



Département d'évaluation
de la recherche

Unité de recherche
Dossier d'évaluation

CAMPAGNE D'EVALUATION 2018-2019
VAGUE E

N.-B. : on renseignera ce dossier d'autoévaluation en s'appuyant sur l'« Aide à la rédaction du dossier d'autoévaluation d'une unité de recherche.»

INFORMATIONS GÉNÉRALES

Nom de l'unité : Données et Algorithmes pour la Ville Intelligente et Durable

Acronyme : DAVID

Domaine et sous-domaine dans la nomenclature du Hcéres : ST6 Sciences et technologies de l'information et de la communication - STIC ; ST6__1 Informatique

Directeur pour le contrat en cours : Dominique Barth

Directeur (ou du porteur de projet) pour le contrat à venir : Dominique Barth

Type de demande :

Renouvellement à l'identique

Etablissement et organismes de rattachement :

Liste des établissements et organismes tutelles de l'unité de recherche pour le contrat en cours et pour le prochain contrat (tutelles).

Contrat en cours :

Université de Versailles - St Quentin

Prochain contrat :

Université de Versailles - St Quentin

Choix de l'évaluation interdisciplinaire de l'unité de recherche (ou d'une ou plusieurs équipes internes) :

NON

Table des matières

1	Présentation de l'unité	5
1.1	Introduction	5
1.1.1	Création et structuration	5
1.1.2	Gouvernance	5
1.1.3	Localisation	5
1.1.4	Effectifs et budget	5
1.2	Politique scientifique	6
1.2.1	Projet scientifique 2016-2021	6
1.2.2	Stratégie du laboratoire	7
2	Présentation de l'écosystème recherche de l'unité	8
2.1	Ecosystème au sein de l'UVSQ et de l'université Paris-Saclay	8
2.2	Interdisciplinarité avec les Sciences Humaines et Sociales	9
2.3	Axe de recherche transverse sur les mobilités	10
2.4	Partenariats avec des TPEs innovantes	10
2.5	Ecosystème national et international	11
2.6	Formation doctorale	11
2.7	Implication dans les formations de masters	12
3	Faits marquants	13
4	Produits et activités de recherche	15
4.1	Synthèse des activités du laboratoire	15
4.2	Produits et activités de l'équipe ADAM	16
4.2.1	Bilan Scientifique	16
4.2.2	Liste des thèses par domaine d'expertise et/ou/domaine d'application (14)	19
4.2.3	Sélection des produits	21
4.2.4	Données chiffrées	23
4.2.5	Activité entre Janvier 2013 et Aout 2015	23
4.3	Produits et activités de l'équipe ALMOST	25
4.3.1	Bilan scientifique	25
4.3.2	Liste des thèses par domaines d'expertises et/ou domaines d'application (20)	28
4.3.3	Sélection des produits et des activités de recherche	31
4.3.4	Données chiffrées	33
4.3.5	Activités entre 2013 et 2015	34
4.4	Produits et activités de l'équipe PETRUS	35
4.4.1	Bilan scientifique	35
4.4.2	Liste des thèses au sein de l'équipe (7)	38
4.4.3	Sélection des produits et des activités de recherche	39
4.4.4	Données chiffrées	40
4.4.5	Activités entre 2013 et 2015	41
5	Organisation et vie de l'unité	42
5.1	Pilotage, animation, organisation de l'unité	42
5.2	Parité	42
5.3	Intégrité scientifique	42
5.4	Protection et sécurité	43

6	Analyse SWOT	43
6.1	Forces	43
6.2	Faiblesses	43
6.3	Opportunités	43
6.4	Menaces	44
7	Projet scientifique à cinq ans	45
7.1	Equipe ADAM	46
7.2	Equipe ALMOST	47
7.3	Equipe PETRUS	48
8	Annexe : participation à des comités de programmes	50

1 Présentation de l'unité

1.1 Introduction

1.1.1 Création et structuration

La création du laboratoire DAVID a été proposée et votée en conseil d'administration de l'université de Versailles - St Quentin (UVSQ) en juillet 2015, puis validée en tant qu'Equipe d'Accueil EA 7431 par le Ministère de la Recherche et de l'Enseignement Supérieur en janvier 2016. Cette création est le résultat de l'élaboration d'un projet scientifique par l'ensemble des membres du laboratoire présents en 2015. Ce projet s'appuie sur des activités menées avant cette date au sein de trois équipes de recherche déjà existantes et leur synergie ; il les a aussi amenées d'une part à proposer et investir de nouveaux axes de recherche et d'autre part à initier des actions de recherche liant données et algorithmes. Ces trois équipes, structurant le laboratoire DAVID sont en 2018 :

- L'équipe ADAM (*Ambient Data Access and Mining*), dirigée par Karine Zeitouni (Pr UVSQ) et qui compte 9 enseignants-chercheurs et 9 doctorants en 2018.
- L'équipe ALMOST (*Algorithms and Stochastic Models*), dirigée par Dominique Barth (Pr UVSQ) qui compte 12 enseignants-chercheurs et 16 doctorants en 2018.
- L'équipe PETRUS (*Personnal and Trusted Cloud*), équipe mixte INRIA-UVSQ dirigée par Nicolas Anciaux (CR INRIA), qui compte 7 enseignants-chercheurs, chercheurs et ingénieurs et 6 doctorants en 2018.

1.1.2 Gouvernance

Dès sa création, le laboratoire s'est doté d'un règlement intérieur, voté en assemblée générale du laboratoire en avril 2016, puis validé par la direction de l'université. Sa gouvernance est comme suit :

- Un directeur élu par les membres et nommé par l'université (Dominique Barth),
- Un bureau paritaire élu sur proposition du directeur : 3 membres de rang A (un par équipe), 3 membres de rang B (un par équipe), avec une parité F/H. Le bureau actuel est constitué de Nicolas Anciaux (B, PETRUS), Danièle Gardy (A, ALMOST), Zoubida Kedad (B, ADAM), Philippe Pucheral (A, PETRUS), Yann Strozecki (B, ALMOST), Karine Zeitouni (A, ADAM).
- Une représentante des doctorants élue par ses pairs, Stefi Nouhelo (ALMOST), qui participe notamment au bureau sur invitation du directeur.
- Des décisions prises en assemblée générale du laboratoire, avec un quorum du 2/3 du nombre de membres titulaires. Le rythme de réunion est depuis sa création de deux assemblées générales par trimestre.

La taille du laboratoire a permis une gouvernance agile et consensuelle, impliquant tous ses membres, ce pourquoi nous n'avons collectivement pas souhaité nous munir d'un conseil de laboratoire.

1.1.3 Localisation

Le laboratoire est entièrement situé sur le site de l'UFR des Sciences de l'UVSQ. Les équipes ADAM et ALMOST sont situées au troisième étage du bâtiment DESCARTES, et l'équipe mixte UVSQ/INRIA PETRUS est située au cinquième étage du bâtiment BUFFON. Les deux bâtiments sont séparés d'une centaine de mètres.

1.1.4 Effectifs et budget

En termes d'effectifs, le laboratoire compte en décembre 2017 vingt six enseignants-chercheurs et chercheurs titulaires. Le recrutement d'un nouveau maître de conférences a eu lieu en 2016 au sein de l'équipe PETRUS. Depuis sa création, trois membres titulaires membres d'ALMOST ont quitté le

laboratoire : Amélie Coulbaut-Lazarini, chercheuse contractuelle en sociologie qui a obtenu un poste de maître de conférences à Nice ; Florence Perronin, maître de conférence à l'université Joseph Fourier de Grenoble qui a réintégré son université d'origine en septembre 2016 ; Christan Cadéré, PAST UVSQ jusqu'en septembre 2017 et qui a changé d'emploi.

Le budget du laboratoire procède très majoritairement du financement de contrats et projets de recherche obtenus suite à des réponses à appels à projets (94,4% du budget total en 2017), ce qui reflète également l'importance et la richesse de l'écosystème de DAVID qui sera présenté dans la section 2.

Tableau des effectifs et moyens de l'unité (en euros)

Type de financement	2015	2016	2017
Appels à projet ANR (hors PIA)	177240	211784	172584
Programme Investissement d'Avenir	126649	219964	198609
Collectivités territoriales			10000
Autres financements publics sur appels à projets	15116	74716	104516
Fondations, associations, mécénats	30000	30000	30000
Contrats de recherche industriels	61800	11771	11471
Dotations UVSQ (hors allocations doctorales)		39600	31572
TOTAL	410805	587835	558752

(nb : suite aux instructions données par la tutelle, le budget total alloué à un projet durant plusieurs années comptables est entièrement pris en compte pour chacune des années).

1.2 Politique scientifique

1.2.1 Projet scientifique 2016-2021

L'objectif de ce jeune laboratoire est de mener des activités de recherche à travers une logique collaborative alliant les données massives (*Bigdata*) et l'extraction de connaissances de qualité, la sécurité et la confidentialité des données, la modélisation et l'algorithmique, afin de proposer des applications innovantes dans le contexte de la ville intelligente et durable. L'enjeu est à la fois de contribuer à des domaines de recherche disciplinaires en informatique alliant données et algorithmes, et d'adresser des défis sociétaux interdisciplinaires en s'appuyant sur des applications scientifiques innovantes pour la ville de demain. DAVID repose sur la collaboration de trois équipes de recherche reconnues dont l'expertise globale couvre les différents maillons de la chaîne scientifique abordée. Tout d'abord, l'expertise de l'équipe ADAM concerne la modélisation des données hétérogènes (notamment imparfaites ou incomplètes), l'intégration et la fusion sémantique de ces données, ainsi que l'extraction de connaissances à l'aide de requêtes complexes et de techniques de fouille. Ensuite, l'expertise de l'équipe ALMOST concerne la modélisation (notamment stochastique), la résolution algorithmique (complexité, algorithmique de graphes et des jeux, *machine learning*, optimisation combinatoire) et l'évaluation de performances pour des problèmes complexes et/ou de grande taille dans divers domaines d'application (*smartcities*, télécommunications, analyse moléculaire). Enfin, l'expertise de l'équipe PETRUS concerne les structures et algorithmes de gestion de données personnelles et leur sécurisation, la gestion de données à base de matériel sécurisé, les modèles de partage, les traitements distribués sécurisés à large échelle, et l'intégration de l'ensemble de ces techniques dans des architectures *Privacy-by-Design* permettant à l'individu de gérer son patrimoine numérique sous contrôle.

