)	1. [BCNS13] V. Benzaken, G. Castagna, K. Nguyễn, and J. Siméon: Static and Dynamic Semantics of NoSQL Languages. In POPL '13, 40th ACM Symposium on Principles of Programming Languages, pag. 101—113, 2013. <a href="http://hal.archives-ouvertes.fr/hal-00797956">http://hal.archives-ouvertes.fr/hal-00797956</a>  2. [JGL12] M. Junedi, P. Genevès, N. Layaïda : XML Query-Update Independence Revisited, DocEng 2012, <a href="http://hal.inria.fr/hal-00734668/">http://hal.inria.fr/hal-00734668/</a>
France	Revue à comité de lecture	1. 2.
	Ouvrages ou chapitres d'ouvrage	1. 2.
	Communications (conférence)	1. 2.
Actions de diffusion	Articles de vulgarisation	1. 2.

	<b>Conférences de vulgarisation</b>	<ol style="list-style-type: none"> <li>[Cas12] G. Castagna. Type-checking union, intersection, and negation types",Conférence invité au 6th workshop on Intersection Types and Related Systems (ITRS '12), Dubrovnik, 2012.<i>Présentation des résultats préliminaires du WP2.</i></li> </ol>
	<b>Autres (articles soumis)</b>	<ol style="list-style-type: none"> <li>[CINB13] G. Castagna, H. Im, K. Nguyễn, and V. Benzaken: A Core Calculus for XQuery 3.0, April, 2013.(article soumis à une conférence internationale)</li> <li>[CNXILP13] G. Castagna, K. Nguyễn, Z. Xu, H. Im, S. Lenglet, and L. Padovani: Polymorphic Functions with Set-Theoretic Types. Part 1: Syntax, Semantics, and Evaluation, June, 2013.(article soumis à une conférence internationale)</li> <li>[CNX13] G. Castagna, K. Nguyễn, and Z. Xu: Polymorphic Functions with Set-Theoretic Types. Part 2: Local Type Inference and Type Reconstruction, July, 2013.(article soumis à une conférence internationale)</li> <li>[GGL13] Nils Gesbert, Pierre Genevès, Nabil Layaïda: A Logical Approach To Deciding Semantic Subtyping - Supporting function, intersection negation and polymorphic types (soumis à une revue internationale <a href="http://hal.inria.fr/hal-00848023">http://hal.inria.fr/hal-00848023</a>)</li> </ol>

<b>Liste des publications monopartenaires (impliquant un seul partenaire)</b>		
<b>International</b>	<b>Revue à comité de lecture</b>	<ol style="list-style-type: none"> <li></li> <li></li> </ol>
	<b>Ouvrages ou chapitres d'ouvrage</b>	<ol style="list-style-type: none"> <li></li> <li></li> </ol>
	<b>Communications (conférence)</b>	<ol style="list-style-type: none"> <li>[EGL12] Jérôme Euzenat, Pierre Genevès, Nabil Layaïda. SPARQL Query Containment under RDFS Entailment Regime. Melisachew Wudage Chekol,. In IJCAR 2012, International Joint Conference on Automated Reasoning.</li> <li>[GLQ12] Pierre Genevès, Nabil Layaïda, Vincent Quint. On the Analysis of Cascading Style Sheets, In WWW 2012, The World Wide Web Conference, 2012</li> <li>[GL13] Pierre Genevès, Nabil Layaïda. XML Validation: Looking Backward — Strongly Typed and Flexible XML Processing are not Incompatible (demo). 2nd International World Wide Web Conference (WWW'13), May 2013, Rio de Janeiro, Brazil. <a href="http://hal.inria.fr/hal-00837765/">http://hal.inria.fr/hal-00837765/</a></li> <li>[CEGL12] Melisachew Wudage Chekol, Jérôme Euzenat, Pierre Genevès, Nabil Layaïda. SPARQL Query Containment. In AAAI 2012, Twenty-Sixth AAAI Conference on Artificial Intelligence.</li> <li>[CEGL13] Melisachew Wudage Chekol, Jérôme Euzenat, Pierre Genevès, Nabil Layaïda. "Evaluating and benchmarking SPARQL query containment solvers". In ISWC 2013: The 12th International Semantic Web Conference.</li> <li>[JGL12] Muhammad Junedi, Pierre Geneves, Nabil Layaida, XML Query-Update Independence Analysis Revisited, DOCENG 2012, ACM Symposium on Document Engineering</li> <li>[OGL12] Raquel Oliveira, Pierre Geneves, Nabil Layaida, Toward automated schema-directed code revision,DOCENG 2012, ACM Symposium on Document Engineering.</li> <li>[JGL13] Louis Jachiet, Pierre Genevès, Nabil Layaida. Type Synthesis for the Logical Solver: an Approach based on Query Automata. <a href="http://typex.lri.fr/files/synthesis.pdf">http://typex.lri.fr/files/synthesis.pdf</a></li> <li>[GLV12] Pierre Genevès, Nabil Layaïda, Christine Vanoirbeek. XQTC: A Static Type-Checker for XQuery Using Backward Type Inference. <a href="http://hal.inria.fr/hal-00757867">http://hal.inria.fr/hal-00757867</a></li> <li>[INP13] H. Im, K. Nakata, S. Park : Contractive Signatures with Recursive Types, Type Parameters, and Abstract Types, In <i>Proceedings of the 40th International Colloquium on Automata, Language and Programming (ICALP 2013)</i>, July 2013 <a href="https://www.lri.fr/~im/publications/icalp2013.pdf">https://www.lri.fr/~im/publications/icalp2013.pdf</a></li> </ol>

