

Projet ANR-14-CE28-0005-01

## PRODAQ - Proof Systems for Data Queries

Programme AAPJGC2014

<b>A IDENTIFICATION</b> .....	2
<b>B LIVRABLES ET JALONS</b> .....	2
<b>C RAPPORT D'AVANCEMENT</b> .....	2
C.1 Objectifs initiaux du projet.....	2
C.2 Travaux effectués et résultats atteints sur la période concernée.....	2
C.3 Difficultés rencontrées et solutions.....	2
C.4 Faits et résultats marquants.....	3
C.5 Travaux spécifiques aux entreprises (le cas échéant).....	3
C.6 Réunions du consortium (projets collaboratifs).....	3
C.7 Commentaires libres.....	3
<b>D VALORISATION ET IMPACT DU PROJET DEPUIS LE DÉBUT</b> .....	4
D.1 Publications et communications.....	4
D.2 Autres éléments de valorisation.....	4
D.3 Pôles de compétitivité (projet labellisés).....	5
D.4 Personnels recrutés en CDD (hors stagiaires).....	6
D.5 État financier.....	6

### A IDENTIFICATION

Acronyme du projet	PRODAQ
Titre du projet	Proof Systems for Data Queries
Coordinateur du projet (société/organisme)	Laboratoire Spécification et Vérification (LSV)
Date de début du projet	01/01/2015
Date de fin du projet	31/08/2019
Labels et correspondants des pôles de compétitivité (pôle, nom et courriel du corresp.)	
Site web du projet, le cas échéant	<a href="http://projects.lsv.ens-cachan.fr/prodaq/">http://projects.lsv.ens-cachan.fr/prodaq/</a>

Rédacteur de ce rapport	
Civilité, prénom, nom	M. Sylvain SCHMITZ
Téléphone	01 47 40 75 42
Courriel	<a href="mailto:schmitz@lsv.fr">schmitz@lsv.fr</a>
Date de rédaction	16/09/16
Période faisant l'objet du rapport d'activité	01/01/2015 au 01/07/2016

## B LIVRABLES ET JALONS

N°	Intitulé	Nature*	Date de fourniture			Partenaires (souligner le responsable)
			Prévue initialement	Replanifiée	Livrée	
0	Project website	site web	03/15		03/15	LSV
A4	Benchmark suite	données	07/16	07/17		LSV
A1	Proof theory for data logics	document	12/16	12/17		LSV
A2	Automated reasoning for data logics	document	12/17	08/18		LSV
A3	Prototype	software	12/18			LSV
B1	Data tree decomposition logic	document	12/16			LSV
B2	Proof structures as data structures	document	12/18			LSV
C2	Branching VASS for XPath	document	12/16	07/17		LSV
C1	Data logic capturing XPath	document	12/18			LSV

## C RAPPORT D'AVANCEMENT

### C.1 OBJECTIFS INITIAUX DU PROJET

L'objectif initial du projet PRODAQ est l'étude des liens entre d'une part les requêtes sur des données semi-hiéraphiques, par exemple de requêtes XPath sur des documents XML, et d'autre part les systèmes de preuve. Plus précisément, les objectifs principaux portent sur :

- Le développement de systèmes de preuves pour des logiques de données apparentées à XPath (task A). De tels systèmes permettent de raisonner formellement sur les propriétés des requêtes, et de procéder à diverses optimisations et vérifications. L'espérance est d'obtenir des algorithmes par recherche de preuve pour des fragments utiles en pratique de XPath. De tels systèmes de preuve bénéficient d'heuristiques qui les rendent aptes à être implantés, comparés aux algorithmes énumératifs « naïfs » connus actuellement.
- L'étude de modèles à données pour des logiques sous-structurelles comme la logique linéaire ou la logique de séparation (task B). Il s'agit principalement d'exhumer les liens profonds entre logiques de données et logiques sous-structurelles, mais on peut aussi entrevoir *in fine* l'utilisation de logiques sous-structurelles à la place des logiques de données pour raisonner sur les données.
- Les liens entre systèmes à compteurs et logiques de données (task C). Si des liens ont déjà été établis dans la littérature, il s'agit ici d'affiner et d'étendre les résultats existants.