Le projet scientifique du laboratoire se situe dans un contexte où les sciences et technologies du numérique connaissent un tournant majeur dû à la taille et la diversité des données, à la convergence de l'algorithmique avec la modélisation et l'optimisation ainsi qu'à la fouille de données par des méthodes d'apprentissage statistique et automatique. C'est l'ère du *bigdata*, de la science des données,

de la modélisation et de la simulation. Données et algorithmes constituent les pierres angulaires de ce tournant : si la connaissance est dans les données, l'intelligence est dans les algorithmes. La recherche sur les données pose de nouveaux défis. Tout d'abord la prise en compte de nouvelles sources de données telles que les capteurs et les objets connectés, éventuellement géolocalisés ou mobiles. Ensuite il faut prendre en compte l'hétérogénéité sémantique des données et leur caractère multi-échelle. Il faut aussi garantir le passage à l'échelle pour traiter des données massives dans un environnement distribué. Enfin, le dernier défi réside dans la qualité des données qui peuvent être incomplètes, approximatives, incohérentes, irréalistes, bruitées ou manipulées. Il faut aussi ajouter d'autres défis scientifiques, en lien direct avec des enjeux sociétaux, comme la protection des données et la préservation de la vie privée. Ces défis incluent notamment (i) le design d'architectures de gestion de données personnelles selon l'approche du *privacy-by-design*, (ii) les nouveaux modèles de partage et d'usage des données personnelles, et le non contournement de ces modèles, et (iii) la définition de traitements de confiance permettant d'exploiter des données personnelles distribuées sans fuite d'information.

La recherche en modélisation et en algorithmique a tout d'abord comme ambition, dans le contexte de la ville intelligente et durable, de proposer des modèles et des algorithmes efficaces pour les applications visées, avec la prise en compte des éléments sociologiques, d'usage et d'infrastructures qui interagissent au sein des écosystèmes urbains visés. Il s'agit d'assurer la synergie nécessaire entre données, modèles et algorithmes pour passer du *Bigdata* au *Smartdata*. L'objectif est de proposer des solutions algorithmiques de dimensionnement ou d'évolution d'infrastructures urbaines et d'en étudier la sûreté de fonctionnement, de proposer des modèles et algorithmes pour le pilotage en temps réel (pour différentes granularités temporelles) et l'aide à la décision d'infrastructures urbaines. Ceci se fait notamment via la maîtrise des réseaux de télécommunication interconnectant toutes ces infrastructures et équipements. Pour comprendre et prédire les usages et les comportements dans leur dynamique, il est aussi nécessaire de proposer et d'évaluer, en particulier par simulation, les méthodes d'apprentissage les mieux adaptées (supervisées ou non, par renforcement, minimisation de regret, etc). Les préoccupations environnementales, la santé et la conception de matériaux innovants nécessiteront aussi la conception d'algorithmes d'analyse de données, de l'échelle moléculaire à l'échelle de données perçues par l'utilisateur.

Enfin, le projet scientifique de DAVID est centré sur les nouveaux défis du numérique dans la ville de demain. Ce thème cristallise l'attention sur plusieurs domaines interdépendants : les systèmes urbains et le patrimoine architectural, le transport et la mobilité, l'énergie et la pollution, la santé et le bien-être, la sécurité et la protection des personnes et de leur vie privée. La compréhension de ces domaines avec leurs interactions, la prise de décision sur de nouveaux critères (comme l'utilité sociale, la responsabilité, l'éthique ou l'environnement) et la gestion de ressources selon des paramètres conflictuels nécessitent une approche interdisciplinaire entre l'informatique, les sciences humaines et sociales, et l'environnement. En particulier, l'acquisition, l'exploration et l'analyse de grandes masses de données, la modélisation et l'optimisation (réseaux de transport, de distribution, ou de télécommunication), la conception d'applications sûres préservant la vie privée sont des thèmes de recherche centraux susceptibles d'apporter des réponses originales aux problèmes précédents. Ce positionnement interdisciplinaire fait du laboratoire DAVID une structure ouverte, susceptible de créer une force centripète autour d'un défi sociétal majeur et d'une recherche ciblée au sein de l'université Paris-Saclay.

1.2.2 Stratégie du laboratoire

La stratégie du laboratoire est tout d'abord de maintenir une excellence dans les domaines d'expertise de chacune de ses équipes et de faire vivre et évoluer le projet scientifique décrit ci-dessus avec l'appui de l'UVSQ et d'Inria Saclay. L'implication de tous les membres du laboratoire dans la définition du projet et son évolution est un élément important de la cohésion de DAVID, reflétée par le fonctionnement

souhaité en assemblée générale. Pour faire vivre ce projet, la stratégie du laboratoire décline aussi quatre éléments principaux :

- un fort investissement dans l'interdisciplinarité avec les sciences humaines et sociales, en particulier dans le cadre de la fédération SIHS du CNRS (voir section suivante), et le maintien de l'interdisciplinarité fructueuse avec d'autres disciplines (biologie, chimie, mathématiques),
- une incitation et un accompagnement à l'émergence d'activités contractuelles de recherche partenariales (nationales et internationales), en particulier publique-privée, avec un focus sur les collaborations avec des startups innovantes, et à l'établissement de relations ciblées et durables avec des acteurs territoriaux en lien avec le projet scientifique du laboratoire,
- le rayonnement scientifique du laboratoire par le soutien, notamment financier, à l'activité de publication d'excellence, en revues et conférences internationales de premier plan, par le soutien à la réponse à des appels à projets de recherche contractuels et par la participation de DAVID à des manifestations nationales et européennes à large audience concernant la ville numérique,
- une implication forte dans la formation doctorale, au sein de l'école doctorale STIC de Paris-Saclay, et dans les formations de masters, ce qui permet d'attirer de bons doctorants potentiels, avec de plus un soutien au financement de stagiaires de master 2 ou équivalent. Un des objectifs est notamment de permettre à chaque jeune maître de conférences de pouvoir co-encadrer au moins un doctorant, objectif actuellement atteint.

Enfin, le laboratoire investit en lien avec la fédération SIHS trois axes transverses stratégiques de recherche : Mobilités et Smartcities ; Données, justice et société ; Energie et villes durables. Ces axes seront développés dans la partie projet du présent document.

2 Présentation de l'écosystème recherche de l'unité

Le laboratoire DAVID participe à un écosystème de recherche très large, à la fois au niveau de son environnement proche, mais aussi régional, national et international. Cet écosystème est ici présenté de façon thématique.

(nb : sont soulignées dans la suite les participations de membres de DAVID à la gouvernance des structures indiquées.)

2.1 Ecosystème au sein de l'UVSQ et de l'université Paris-Saclay

Le laboratoire DAVID est un laboratoire de l'UFR des Sciences de l'université de Versailles - St Quentin. Il participe activement aux actions du **département STIC de Paris-Saclay** (Dominique Barth est membre du bureau du conseil de ce département depuis 2016). Au sein de ce département, l'algorithme, les réseaux, l'optimisation et la gestion des données, au cœur de l'unité DAVID, sont aussi des thématiques fortes que l'on retrouve au LRI, LIX, LTCI, SAMOVAR, L2S et LSV, souvent dans des équipes communes avec Inria ou le CEA. Une coopération ancienne avec ces équipes existe au sein des équipes ADAM, ALMOST et PETRUS.

Dans le cadre de l'**IDEX Paris-Saclay** et du Programme d'Investissement d'Avenir, les équipes de DAVID participent également aux **Instituts de Recherche Stratégiques** (IRS, <https://www.universite-paris-saclay.fr/fr/recherche/initiatives-de-recherche-strategique>), projets internes à l>IDEX Paris-Saclay labellisés et financés par l'Université Paris-Saclay afin de faire émerger des communautés scientifiques thématiques et des projets de recherche partenariaux : les IRS CDS2.0 (*Center for Data Sciences 2*) et ACE-ICSEN (*Adaptation to Environmental Challenges : a multi-scale and interdisciplinaire challenge*) qui impliquent l'équipe ADAM, et l'IRS B5GI (*Beyond 5G Internet*, avec le soutien de quatre acteurs industriels majeurs du domaine) qui implique ALMOST. Le laboratoire avait aupa-

ravant participé à deux **Lidex** (structures similaires aux IRS jusqu'à 2017) : l'Institut des Sciences du Numérique mêlant STIC et SHS, dans lequel un membre de DAVID, Nicolas Anciaux (PETRUS), était co-responsable du pôle "vie privée", et le Lidex Bio-info, alliant bio-informaticiens et biologistes, auquel participait ALMOST.

Le laboratoire est aussi engagé de façon très active dans différents programmes d'investissements d'avenir au sein de l'écosystème de l'université Paris-Saclay en lien avec son projet scientifique interdisciplinaire.