	<b>Thèses et rapports de stage</b>	<ol style="list-style-type: none"> <li>[Xu13] Zhiwu Xu. Parametric Polymorphism for XML Processing Languages, PhD Université Paris Diderot and Chinese Academy of Science. This thesis provides the foundations for a polymorphic extension of CDuce and concerns the Task 2. (Formellement il s'agit d'un travail d'un seul partenaire mais il contient des contributions du partenaire LRI)</li> <li>[Hir12] Lucca Hirshi. Alternating Tree Automata for XPath Queries. Stage M1, ENS Lyon.</li> <li>[Lay13] Nabil Layaida. Représentation et analyses des contenus et des traitements sur le Web. Habilitation à diriger les recherches. Université de Grenoble. 23 Avril 2013.</li> <li>[Che13] Melisachew Wudage Chekol. Analyse statique de requêtes pour le Web Sémantique. Thèse de doctorat. Université de Grenoble. Décembre 2012. <a href="http://tel.archives-ouvertes.fr/tel-00834448/">http://tel.archives-ouvertes.fr/tel-00834448/</a></li> </ol>
<b>France</b>	<b>Revue à comité de lecture</b>	
	<b>Ouvrages ou chapitres d'ouvrage</b>	<ol style="list-style-type: none"> <li></li> <li></li> </ol>
	<b>Communications (conférence)</b>	<ol style="list-style-type: none"> <li></li> <li></li> </ol>
<b>Actions de diffusion</b>	<b>Articles de vulgarisation</b>	<ol style="list-style-type: none"> <li></li> <li></li> </ol>
	<b>Conférences de vulgarisation</b>	<ol style="list-style-type: none"> <li></li> <li></li> </ol>
	<b>Autres</b>	<ol style="list-style-type: none"> <li></li> <li></li> <li></li> </ol>

## D.2 AUTRES ÉLÉMENTS DE VALORISATION

Nous avons entamé un partenariat avec le Centre de Recherche T.J. Watson de IBM à New York pour appliquer les techniques de TYPEX aux langages NoSQL. Ceci a déjà donné une publication commune [BCNS13] (le livrable D.6.a) et au moment de l'écriture de ce rapport le partenariat se poursuit par l'encadrement conjoint d'un élève normalien (1 mois en France, 1 mois aux États Unis) sur les mêmes thématiques.