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

#### Task A : Systèmes de preuve pour logiques de données

Les travaux sur cet objectif ont été obtenus lors de deux stages de Master de 5 mois chacun de Simon LUNEL en 2015 et d'Anthony LICK en 2016, tous deux encadrés par David BAELDE et Sylvain SCHMITZ. Nous avons attaqué le problème en commençant par des fragments restreints de XPath, qui peuvent s'exprimer dans des logiques modales pour des modèles d'arbres et de mots de données. Plus précisément :

- Nous avons développé un système de preuve pour la logique de données sur les arbres DataGL, un fragment de XPath. La recherche de preuve dans ce système correct et complet peut se faire en espace polynomial, ce qui est optimal pour cette logique [4,5].
- Nous nous sommes ensuite intéressés au cas des mots de données, et avons étudié des systèmes de preuve pour des logiques modales bidirectionnelles (à la LTL) sur les mots. Ces résultats améliorent les procédures de recherche de preuve connus d'INDRZEJCZAK (BSL 2016), qui n'apportaient pas de garantie de terminaison. Un article est en

préparation sur la base du mémoire de Master [6].

#### Task B : Modèles à données pour logiques sous-structurelles

Les travaux sur cet objectif ont été obtenus lors d'un stage de 2 mois de niveau Licence de Manoj KILARU en 2016, co-encadré par Étienne LOZES (LSV) et Sylvain SCHMITZ.

- Ces travaux, encore à un stade préliminaire, ont porté sur une variante de la logique de séparation capable de raisonner sur les mots de données. Nous avons pu montrer l'indécidabilité de cette logique et comparer son expressivité avec la logique  $FO^2$  (une logique proche de XPath) sur les mots de données dans un article en cours de préparation.

#### Task C : Systèmes à compteurs et données

Les travaux sur cet objectif ont été obtenus lors de collaborations entre Piotr HOFMAN (LSV, Cachan), Sławomir LASOTA (U. Varsovie), Jérôme LEROUX (LaBRI, Bordeaux), Ranko LAZIĆ (U. Warwick), Sylvain SCHMITZ, et Patrick TOTZKE (U. Edimbourg).

- D'une part, la première borne supérieure de complexité pour le problème d'accessibilité des réseaux de PETRI [1]. Depuis les preuves de décidabilité par MAYR (STOC 1981) et KOSARAJU (STOC 1982), il n'existait aucune borne de complexité pour ce problème central dans de nombreux domaines en informatique. En particulier, la même borne de complexité s'applique à la satisfaisabilité pour la logique  $FO^2$  sur les mots de données, de part les liens exhibés par BOJAŃCZYK *et al.* dans leur article séminial (PODS 2006, qui a reçu cette année le *test of time award* de la conférence PODS).
- D'autre part, l'étude de problèmes de décision pour des extensions des réseaux de Petri avec données [2,3]. De tels réseaux de PETRI sont à même de modéliser des versions simples des systèmes dynamiques centrés sur les données à la HARIRI *et al.* (PODS 2013, KR 2014). Nous montrons que les questions de finitude, par exemple de l'ensemble des données différentes utilisées par le système, sont décidables pour le modèle le plus simple [2], et caractérisons exactement la complexité de la vérification de propriétés de sûreté pour un modèle plus expressif, apte à garantir l'utilisation de données fraîches [3], résolvant par là même plusieurs problèmes ouverts du domaine.

### C.3 DIFFICULTÉS RENCONTRÉES ET SOLUTIONS

La principale difficulté rencontrée par le projet est la difficulté d'embaucher un doctorant. Le stage de Master de Simon LUNEL s'est très bien déroulé, mais pour des raisons personnelles, celui-ci n'a pas souhaité continuer en thèse en région parisienne. Les appels à candidature lancés à l'automne 2015 sont tombés « hors saison » et n'ont pas été fructueux. De ce fait, l'avancement sur la Task A est en retard d'environ une année. En l'absence d'un doctorant, nous avons préféré travailler sur les autres objectifs du projet, pour lesquels l'interaction avec un doctorant est moins primordiale.