- le **Labex DIGICOSME** (*Digital worlds : distributed data, programs and architectures*) regroupe une grande partie des équipes de laboratoires du département STIC de l'Université Paris-Saclay. Les axes scientifiques du Labex impliquent les trois équipes de DAVID. Le Labex a financé trois contrats doctoraux au sein du laboratoire, et un membre de DAVID, Franck Quessette (ALMOST) est co-responsable de la commission formation et siège au comité exécutif du Labex. L'équipe ALMOST participe également aux groupes de travail *Future Access Networks* et *Optical Communications in Comex* du Labex.
- le **Labex CHARMMMAT** (*Chimie des Architectures Moléculaires Multifonctionnelles et des MATériaux*) réunit des chimistes, des physiciens et des informaticiens (principalement issus de DAVID). Tous les membres de l'actuelle université Paris-Saclay, dont des organismes comme le CEA et le CNRS, y sont impliqués. L'équipe ALMOST participe activement à ce Labex. Dominique Barth y a assuré durant quatre ans la responsabilité de l'axe transverse Modélisation et Caractérisation, qui a permis le financement d'un contrat doctoral co-encadré avec le laboratoire LAMBE de l'université d'Evry, d'un contrat post-doctoral et d'un stage de master 2.
- le **Labex PATRIMA** (*Patrimoines matériels : savoirs, conservation, transmission*) est porté par l'université de Cergy-Pontoise (UCP) et l'UVSQ. Il rassemble des équipes de recherche en sciences humaines et sociales et en sciences exactes afin de développer des actions de recherche et de formation dans le domaine du patrimoine matériel, sa conservation, sa compréhension et sa valorisation. L'équipe ADAM est fortement impliquée dans les actions de ce Labex qui a permis le financement de deux doctorants. En particulier, Karine Zeitouni (ADAM) est membre du comité scientifique de la Fondation des Sciences du Patrimoine depuis 2014, renouvelée en février 2017.

2.2 Interdisciplinarité avec les Sciences Humaines et Sociales

Les collaborations avec les sciences humaines et sociales ont d'abord été initiées au sein du Labex Patrima qui se consacre notamment à rendre le patrimoine culturel et historique plus accessible aux citoyens via les STIC. L'équipe ADAM y participe activement depuis six ans en collaboration avec plusieurs laboratoires et organismes du domaine SHS (C2RMF, LRMH, Centre de Recherche du château de Versailles, BNF, Musée du Louvre). Les collaborations avec les SHS ont ensuite été étendues par des actions communes menées par l'équipe PETRUS (anciennement SMIS) avec les laboratoires DANTE, CERDI et RITM sur la protection des données personnelles ; collaborations qui se sont poursuivies au sein du LIDEX "Institut de la Société Numérique", dont l'enjeu était d'aborder les défis sociétaux et juridiques liés à l'émergence des STIC.

Une stratégie majeure du laboratoire vise l'interdisciplinarité avec les sciences humaines et sociales. Ainsi, le CNRS a créé en janvier 2017 une **fédération de recherche FR2002 SIHS : Sciences Informatiques, Humaines et Sociales**. Cette fédération regroupe cinq laboratoires de l'UVSQ : CESDIP (UMR CNRS, Sociologie et sciences juridiques), CSHSC (EA UVSQ, histoire), DANTE (EA UVSQ, droit et sciences juridiques), DAVID et PRINTEMPS (UMR CNRS, Sociologie) ; elle a comme directeur Dominique Barth (DAVID) et comme directeur adjoint Jacques de Maillard (CESDIP). L'ambition de cette fédération de recherche est de susciter et de développer une réelle interdisciplina-

rité entre différentes disciplines des sciences humaines et sociales (droit, histoire, sociologie, sciences politiques), et des disciplines des sciences et technologies de l'information et de la communication, afin de pouvoir mener des projets scientifiques en vue de relever des défis sociétaux majeurs.

Le laboratoire DAVID participe activement, via l'UVSQ et l'INRIA, au nouvel **Institut de Convergence DATA-IA** dédié aux sciences des données et à leurs défis socio-économiques en France, avec une forte interdisciplinarité avec les sciences humaines et sociales. Nicolas Anciaux, responsable de l'équipe PETRUS, siège en suppléant d'une membre de la fédération SIHS au comité de programme de l'institut.

2.3 Axe de recherche transverse sur les mobilités

Le laboratoire a depuis sa création investi un axe de recherche interdisciplinaire concernant les mobilités urbaines et péri-urbaines. Une des thématiques de recherche de la fédération SIHS (voir plus haut), sous la responsabilité de Leïla Kloul (ALMOST) et de Yoan Demoly (lab. PRINTEMPS), donne lieu à des groupes de travail réguliers et actifs. Cet axe de recherche a amené le laboratoire à s'impliquer dans un large écosystème consacré à la mobilité.

- **VéDéCom** (*Véhicules décarbonnés et Communicants*) est un institut pour la transition énergétique (ITE du PIA), fondation partenariale publique-privée de l'UVSQ menant des activités de recherche et de formation consacrées aux véhicules et mobilités de demain. Le laboratoire est fortement impliqué dans deux des trois domaines de recherche de l'institut (Délégation de Conduite et la Connectivité, Mobilité et Energie partagées) en y menant des projets liés à l'analyse de données de mobilité (équipe ADAM), à l'aide à la conduite et à la gestion de flottes de véhicules autonomes (équipe ALMOST). Un membre de DAVID, Dominique Barth, a été membre du conseil d'administration de Védécom en 2015 et siège au sein du comité de pilotage du domaine "Mobilité et énergie partagées" depuis 2016. Quatre doctorants de Védécom sont ou ont été encadrés au sein du laboratoire.
- **L'Institut de Recherche Technologique (IRT) SystemX** est dédié à l'ingénierie numérique des systèmes du futur dont l'UVSQ est un partenaire. L'équipe ALMOST y mène une action de recherche concernant la sûreté de fonctionnement des véhicules électriques et autonomes à travers l'encadrement d'un contrat doctoral de l'IRT au sein de l'équipe ALMOST.

Enfin, dans le cadre de sa stratégie de recherche concernant les mobilités, le laboratoire s'implique dans la participation de l'UVSQ à la proposition d'une possible future KIC (*Knowledge Innovation Community*) sur la mobilité, et à l'élaboration d'un programme de recherche et d'innovation sur les mobilités impliquant les trois territoires de l'université Paris-Saclay. Au sein du GdR MAGIS (*Méthodes et Applications pour la Géomatique et l'Information Spatiale*), soutenu par les instituts INSHS et INS2I du CNRS (institut au sein duquel Mokrane Bouzeghoub (ADAM) assure une Direction Adjointe Scientifique), Karine Zeitouni (ADAM) a été co-animatrice de l'action prospective "Transport et mobilité" entre 2013 et 2016. Le sujet des mobilités urbaines est aussi un sujet d'innovation soutenu par le laboratoire au sein du groupe de travail "Ville Numérique" du **pôle de compétitivité SYSTEM@TIC** (Dominique Barth est membre du comité de pilotage de ce GT depuis 2015).

2.4 Partenariats avec des TPEs innovantes

Dans sa volonté d'intensifier ses partenariats de recherche avec les startups innovantes, le laboratoire interagit avec le **Versailles Sciences Lab** de l'UFR des Sciences de l'UVSQ qui initie et supporte, sous la direction de Dominique Barth, des projets associant startups innovantes, étudiants et laboratoires. Cette interaction a permis l'établissement de partenariats de recherche avec plusieurs startups

(contrats doctoraux CIFRE, réponses à des AAP).

En lien avec des structures de valorisation du territoire (pépinières d'entreprise de Versailles Grand Parc, Incubateur de Saint-Quentin en Yvelines), le laboratoire s'appuie sur différents modes de partenariats avec des TPE innovantes, en particulier l'ANRT pour le financement CIFRE de doctorants, le dépôt conjoint de projets de type FEDER et l'implication de certaines de ces TPE dans des réponses à des appels à projets nationaux.

2.5 Ecosystème national et international

DAVID est un des laboratoires partenaires du **Réseau Francilien en Sciences Informatiques** (Domaine d'Intérêt Majeur d'Ile de France) au bureau duquel siège le directeur de DAVID. En Ile de France, au delà de l'université Paris-Saclay, le laboratoire collabore avec plusieurs autres laboratoires. DAVID finalise également un accord cadre de recherche avec le LINEACT du CESI à Nanterre, avec qui une thèse en co-encadrement a débuté fin 2017 sur le sujet des bâtiments intelligents pour la maîtrise énergétique. Le laboratoire interagit aussi avec différentes communautés d'agglomération à travers des actions et projets de recherche : Versailles Grand Parc (projet ANR polluscope avec ADAM, co-organisation d'événements en lien avec la mobilité urbaine), Saint Quentin en Yvelines (un partenariat en discussion sur le sujet de la maîtrise énergétique des bâtiments), la ville de Paris (via l'association AIRPARIF). Enfin, le laboratoire participe au *Séminaire francilien de sûreté de fonctionnement* dont Leïla Kloul (ALMOST) a été co-responsable entre 2015 et 2017.

Au niveau national, l'équipe ADAM participe aussi aux GdRs MADICS (*Masses de Données, Informations et Connaissances en Sciences*) et MAGIS, déjà évoqué dans le paragraphe consacré à la mobilité. L'équipe ALMOST participe quant à elle au GDR RO (*Recherche Opérationnelle*), dont le comité scientifique compte Jean-Michel Fourneau (ALMOST), et au groupe de travail (*ALEA*) du GDR IM. De plus, comme le montrera la description des activités de recherche de chaque équipe, le laboratoire a obtenu le soutien de plusieurs organismes, agences et programmes français comme le **CNRS** (deux projets PEPS), l'**ANR** (6 projets dont 4 en cours), **Inria** pour mener ses activités de recherche partenariales nationales.