Liste des éléments. Préciser les titres, années et commentaires	
<b>Brevets internationaux obtenus</b>	<ol style="list-style-type: none"> <li></li> <li></li> </ol>
<b>Brevet internationaux en cours d'obtention</b>	<ol style="list-style-type: none"> <li></li> <li></li> </ol>
<b>Brevets nationaux obtenus</b>	<ol style="list-style-type: none"> <li></li> <li></li> </ol>
<b>Brevet nationaux en cours d'obtention</b>	<ol style="list-style-type: none"> <li></li> <li></li> </ol>
<b>Licences d'exploitation (obtention / cession)</b>	<ol style="list-style-type: none"> <li></li> <li></li> </ol>
<b>Créations d'entreprises ou essaimage</b>	<ol style="list-style-type: none"> <li></li> <li></li> </ol>
<b>Nouveaux projets collaboratifs</b>	<ol style="list-style-type: none"> <li></li> <li></li> </ol>
<b>Colloques scientifiques</b>	<ol style="list-style-type: none"> <li></li> <li></li> </ol>
<b>Autres (préciser)</b>	<ol style="list-style-type: none"> <li>[N13] TAToo Tree Automata Toolkit : moteur XPath basé sur des automates d'arbres : <a href="http://typex.lri.fr/software.html">http://typex.lri.fr/software.html</a></li> <li>[GLS13a] : XML Reasoning Solver Project : moteur de résolution logique du polymorphisme paramétrique. <a href="http://typex.lri.fr/software.html">http://typex.lri.fr/software.html</a></li> <li></li> </ol>

## D.3 PÔLES DE COMPÉTITIVITÉ (PROJET LABELLISÉS)

### Collaboration du projet avec le(s) pôle(s) ayant labellisé

Préparation d'une fiche de présentation pour le « Book » des projets 2011.
--

## Activités financées par le complément de pôle (laboratoires publics uniquement)

<b>Montant du complément accordé par l'ANR (pour chaque labo public)</b>	<ul style="list-style-type: none"> <li>- Partenaire LRI : 6102 €</li> <li>- Partenaire PPS : 5985 €</li> </ul>
--	--

Type d'action menée	Détails (exemples non limitatifs)	Dépenses complément de pôle*
<b>Actions contribuant à la réflexion stratégique et à la programmation scientifique du pôle</b>		
<b>Actions de communication scientifique et publique bénéficiant à la notoriété du pôle</b>	Présentation des résultats du projet suite à une invitation.	PPS: 1048.30 €
<b>Développement de la recherche partenariale (recherche de partenaires, frais de gestion du partenariat, ingénierie de projets,...)</b>	Veille technologique : Participation au <i>Research Projects Symposium</i> ainsi qu'à la multi-conférence ECMFA-ECOOP-ECSA en Montpellier.	LRI: 850 € PPS: 950 € (engagés mais no réalisés)
<b>Valorisation de la recherche et transfert vers le monde industriel</b>		

\* Estimation des dépenses imputées sur le complément de financement accordé au titre de la labellisation par un pôle de compétitivité, partenaires publics seulement.

## D.4 PERSONNELS RECRUTÉS EN CDD (HORS STAGIAIRES)

Identification				Avant le recrutement sur le projet			Recrutement sur le projet			
Nom et prénom	Sexe H/F	Adresse email (1)	Date des dernières nouvelles	Dernier diplôme obtenu au moment du recrutement	Lieu d'études (France, UE, hors UE)	Expérience prof. antérieure (ans)	Partenaire ayant embauché la personne	Poste dans le projet (2)	Date de recrutement	Durée missions (mois) (3)
Lenglet Serguei	H	serguei.lenglet@gmail.com	Toujours en contact	PhD	France	1	PPS	Post-doc	5/1/2012	8 (sur 18 prévus)
Im Hyeonseung	H	Hyeonseung Im <Hyeonseung.im@lri.fr>	Toujours en contact	PhD	Corée		LRI	Post-doc	2/11/2012	8 (sur 18 engagés)

## D.5 ÉTAT FINANCIER

Nom du partenaire	Crédits consommés (en %)	Commentaire éventuel
LRI	39.4 %	Personnel 35942€ (45%) Missions (33%)
PPS	22.6 %	Personnel 25862€ (35%) Missions 2651€ (10%)
WAM	4.33%	Personnel 0€ (0%), Missions 4372€ (15%)

Il s'agit des montants *réalisés* au 5 juillet 2013 (les montants engagés sont supérieurs, notamment pour ce qui concerne le personnel)

## E ANNEXES ÉVENTUELLES