Avec l'embauche en thèse d'Anthony LICK en septembre 2016, nous sommes confiants pour la poursuite du projet sur un calendrier révisé pour tenir compte de ce retard. Plus précisément, nous nous attendons à avoir une année de retard sur tous les travaux liés à la Task A.

Sur la Task C, nous avons quelque peu dévié par rapport à la proposition initiale. Nous avons

des collaborations en cours avec Diego FIGUEIRA (LabRI) et Ranko LAZIĆ (U. Warwick) sur le sujet du livrable C2, mais du fait de l'avancement rapide sur le sujet (proche) des réseaux de PETRI avec données, nous nous sommes plutôt concentrés sur ces derniers, avec succès.

#### C.4 FAITS ET RÉSULTATS MARQUANTS

L'article [1] présente la première borne supérieure de complexité pour le problème d'accessibilité dans les réseaux de PETRI. Cette question, liée à de très nombreux problèmes en informatique théorique, était ouverte depuis 25 ans, et sa résolution a fait l'objet depuis d'articles invités à la conférence STACS 2016 et dans le journal SIGLOG News.

#### C.5 COMMENTAIRES LIBRES

### D VALORISATION ET IMPACT DU PROJET DEPUIS LE DÉBUT

#### D.1 PUBLICATIONS ET COMMUNICATIONS

*Note : le projet Prodaq est mono-partenaire.*

Liste des publications monopartenaires (impliquant un seul partenaire)		
<b>International</b>	<b>Communications (conférence)</b>	<p>[1] Jérôme LEROUX et Sylvain SCHMITZ. <a href="#">Demystifying Reachability in Vector Addition Systems</a>. Actes de <b>LICS'15</b>, pages 56-67. IEEE Press, 2015.</p> <p>[2] Piotr HOFMAN, Sławomir LASOTA, Ranko LAZIĆ, Jérôme LEROUX, Sylvain SCHMITZ et Patrick TOTZKE. <a href="#">Coverability Trees for Petri Nets with Unordered Data</a>. Actes de <b>FoSSaCS'16</b>, LNCS 9634, pages 445-461. Springer, 2016.</p> <p>[3] Ranko LAZIĆ et Sylvain SCHMITZ. <a href="#">The Complexity of Coverability in y-Petri Nets</a>. Actes de <b>LICS'16</b>. ACM Press, 2016.</p> <p>[4] David BAELDE, Simon LUNEL et Sylvain SCHMITZ. <a href="#">A Sequent Calculus for a Modal Logic on Finite Data Trees</a>. Actes de <b>CSL'16</b>, Leibniz International Proceedings in Informatics. 2016.</p>
<b>France</b>	<b>Thèses de Master</b>	<p>[5] Simon LUNEL. Systèmes de preuves pour logiques modales à données. <i>Mémoire de Master</i>, LMFI, Université Paris-Diderot, septembre 2015.</p> <p>[6] Anthony LICK. Systèmes de preuves pour logiques modales. <i>Mémoire de Master</i>, MPRI, ENS Cachan, septembre 2016.</p>

#### D.2 AUTRES ÉLÉMENTS DE VALORISATION

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

Pas de labellisation par un pôle.

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

Tous les recrutements jusqu'ici ont été des stages sous convention, à savoir pour Simon LUNEL (5 mois en 2015), Manoj KILARU (2 mois en 2016) et Anthony LICK (5 mois en 2016). La thèse d'Anthony LICK démarre en septembre 2016.

#### D.5 ÉTAT FINANCIER

Nom du partenaire	Crédits consommés (en %)	Commentaire éventuel
LSV - UMR8643 - CNRS	9,27%	Consommation au 20/06/16 = 12927,37€ 11 125,53€ en fonctionnement/Missions 1 801,84€ en personnel (gratification stage)