Enfin, DAVID entretient dans la durée de nombreux partenariats internationaux, plusieurs supportés financièrement par différents organismes. Ainsi, en plus des nombreux contacts informels, souvent concrétisés par des publications communes, voire des thèses en co-tutelle, le laboratoire a des collaborations contractuelles avec l'IT University of Copenhagen, le New Jersey Institute of Technology (NJIT), la Frei Universität de Berlin, l'Université de Ca' Foscari de Venise, l'Imperial College de Londres, l'Université de Technologie de Vienne et l'Université Fédérale de Ceara (à Fortaleza). Nombre de ces collaborations ont été financées par des PHC (un PHC Procope, un PHC Amadeus), un projet CNRS-FUNCAP, un projet ANR International, un projet PICS-CNRS et un projet CNRS-FUNCAP. De plus, le laboratoire participe au **GDRI "Innovative research on Web of Science"** en collaboration avec plusieurs Universités et organismes de recherches brésiliens. (2016 à 2020). Enfin, depuis 2014, Karine Zeitouni (ADAM) est membre du comité de gestion de l'action **COST TD1403** Big-Sky-Earth "Big Data Era in Sky and Earth Observation" (http://www.cost.eu/COST_Actions/tdp/TD1403_management).

2.6 Formation doctorale

En matière de formation doctorale, le laboratoire est très investi dans la gouvernance de l'**école doctorale STIC de l'université Paris-Saclay** depuis sa création en 2015. Cette école est composée de quatre pôles :

- Pôle 1 : Automatique, Traitement du Signal, Traitement des Images, Robotique
- Pôle 2 : Réseaux, Information et Communications
- Pôle 3 : Données, connaissances, apprentissage et interactions
- Pôle 4 : Programmation : modèles, algorithmes, langages, architecture

Les équipes ADAM et PETRUS sont rattachées au pôle 3 et l'équipe ALMOST est principalement rattachée au pôle 4 et secondairement au pôle 2. Dominique Barth puis Leïla Kloul (ALMOST) depuis septembre 2016 assurent la direction du pôle 4 au sein de l'ED, pôle qui compte environ 120 HdR et plus de 180 doctorants. Jean-Michel Fourneau (ALMOST) et Philippe Pucheral (PETRUS) sont respectivement membres des comités des pôles 2 et 3. De plus, Nicolas Anciaux (PETRUS) siège au conseil du collège doctoral de l'université Paris-Saclay, auquel Dominique Barth a aussi siégé durant un an.

Au sein de cette école doctorale, le nombre de thèses en cours et soutenues au sein de DAVID est important (30 doctorants). Les financements de ces thèses se font principalement via des actions et projets de recherche contractés par les membres et équipes du laboratoire (4 financements UVSQ, 26 financements hors UVSQ). L'effort d'encadrement d'étudiants de niveau Master 2 ou équivalent est également conséquent et soutenu par le laboratoire.

2.7 Implication dans les formations de masters

Le laboratoire accompagne et s'appuie sur quatre parcours de master 2 au sein de mentions de l'université Paris-Saclay, à des degrés divers d'implication :

- Le parcours DATASCALE (*Data Management in a Digital World*) de la mention de master Informatique de Paris-Saclay (Réfèrent : UVSQ ; opérateurs : Université Paris-Sud ; ENSIEE, Télécom Sud-Paris ; Responsables : Philippe Pucheral (PETRUS), Zoubida Kedad (ADAM)). Ce parcours offre une formation centrée sur la gestion de données à large échelle. Elle se distingue des autres formations du même type, dans l'Université Paris-Saclay comme au niveau national, par une approche holistique de cette problématique : (1) architectures pour le big data, le cloud et l'IoT, sécurité et performance, (2) intégration et qualité des données, fouille de données et (3) développement et déploiement de services et d'applications data-driven. Cette spécificité est rendue possible par la complémentarité des trois équipes DAVID et le renfort des autres opérateurs.
- Le parcours AMIS (*Algorithmique et Modélisation à l'Interface des Sciences*) de la mention de master Informatique de Paris-Saclay (Réfèrent : UVSQ ; opérateur : Télécom Sud Paris ; responsables : Dominique Barth, Sandrine Vial (ALMOST)). L'objectif de ce parcours est d'offrir une formation fondamentale de haut niveau aux techniques de modélisation stochastiques, probabilistes et algorithmiques avec une ouverture vers des applications pour la résolution de problèmes énergétiques (réseaux de télécommunication, *smartcities*,...) et de problèmes liés à la modélisation moléculaire (chémo-informatique, bio-informatique). Les membres de l'équipe ALMOST forment le principal support pédagogique de ce parcours de master, pour ce qui concerne les domaines fondamentaux et les domaines d'application.
- Le parcours AMI2B (*Analyse, Modélisation et Ingénierie de l'Information Biologique et Médicale*) de la mention Bioinformatique de Paris-Saclay (Réfèrent UPSud ; opérateurs : UVSQ, Ecole Polytechnique). Ce parcours a pour objectif la formation de spécialistes à l'interface de trois disciplines : Biologie, Informatique et Mathématiques. Les membres de DAVID interviennent dans ce parcours dans des UEs concernant l'algorithmique et les réseaux biologiques.
- Le parcours CCSC (*Commande des Systèmes Critiques*) de la mention Ingénierie des Systèmes Complexes de Paris-Saclay (Réfèrent ENS Paris-Saclay ; opérateurs : ENSTA, SUPMECA, UVSQ), qui vise la maîtrise des concepts, techniques et outils permettant la conception, l'analyse et l'amélioration de la sûreté de fonctionnement de systèmes complexes. Les membres

de DAVID y assurent une UE concernant les outils mathématiques pour la sûreté de fonctionnement et participent à une UE de Modélisation dysfonctionnelle et une UE de méthodes d'élimination des fautes.

Depuis 2015, sept doctorants du laboratoire sont issus du master AMIS et cinq sont issus du master DataScale.

La responsabilité de la mention Informatique de l'université Paris-Saclay est assurée depuis avril 2018 par Sandrine Vial (ALMOST).

3 Faits marquants

Hormis la qualité de sa production scientifique, qui sera démontrée dans la section suivante, quatre grands faits marquants peuvent être retenus dans l'activité de ce jeune laboratoire.

Le premier fait marquant est que le laboratoire créé il y a seulement deux ans, à partir d'un projet scientifique conçu dans une période alors incertaine pour les trois équipes, est aujourd'hui **un laboratoire pérenne tant par son activité scientifique que par sa forte visibilité** par les acteurs académiques, territoriaux et industriels dans le domaine de la ville numérique. On peut particulièrement noter la création en juillet 2017 de l'équipe mixte Inria-UVSQ PETRUS au sein de DAVID après une expertise internationale très positive menée par la direction d'Inria, projet faisant suite au projet SMIS. Preuve en est aussi non seulement le nombre de projets de recherche menés en partenariat avec de tels acteurs, mais aussi les invitations de membres de DAVID à participer à des manifestations spécifiques (Think Smart Grids, I.A. meeting ; Pleinières Ville numérique de SYSTEM@TIC ; Séminaire Campus 2017 du M.E.S.R.I., Smart Building, ; NOKIA Open Days, Mobilités ; etc). Le nombre important de contrats doctoraux obtenus en partenariat avec des acteurs publics et privés en est aussi une preuve, dans un contexte il est vrai de raréfaction du nombre d'allocations de recherche pour les STICs à l'UVSQ.

Un second fait marquant du laboratoire est, dans le cadre de son projet de recherche et de sa stratégie visant l'interdisciplinarité avec les sciences humaines et sociales, sa forte implication dans la création et l'animation de la **fédération de recherche FR2002 SIHS : Sciences Informatiques, Humaines et Sociales** qui, comme indiqué dans la section 2, implique cinq laboratoires de l'UVSQ : CESDIP, CSHSC, DANTE, DAVID et PRINTEMPS. Au sein de cette fédération, l'interdisciplinarité entre SHS et STIC concerne quatre premiers thèmes de recherche pluri-disciplinaires visant des défis sociétaux majeurs : Territoires et environnement, Mobilités, Société réelle et société virtuelle, et Culture et patrimoine. Chacun de ces thèmes est animé par un groupe de travail ; deux premiers axes de recherche, concernant d'une part la mobilité urbaine et d'autre part les données de la justice, ont permis d'initier des collaborations inter-disciplinaires et de répondre à des appels à projets. La fédération a déjà financé des soutiens logistiques d'activités de recherche et l'organisation de groupes de travail et de séminaires à l'échelle de l'UVSQ et de l'université Paris-Saclay, notamment en partenariat avec la Maison des Sciences de l'Homme de Paris-Saclay.

Le troisième fait marquant est la forte **activité de recherche et d'innovation en partenariat avec des startups innovantes du territoire**. Des partenariats de recherche pérennes et fructueux existent avec plusieurs startups. L'équipe Petrus a développé une collaboration étroite et durable avec la startup Cozy Cloud (<https://cozy.io/fr/>), leader national du Cloud personnel. Plusieurs projets de recherche communs ainsi que 2 thèses CIFRE sont en cours sur des sujets relatifs au partage sécurisé de données personnelles, à l'exécution de requêtes distribuées sans fuite d'information et à la définition de plateformes de cloud personnel "Privacy-by-Design". L'équipe PETRUS collabore également avec

la PME Hippocad (<http://hippocad.com/>) avec laquelle Inria et l'UVSQ fondent actuellement un IILab (Inria-Innovation Lab); cet IILab développé par Inria, l'UVSQ et la *startup* Hippocad vise le transfert industriel de travaux de recherche concernant la conception d'une box sécurisée faisant office de cloud personnel dédié au médico-social. L'équipe ALMOST collabore depuis sa création avec l'entreprise DCBrain sur des questions de maintenance et de pilotage de réseaux énergétiques, à travers une bourse CIFRE, le dépôt d'un projet FEDER et d'un nouveau projet CIFRE. L'équipe a aussi collaboré avec l'entreprise Bionext concernant le docking de molécules dans le domaine pharmaceutique (un doctorat CIFRE soutenu). L'équipe ADAM collabore avec COGNITUS, TPE de conseils en sciences de données, qui apporte son expertise pour le développement et le déploiement de l'architecture technique de la plateforme Polluscope en 2017. Enfin, le laboratoire collabore avec plusieurs *startups* innovantes via le co-encadrement de projets de master 1 et 2 au sein de Versailles Sciences Lab (Flex Services, LUMI-THD, TakeNB, Bservice, Jollyclik), *startups* avec lesquelles des collaborations de recherche sont envisagées.

Enfin, un dernier fait marquant concerne le **partenariat de DAVID avec différents acteurs territoriaux pour l'expérimentation et le déploiement de ses activités de recherche**. Citons par exemple le département des Yvelines, concernant le déploiement auprès de 10.000 patients ciblés par le Conseil Département du projet de PETRUS de concevoir la box sécurisée citée ci-dessus faisant office de cloud personnel dédié au médico-social, pour faciliter la coordination médico-sociale au chevet des personnes dépendantes. Le projet Poluscope mené par l'équipe ADAM concernant la qualité de l'air en zone urbaine implique l'association AIRPARIF et fait l'objet d'une convention avec la communauté d'agglomération de Versailles Grand Parc. L'équipe ALMOST a mené une action de recherche sur l'aide au stationnement urbain avec Védécom qui a impliqué la communauté d'agglomération de Versailles Grand Parc, ce qui a notamment permis l'obtention de données réelles.

4 Produits et activités de recherche

4.1 Synthèse des activités du laboratoire

Le laboratoire maintient depuis sa création en juillet 2015 un bon équilibre entre activités de recherche d'une part, comme le montre le nombre et la qualité des publications en revues et conférences internationales, et applications de leur recherche d'autre part, avec des partenariats interdisciplinaires, des acteurs territoriaux et industriels, et par les actions concrètes menées dans différents domaines. Le laboratoire accueille ou a accueilli 41 doctorants depuis sa création.

L'importance de l'interdisciplinarité au sein des actions initiées depuis la création de DAVID a permis, on l'a vu la création de la fédération SIHS au sein de laquelle des projets de recherche ont été initiés, concernant principalement la mobilité (en lien avec les laboratoires PRINTEMPS et le DANTE), l'analyse et la prédiction de données sociales et de justice (avec le PRINTEMPS, le CESDIP et le DANTE) et le statut juridique de la donnée (à travers une collaboration déjà fructueuse entre PETRUS et le DANTE). L'équipe ADAM mène aussi des projets de recherche interdisciplinaires avec les SHS, par exemple avec aussi le musée du Louvre, concernant la préservation et la valorisation du patrimoine culturel. Enfin, l'équipe ALMOST a obtenu des résultats significatifs en chémo-informatique concernant l'analyse de trajectoire de dynamiques moléculaires et la conception de cages moléculaires organiques, par ses collaborations avec les laboratoires de chimie LAMBE de l'université d'Evry et ILV de l'UVSQ.

En matière de rayonnement, hormis son importante activité de publication et de participation à des projets de recherche, le laboratoire participe à l'organisation de nombreux événements scientifique destinés à différents publics (acteurs de la recherche, entreprises, acteurs institutionnels). Enfin, comme le montre l'Annexe de ce document, les membres du laboratoire participent à un grand nombre de comité de programmes et/ou d'organisation de conférences internationales qui couvrent tous les domaines d'expertises des équipes du laboratoire.

Sont à présents détaillés les indicateurs de publications du laboratoire depuis 2015.

	2015	2016	2017	2018	Total
Doctorats	3	8	8	2	21
Livres, chapitres	—	2	3	1	6
Journaux internationaux	10	18	14	-	42
Journaux nationaux	1	6	3	—	10
Edition d'actes de conférences	—	—	1	—	1
Conférences et workshop internationaux avec actes publiés	25	34	42	3	104
Autres conférences	14	6	5	—	25
Conférences invitées	3	—	1	—	4
Posters	1	2	2	—	5

Les bilans scientifiques de chacune des trois équipes du laboratoires sont présentés dans la suite de cette section, en mettant en particulier l'accent sur l'importante activité d'encadrement doctoral, la qualité des publications scientifiques et la participation active à des projets de recherche contractuels.

4.2 Produits et activités de l'équipe ADAM

Nous présentons tout d'abord les membres titulaires et temporaires de l'équipe en 2018. Les doctorants seront présentés dans une prochaine section.

Membres titulaires

Bouzeghoub Mokrane (PR UVSQ)	Preda Nicoleta (MCF UVSQ)
Finance Béatrice (MCF HDR UVSQ)	Taher Yehia (MCF UVSQ)
Kedad Zoubida (MCF HDR UVSQ)	Yeh Laurent (MCF UVSQ)
Lopes Stéphane (MCF UVSQ)	Zeitouni Karine (PR UVSQ, <u>Resp.</u>)
Loyer Yann (MCF UVSQ)	

Ingénieurs contractuels et post-doctorants

Rafiq Haque (Déc. 2016, Avr. 2017)

4.2.1 Bilan Scientifique

Expertises et thèmes de recherche

La donnée est devenue aujourd'hui un matériau incontournable dans tous les domaines et dans notre vie quotidienne. Avec le tout numérique, la donnée est partout, multi-forme, parfois ouverte, en flux et massive. Si cela ouvre un champ d'application vaste, cela amène également à de nouveaux verrous scientifiques et technologiques que la recherche et l'industrie du big data tente de lever. L'équipe Ambient Data Access and Mining (ADAM) adresse en particulier les problématiques d'intégration, de requêtage et de fouille de données ambiantes de formes diverses allant des données ouvertes à celle générées par des capteurs en passant par les données du web profond. Ces axes thématiques ont fait l'objet d'applications principalement dans les domaines de la ville intelligente.

Intégration, agrégation et enrichissement de données

Une partie des travaux de l'équipe ADAM porte sur l'intégration, l'agrégation et l'enrichissement de données. Nous nous sommes en particulier intéressés à l'intégration de services Web dans une base de connaissance, à l'exploration de données sous forme de graphe ainsi à l'agrégation de données pour permettre des recherches par mots-clés, et à l'intégration contextuelle de données dans des environnements intelligents; nous nous sommes également intéressés à caractériser le contenu de sources de données du Web sémantique, et à extraire automatiquement leur schéma lorsque celui-ci est incomplet ou absent; enfin, nous avons proposé des outils d'intégration spécifiques au domaine des données de transports, notamment pour l'interconnexion ou le matching de sources de données au format RDF. Afin de permettre l'intégration automatique de nouveaux services Web dans une base de connaissance, une vue est calculée de manière automatique pour chaque méthode d'un service Web; cette vue est représentée comme une requête sur la base de connaissances, et nous avons proposé un algorithme qui calcule une fonction de transformation capable de transformer les résultats d'appel au service en un fragment conforme au schéma de la base de connaissance. Cette approche est particulièrement utile dans le cas de service ne permettant pas la publication de schéma, ce qui rend les algorithmes d'alignement existants inopérants.

Pour permettre l'exploration de sources de données RDF, nous avons proposé des approches qui intègrent de la connaissance métier dans le processus d'exploration pour l'adapter aux besoins de l'utilisateur. Pour représenter cette connaissance, nous avons introduit la notion de pattern, qui représente une équivalence entre une expression de propriété et une expression de chemin dans un graphe RDF et nous avons proposé une méthode d'évaluation de ces patterns. Nous avons proposé deux approches d'exploration complémentaires : l'exploration thématique et la recherche par mots clés. La

découverte de thèmes dans une source de données RDF consiste à identifier un ensemble de sous-graphes représentant chacun un ensemble cohérent de ressources sémantiquement liées et définissant un thème selon le point de vue de l'utilisateur. La recherche par mots clés dans une source de données RDF suppose l'indexation des éléments du graphe, l'identification des fragments pertinents pour une requête spécifique, puis l'agrégation de ces fragments pour former un résultat, ainsi que le classement des résultats obtenus.

Nous avons par ailleurs investigué la caractérisation du contenu d'une source de données RDF, dont la découverte automatique d'un schéma à partir de la structure implicite de ses données. Nous avons pour cela utilisé un algorithme de clustering et proposé une approche permettant la découverte des types sous-jacents dans une source de données ainsi qu'un profil probabiliste pour chacun ; ces profils sont notamment utilisés pour attribuer plusieurs types à une même instance, et pour attribuer un type à une nouvelle instance. Notre approche permet également de trouver les liens sémantiques et hiérarchiques entre les types en analysant les profils de types. Nous nous sommes également intéressés à caractériser la co-occurrence entre les propriétés des classes et nous avons proposé SchemaDecrypt, une approche permettant de découvrir en ligne les différentes versions structurelles d'un type.

Enfin, dans le cadre d'un projet avec l'institut VEDECOM, organisme de recherche public/privé dédié à la mobilité individuelle décarbonée et durable, nous nous sommes intéressés à l'intégration de sources de données dans le domaine du transport pour supporter la multi-modalité des déplacements des usagers ; ceci requiert l'intégration de données et/ou de services provenant de systèmes de transport hétérogènes, et ce afin de générer une vue globale du réseau de transport. Dans ce contexte, nous avons proposé : (i) une approche d'appariement de schémas pour les données géospatiales, (ii) une méthode de découverte de connexions offrant la possibilité de définir et de calculer des relations sémantiques riches entre des sources au niveau des instances et (iii) une solution pour la planification d'un trajet multimodal qui intègre les nouveaux services de mobilité avec ceux du transport public.

Requêtes optimisées sur des masses de données

Le traitement efficace des requêtes est un des thèmes de recherche récurrents en bases de données. Néanmoins, la nature des requêtes, la complexité des données et leur volume croissant soulèvent de nouveaux problèmes. Par ailleurs, les technologies de gestion de données ont connu de grands bouleversements avec l'avènement du big data. Dans ce contexte, les travaux de l'équipe ADAM ont concerné à la fois l'aspect fonctionnel des requêtes et leurs performances. Sur le plan fonctionnel, nous avons proposé une méthode pour la résolution de requêtes du web profond. Sur le plan de l'optimisation des performances des requêtes, nous avons proposé des méthodes de partitionnement et d'indexation, des algorithmes distribués et des techniques de compression des données dans le contexte des données spatio-temporelles.

Plus précisément, nous avons proposé d'étendre les requêtes aux Web à accès limité comme par exemple les services Web ou les formulaires. Ces interfaces sont essentiellement des requêtes paramétrées prédéfinies. Seulement, des classes entières de requêtes ne peuvent pas être évaluées efficacement du fait de la méconnaissance des méthodes publiées par le service ou des paramètres en entrée. Nous avons proposé un algorithme qui génère des descriptions pour les méthodes manquantes. Notre solution consiste à utiliser la partie visible du Web comme un Oracle pour deviner les réponses à ces méthodes. Un algorithme d'évaluation de requêtes qui évite les plans inefficaces a également été proposé.

Nos travaux sur l'optimisation des requêtes se sont focalisés sur les requêtes spatio-temporelles. En continuité des travaux sur l'indexation de trajectoires d'objets mobiles, un index de trajectoires a été proposé dans le contexte de mémoire flash. Le partage de traces de mobilité soulève des questions de confidentialité liées à la localisation des porteurs de mobiles. Nous avons proposé une solution de calcul et de partage sécurisés de requêtes agrégats spatio-temporels dans le contexte de collecte participative. L'anonymat est assuré par une architecture distribuée composée de mobiles sécurisés et par un protocole de communication adéquat.

Plus récemment, nous avons revisité la problématique d'organisation physique des données spatiales dans le contexte de systèmes distribués orientés mémoire. Nous avons proposé une nouvelle méthode de partitionnement bâtie sur une méthode d'indexation par courbes de remplissage sphériques et évalué expérimentalement de nouveaux algorithmes de recherche et de jointure spatiale et des stratégies d'optimisation logique et physique de requêtes. Enfin, nous avons proposé une méthode de compression de trajectoires exploitant la structure du réseau sous-jacent.

Fouille de données et extraction de connaissances

L'extraction d'information intelligible à partir des masses de données ambiantes passe par la fouille de ces données en lien avec des données de référence. L'équipe ADAM s'est principalement intéressée à la fouille de données de capteurs, les trajectoires et les bases de connaissances. Les contributions sont à la fois sur le plan méthodologique, qualitatif et des performances de calcul.

En traçant l'état des objets et/ou leur environnements, les capteurs permettent d'observer des situations particulières. La fouille des données ainsi produites offre une opportunité pour expliquer et prédire des événements d'intérêt comme des défaillances ou des situations indésirables. Toutefois, la complexité de la prédiction nous a amené à explorer de nouvelles méthodes en apprentissage et en intelligence artificielle. Ainsi, dans autoCEP, nous avons étendu les algorithmes d'apprentissage par shapelets à des séries temporelles multi-variées afin de répondre au contexte d'observation multi-capteurs. Notre algorithme permet de découvrir des règles prédictives mettant en évidence des relations temporelles complexes entre les différentes séries en entrée. De plus, ces règles sont exécutables directement par un moteur de traitement d'événements complexes (CEP), ce qui permet de bénéficier de l'interopérabilité de ces outils notamment le monitoring avancé de processus métiers prenant en compte les événements internes aux activités instrumentées par capteurs.

Dans le contexte des robots humanoïdes, il s'agit d'analyser des causes d'échecs de certaines actions du robot. Le modèle LEAF extrait les causes potentielles depuis l'historique grâce à un algorithme Multi-Armed Bandit. Ces connaissances sont ensuite injectées dynamiquement dans le réseau hiérarchique de tâches DHTN du robot, afin d'augmenter son efficacité tout en diminuant la consommation d'énergie. La reconnaissance d'activité est un autre volet de nos travaux. Nous avons montré que les connaissances ontologiques permettent l'amélioration des techniques d'apprentissage basées sur la vision pour la reconnaissance d'activité. Nous avons proposé un modèle de perception FSCEP combinant la gestion d'événements complexes, la logique floue et les ontologies pour gérer plusieurs dimensions d'incertitude dans les données ambiantes. Nous avons également proposé d'utiliser les réseaux logiques de Markov et le modèle FSCEP pour la détection de situation anormale et la prise de décision.

S'agissant des trajectoires, il s'agit de découvrir des structures (patterns) caractéristiques de la mobilité à partir de données spatio-temporelles. Pour ce faire, de nouvelles mesures de similarité et de nouveaux algorithmes sont nécessaires. La prise en compte des contraintes imposées par l'espace de mobilité est importante. Après s'être intéressée à la fouille d'ensembles de trajectoires, l'équipe a proposé des solutions pour le clustering de flux de trajectoires. Elle a ainsi proposé des métriques de similarité sur fenêtres glissantes et un algorithme incrémental de clustering par densité. Lorsque les trajectoires sont contraintes par un réseau pré-défini, celui-ci sert de base de représentation et dans le calcul de similarité, tandis que pour les trajectoires non contraintes, un index par courbe de Hilbert est défini pour réduire les calculs de similarités spatio-temporelles.

En outre, nous avons appliqué la fouille de textes aux bases de connaissances dans le but d'aligner les schémas de deux ontologies. Nous avons montré que l'alignement des schémas d'ontologies peut être vu comme un problème d'extraction de règles d'association. La nouveauté de notre approche est qu'elle permet la découverte des alignements complexes qui sont ignorés par les algorithmes existants.

Domaines d'applications La plupart des contributions de l'équipe ADAM ont été appliquées, principalement dans les domaines de la mobilité intelligente, du patrimoine culturel et de l'astronomie.

Sur le volet de la mobilité intelligente, on peut citer l'intégration des données de transport multi-modal de sources hétérogènes, ainsi que la planification dynamique de transport sous toutes ses formes (partagé, transport en commun). La fouille de trajectoires est un outil efficace d'analyse de la mobilité.

4.2.2 Liste des thèses par domaine d'expertise et/ou/domaine d'application (14)

Intégration, agrégation et enrichissement de données (6)

- [10/2012-09/2016]
Nom : Koutraki **Prénom** : Maria
Titre : Approches vers des modèle unifiés pour l'intégration de bases de connaissances.
Etablissement d'inscription : UVSQ
Financement : Labes PATRIMA
Encadrement : Karine Zeitouni (dir.), Dan Vodislav (co-dir., Univ. Cergy Pontoise), Nicoleta Preda.
- [10/2012- 10/2016]
Nom : Ouksili **Prénom** : Hanane
Titre : Exploration de sources de données RDF
Etablissement d'inscription : UVSQ
Financement : Industriel EDF
Encadrement : Mokrane Bouzeghoub, Zoubida Kedad, Stéphane Lopes.
- [10/2013- 10/2017]
Nom : Kellou-Menouer **Prénom** : Kenza
Titre : Découverte de schéma pour les données du Web sémantique
Etablissement d'inscription : UVSQ
Financement : Contrat doctoral Paris-Saclay
Encadrement : Zoubida Kedad.
- [02/2015- 11/2017]
Nom : Masri **Prénom** : Ali
Titre : Intégration multi-réseaux pour la mobilité intelligente.
Etablissement d'inscription : UVSQ
Financement : Institut Vedecom.
Encadrement : Karine Zeitouni (dir.), Zoubida Kedad (co-dir.).
- [09/2017- prévue en 09/2020]
Nom : Bouhamoum **Prénom** : Redouane
Titre : Découverte automatique de schéma pour des données massives.
Etablissement d'inscription : UVSQ
Financement : Contrat doctoral Paris-Saclay
Encadrement : Zoubida Kedad (dir), Stéphane Lopes.
- [03/2017- prévue 03/2020]
Nom : Rihany **Prénom** : Mohamad
Titre : Recherche mots-clés dans des sources de données RDF.
Etablissement d'inscription : UVSQ
Financement : mixte, ANR/bourse gouvernement libanais.
Encadrement : Zoubida Kedad (dir), Stéphane Lopes.

requêtes optimisées sur des masses de données (3)

- [11/2012 - 01/2016]
Nom : Dai Hai **Prénom** : Ton That
Titre : Gestion et Partage Sécurisés des Traces Personnelles Privées.
Etablissement d'inscription : UVSQ
Financement : ANR

Encadrement : Karine Zeitouni (dir.), Iulian Sandu Popa.

— [10/2015 - prévue 02/2018]

Nom : Brahem **Prénom :** Mariem

Titre : Traitement de requêtes spatiales sur des données massives d'observation de l'univers.

Etablissement d'inscription : UVSQ

Financement : mixte, CNES - UVSQ

Encadrement : Karine Zeitouni (dir.), Laurent Yeh.

— [10/2015- prévue 10/208]

Nom : Al Shaer **Prénom :** Mohamed

Titre : Intégration des méthodes CEP et traitement de données massives dans le contexte de l'Internet des objets.

Etablissement d'inscription : Université de Lyon

Financement : bourse gouvernement libanais.

Encadrement : Mohand-Said Hacid (dir. Univ de Lyon), Yehia Taher (UVSQ).

Fouille de données et extraction de connaissances (5)

— [10/2013 - 12/2016]

Nom : Coelho da Silva **Prénom :** Tician Linhares

Titre : Online clustering of trajectory data stream.

Etablissement d'inscription : Université fédérale de Ceara (Brésil)

Financement : Bourse Brésilienne CFPQ

Encadrement : Jose Antonion de Macedo (dir., UFC), Karine Zeitouni (co-dir., UVSQ)

— [10/2014-10/2017]

Nom : Mousheimish **Prénom :** Raef

Titre : Combinaison de l'Internet des objets, du traitement d'évènements complexes et de la classification de séries temporelles pour une gestion proactive de processus métier.

Etablissement d'inscription : UVSQ

Financement : Labex Patrima

Encadrement : Karine Zeitouni (dir.), Yehia Taher.

— [09/2014 - prévue 06/2018]

Nom : RAMOLY **Prénom :** Nathan

Titre : Intégration contextuelle de données dans les environnements intelligents ouverts et opportunistes : application aux robots humanoïdes.

Etablissement d'inscription : Telecom Sud Paris

Financement : Bourse Telecom

Encadrement : Amel Bouzeghoub (dir- TSP- SAMOVAR), Béatrice Finance (co-dir- UVSQ - David).

— [11/2016 - prévue 12/2019]

Nom : Kontarinis **Prénom :** Alexandros **Titre :** Enrichissement et fouille de trajectoires en environnement intérieur.

Etablissement d'inscription : Université de Cergy Pontoise

Financement : Labex Patrima

Encadrement : Dimitris Kotzinos (dir., UCP), Karine Zeitouni (co-dir., UVSQ)

— [10/2017- prévue 10/2020]

Nom : Romero **Prénom :** Julien

Titre : Fouille de schémas dans les bases de connaissances.

Etablissement d'inscription : Telecom ParisTech

Financement : ANR

Encadrement : Fabian Suchanek (Dir. TPT), Nicoleta Preda (UVSQ)

4.2.3 Sélection des produits

	2015	2016	2017	2018	Total
Doctorats	—	3	3	—	6
Livres, chapitres	—	1	1	—	2
Journaux internationaux	3	5	5	—	13
Journaux nationaux	—	2	—	—	2
Edition d'actes de conférences	—	—	1	—	1
Conf. et workshop Intern. avec actes publiés	10	20	24	1	55
Autres conférences (nat. et intern.)	7	2	4	—	13
Conférences invitées	1	—	1	—	2
Posters	1	2	1	—	4

Nous listons à présent une sélection des publications de l'équipe durant cette période :

Revue Internationale (4)

- [SP14] Fabian M. SUCHANEK et Nicoleta PREDA. "Semantic Culturomics (vision paper)". In : *PVLDB* 7.12 (2014), p. 1215–1218.
- [TSZ15b] Dai Hai TON THAT, Iulian SANDU POPA et Karine ZEITOUNI. "TRIFL : A Generic Trajectory Index for Flash Storage". In : *ACM Trans. Spatial Algorithms Syst.* 1.2 (juil. 2015), 6 :1–6 :44. ISSN : 2374-0353. DOI : 10.1145/2786758. URL : <http://doi.acm.org/10.1145/2786758>.
- [Bou+16] Mohamed Reda BOUADJENEK, Hakim HACID, Mokrane BOUZEGHOUB et Athena VAKALI. "PerSaDoR : Personalized social document representation for improving web search". In : *Information Sciences* 369 (2016), p. 614–633.
- [Ouk+17] Hanane OUKSILI, Zoubida KEDAD, Stéphane LOPES et Sylvaine NUGIER. "Pattern oriented RDF graphs exploration". In : *Data & Knowledge Engineering* (2017). ISSN : 0169-023X. DOI : <http://dx.doi.org/10.1016/j.datak.2017.06.003>. URL : <http://www.sciencedirect.com/science/article/pii/S0169023X17301118>.

Conférences Internationales avec PC et Actes (11)

- [KK15] Kenza KELLOU-MENOUER et Zoubida KEDAD. "Schema Discovery in RDF Data Sources". In : *Proceedings of the 34th International Conference on Conceptual Modeling (ER)*. 2015, p. 481–495. DOI : 10.1007/978-3-319-25264-3_36. URL : https://doi.org/10.1007/978-3-319-25264-3_36.
- [Jar+16] Amina JARRAYA, Nathan RAMOLY, Amel BOUZEGHOUB, Khedija AROUR, Amel BORG et Béatrice FINANCE. "A Fuzzy Semantic CEP Model for Situation Identification in Smart Homes". In : *Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI)*. 2016, p. 1678–1679.
- [MTZ16a] Raef MOUSHEIMISH, Yehia TAHER et Karine ZEITOUNI. "Autocep : automatic learning of predictive rules for complex event processing". In : *International Conference on Service-Oriented Computing*. Springer. 2016, p. 586–593.
- [MTZ16b] Raef MOUSHEIMISH, Yehia TAHER et Karine ZEITOUNI. "The Butterfly : An Intelligent Framework for Violation Prediction within Business Processes". In : *Proceedings of the 20th International Database Engineering & Applications Symposium (IDEAS)*. ACM. 2016.

- [Sil+16] Ticiania L. Coelho da SILVA, Karine ZEITOUNI, José Antônio Fernandes de MACÊDO et Marco A. CASANOVA. "CUTiS* : optimized online Clustering of Trajectory data Stream". In : *Proceedings of the 20th International Database Engineering & Applications Symposium (IDEAS)*. ACM. 2016.
- [Tah+16] Yehia TAHER, Rafiqul HAQUE, Mohammed ALSHAER, Willem-Jan van den HEUVEL, Karine ZEITOUNI, Renata ARAUJO, Mohand-Said HACID et Mohamed DBOUK. "A Service-Based System for Sentiment Analysis and Visualization of Twitter Data in Realtime". In : *International Conference on Service-Oriented Computing. Demo*. Springer. 2016, p. 199–202.
- [KK17] KENZA KELLOU-MENOUER et Zoubida KEDAD. "On-line Versioned Schema Inference for Large Semantic Web Data Sources". In : *Proceedings of the 29th International Conference on Scientific and Statistical Database Management, Chicago, IL, USA, June 27-29, 2017*. 2017, 9 :1–9 :12. DOI : 10.1145/3085504.3085513. URL : <http://doi.acm.org/10.1145/3085504.3085513>.
- [KPV17] Maria KOUTRAKI, Nicoleta PREDA et Dan VODISLAV. "Online Relation Alignment for Linked Datasets". In : *The Semantic Web - 14th International Conference, ESWC Proceedings, Part I*. 2017, p. 152–168. DOI : 10.1007/978-3-319-58068-5_10. URL : https://doi.org/10.1007/978-3-319-58068-5_10.
- [MTZ17] Raef MOUSHEIMISH, Yehia TAHER et Karine ZEITOUNI. "Automatic learning of predictive cep rules : bridging the gap between data mining and complex event processing". In : *Proceedings of the 11th ACM International Conference on Distributed and Event-based Systems (DEBS)*. ACM. 2017, p. 158–169.
- [RBF17] Nathan RAMOLY, Amel BOUZEGHOUB et Beatrice FINANCE. "A Causal Multi-Armed Bandit Approach for Domestic Robots' Failure Avoidance". In : *Proceedings of the 24th International Conference On Neural Information Processing (ICONIP)*. Springer. 2017, p. 90–99.
- [Ram+17] Nathan RAMOLY, Hela SFAR, Amel BOUZEGHOUB et Beatrice FINANCE. "LEAF : Using Semantic Based Experience to Prevent Task Failures". In : *Proceedings of the 11th Conference on Field and Service Robotics (FSR)*. Springer. 2017, p. 681–697.

4.2.4 Données chiffrées

Titre	Type	Budget	Date DEBUT	date de fin	Statut
Master : Multiple Aspect TrajEctoRy management and analysis	Horizon 2020 - Marie Curie RISE	62730	01/03/2018	28/02/2022	Partenaire
ACE ICSN : Adaptation aux Changements Environnementaux - Institut des Changements Socio-ENVironnementaux	IDEX-IRS	43100	2016	2018	Partenaire
POLLUSCOPE : Observatoire participatif pour la surveillance de l'exposition individuelle à la pollution de l'air en lien avec la santé	ANR-15-CEE22-0018-01	172721	01/04/2016	31/03/2021	Porteur
Stage de Master2	LABEX PATRIMA	3592	01/04/2016	30/09/2016	
Multi-Network Modeling for an Intelligent Mobility	PIA IET-VEDECOM	45000	01/02/2015	31/01/2018	Partenaire

4.2.5 Activité entre Janvier 2013 et Aout 2015

Nos contributions antérieures à la création du laboratoire DAVID avaient permis d'achever des travaux, d'une part sur l'extension des requêtes pour améliorer la recherche d'information et leur optimisation, et d'autre part sur l'extraction de connaissances. Ainsi, nous avons proposé une extension des requêtes de préférences multi-critères, afin de réduire et classer les solutions. La solution a consisté à introduire des critères objectifs permettant la comparabilité et le classement des solutions, de sorte à faciliter le choix. L'implémentation optimisée de ce concept a montré son intérêt pour la prise de décision par rapport à l'opérateur *skyline* classique, ainsi que sa supériorité en termes de performances de calcul. Nous avons, par ailleurs, proposé la personnalisation de la recherche d'information sur le Web en exploitant des réseaux sociaux. Cela s'est traduit par la proposition d'une méthode d'expansion de requêtes basées sur le vocabulaire social des utilisateurs et de plusieurs algorithmes d'analyse textuelles de micro-blogs.

Sur le plan de l'extraction de connaissances, nous avons proposé, dans le projet Semantic Culturomics, d'utiliser les bases de connaissances conjointement au texte pour découvrir des tendances dans l'histoire ou la culture, et ce dans le but d'extraire des règles expliquant ces tendances permettant de faire des prédictions. Dans le domaine de la fouille de données spatio-temporelles, nous avons achevé les travaux sur le clustering de trajectoires historiques, avant d'étendre la problématique aux trajectoires en flux. L'équipe ADAM a proposé et animé de nombreux travaux patrimoine culturel, que ce soit d'un point de vue des bases de connaissances et des corpus métiers, ou plus amont dans la conservation et le suivi (par capteurs) de l'environnement des objets du patrimoine.

	2013	2014	2015	Total
Doctorats	2	4	—	6
Livres, chapitres	1	—	—	1
Journaux internationaux	1	1	3	5
Journaux nationaux	—	1	—	1
Conf. et workshop Intern. avec actes publiés	16	5	10	31
Autres conférences (nat. et intern.)	3	3	7	13
Conférences invitées	—	—	1	1
Posters	—	1	1	2

4.3 Produits et activités de l'équipe ALMOST

Nous présentons tout d'abord les membres titulaires et temporaires de l'équipe en 2018. Les doctorants seront présentés dans une prochaine section.

Membres titulaires

David Auger (MdC UVSQ)	Dana Marinca (MdC UVSQ)
Dominique Barth (PR UVSQ, Resp.)	Laurent Marsan (MdC UVSQ)
Pierre Coucheney (MdC UVSQ)	Thierry Mautor (MdC HdR UVSQ)
Jean-Michel Fourneau (PR UVSQ)	Franck Quessette (MdC UVSQ)
Danièle Gardy (PR UVSQ)	Yann Strozecki (MdC UVSQ)
Leïla Kloul (MdC HdR UVSQ)	Sandrine Vial (MdC UVSQ)

Anciens membres titulaires

Christian Cadéré (DT Serenipay)	Florence Peronin (MdC INPG)
Amélie Coulbault-Lazarini (MdC Univ. Nice)	

Ingénieurs contractuels et post-doctorants

Mohamed Hadsı (sept 2107 - sept 2018)	Mael Guiraud (Janv. 2017 - Fev. 2018)
---------------------------------------	---------------------------------------

4.3.1 Bilan scientifique

La spécificité de l'équipe ALMOST est d'associer différents champs d'expertise concernant la modélisation algébrique et stochastique d'une part et l'algorithmique d'autre part, au sein de projets fédérateurs. Les membres de l'équipe ont développé depuis plusieurs années la capacité à mettre leurs diverses compétences en commun sur des mêmes projets applicatifs, qui touchent trois domaines : Mobilité et smartcities, maîtrise énergétique des infrastructures pour les *smartcities*, Analyse et prédiction de structures moléculaires. Nous présentons ci-après les activités scientifiques menées dans ces deux champs d'expertise et ces trois domaines d'application.

Algorithmique. L'équipe ALMOST développe par son activité et par la compétence de ses membres une expertise forte dans divers domaines de recherche en algorithmique discrète. Tout d'abord, l'étude de la complexité et de l'approximabilité de problèmes fondamentaux, nouveaux ou variantes de problèmes classiques, en particulier des problèmes d'énumération (au sein du projet ANR AGGREG visant l'aggrégation de requêtes de bases de données) ou de la théorie des graphes [MS16]. L'approche scientifique de l'équipe consiste à étudier chacun des problèmes tout d'abord sous l'angle de la complexité, puis de l'approximabilité avant d'en proposer des méthodes de résolution validées par l'évaluation de leurs performances. L'équipe mène par exemple une activité de recherche sur la complexité de variantes du problème de Steiner en lien avec des chercheurs d'autres laboratoires de l'université Paris-Saclay [Wat+16]. En matière de résolution de problèmes complexes de grande taille, l'équipe maintient une expertise en optimisation combinatoire, en particulier par l'utilisation de méthodes exactes et de méta-heuristiques pour des problèmes d'allocation de ressources ou d'ordonnancement [LeC+15]. L'équipe s'intéresse également à l'analyse et la gestion de systèmes dans lesquels l'ensemble des acteurs ou agents ont des intérêts propres et souvent divergents. Il ne s'agit donc pas ici de calculer une optimisation globale du système mais plutôt d'en étudier les dynamiques et les états d'équilibre, ce qui est l'objectif de la théorie algorithmique des jeux, un domaine d'expertise de l'équipe. En particulier, l'équipe s'intéresse à la complexité du calcul des équilibres de jeux stochastiques à information complète par l'obtention d'algorithmes efficaces, en particulier à complexité paramétrée, et à l'identification de classes de graphes où les stratégies optimales peuvent être calculées efficacement. En pratique, pour différents modèles de jeux à information incomplète, le calcul des équilibres est envisagé

par l'utilisation d'algorithmes d'apprentissage par renforcement ou par des méthodes de minimisation de regrets.

Il est à noter que les études menées sur les jeux stochastiques offrent un axe de recherche transverse aux deux domaines d'expertise de l'équipe puisque ces jeux constituent une extension du cadre des Processus de Décision Markoviens au cadre adversarial, permettant d'envisager les pires cas pour éviter les prises de risques dans les modèles de décisions stochastiques.

Modèles algébriques et stochastiques. Nous cherchons à construire et à analyser des modèles de systèmes et de réseaux ayant des ressources aléatoires en durée ou en taille. Les techniques employées peuvent être analytiques pour obtenir des solutions à forme close (par exemple à forme produit) ou consister à simuler le modèle en développant des algorithmes numériques efficaces aussi bien dans le domaine de l'analyse du régime stationnaire ou transitoire, de la durée de vie résiduelle (DFT et plus généralement fiabilité), ou de la politique optimale (jeux stochastiques). Dans le domaine analytique, les principaux résultats portent sur l'analyse de réseaux de files d'attente avec plusieurs classes de clients et des disciplines de service permettant de modéliser le partage de charge (Réseaux de Whittle).

Les techniques algorithmiques sur les distributions de probabilités ont conduit à des algorithmes de compression. Associées à des propriétés de monotonie stochastique, ces méthodes ont permis d'agréger des distributions empiriques de données en conservant des garanties sur les bornes calculées, par exemple pour analyser des éléments de réseaux soumis à des flux réels [Ait+16].

On s'intéresse également au développement de techniques formelles pour l'analyse de la sûreté de fonctionnement des systèmes industriels et, en particulier, à des approches, qui permettent une plus grande expressivité, une structuration hiérarchique des modèles en fonction des composants du système, et une modélisation basée sur une sémantique formelle. En collaboration avec A. Rauzy (NUST Trondheim, Norvège), nous avons développé à travers un doctorat CIFRE avec SIEMENS le formalisme *SCOLA* (*SCenario Oriented LAnguage*), en partie basé sur une algèbre des processus de type CSP, et qui permet la modélisation d'un système à partir de ses scénarios fonctionnels [IKR18]. Des mécanismes permettant l'analyse de la fiabilité des systèmes modélisés ont été associés à ce formalisme (notamment une *Failure Mode Effect Analysis* basé sur les scénarios). Par ailleurs, nous avons développé une nouvelle technique de modélisation (les *Arbres de Production AP*) permettant l'analyse de la disponibilité de la production d'un système, visant les flux en amont et en aval d'une ligne de production.

Enfin en collaboration avec d'IRIF, du Lynx et de la TU Wien, une modélisation des graphes par la combinatoire analytique qui a permis d'obtenir, outre des preuves alternatives de résultats connus, des résultats nouveaux dans le problème pourtant classique du seuil d'apparition et du nombre d'occurrences de sous-graphes dans des graphes simples ou des multi-graphes aléatoires suivant les modèles classiques d'Erdos-Renyi, où les degrés des sommets sont connus, ou bien suivent une loi de distribution donnée.

Maîtrise énergétique des infrastructures pour les *smartcities*. Ce domaine d'application vise la conception et l'analyse de modèles et de méthodes technico-économiques de gestion d'infrastructures avec comme but la maîtrise de leur consommation énergétique.

Le premier type d'infrastructures considéré par l'équipe concerne les réseaux de télécommunications, qui proposent des services garantis fixes ou mobiles, avec une maîtrise énergétique des solutions technologiques envisagées et modélisées [FMB16, Fou+17]. Qu'elles soient centralisées ou distribuées, les méthodes proposées doivent pouvoir évoluer dynamiquement selon les comportements, usages et besoins des applications et des utilisateurs, afin d'avoir une utilisation de ressources la mieux adaptée et donc la mieux énergétiquement maîtrisée. Dans ce but, l'équipe a participé au projet ANR NET-LEARN dont le but était de proposer et d'analyser des méthodes d'apprentissage automatique et leur

orchestration pour résoudre des problèmes d'allocation dynamique de ressources de réseaux de télécommunication, en particulier pour la gestion de ressources radio-mobiles [CKC15] ou pour la gestion de caches CDN (*Content Delivering Networks*) [Ben+15, Has+16], ce qui a donné lieu à la co-direction d'une doctorante avec Inria Paris. La maîtrise énergétique des réseaux 5G repose aussi sur le déport et le regroupement de certaines fonctionnalités d'éléments terminaux du réseau vers le cœur du réseau, en particulier dans le *cloud*. Il faut alors pouvoir garantir une latence déterministe et périodique des échanges de données entre le terminal et la fonction déportée. En partenariat avec Nokia Bell Labs France, l'équipe a défini les problèmes de routage déterministe périodiques associés et en a étudié la complexité (en lien avec le sujet d'un doctorat CIFRE en cours avec Nokia Bell Labs France). Elle a aussi proposé des heuristiques adaptées au problème spécifique du *Cloud RAN* (concernant la gestion des ressources radios pour les réseaux mobiles), à travers un doctorat co-encadré avec Télécom Paris-Tech. Une application de cette approche dans l'architecture de l'anneau optique a aussi été proposée au sein du projet ANR N-Green.

Plus récemment, l'équipe a considéré la gestion de la maîtrise énergétique des *DataCenters*, en menant des actions de recherche communes avec la *startup DCBrain* [BKR17]. Une thèse se termine actuellement sur le problème de la découverte de réseaux de distribution d'énergie à partir de mesures de ses flots.

Enfin, l'équipe a initié en lien avec le laboratoire LINEACT du CESI, avec lequel DAVID a une convention cadre de partenariat de recherche, une étude sur l'utilisation d'algorithmes d'apprentissage par renforcement pour la maîtrise énergétique d'un bâtiment, en particulier sur le contrôle du niveau de luminosité en fonction des usages (un doctorat co-encadré par le CESI vient de débiter sur ce sujet); l'étude utilisera notamment les données issues d'un bâtiment pilote construit par le CESI à Nanterre.

Mobilité et *smartcities*.

Les nouvelles technologies et besoins des grandes métropoles contraignent les concepteurs des systèmes de transport intelligents qu'ils soient véhiculaires ou ferroviaires à intégrer de plus en plus de fonctionnalités. Ces fonctionnalités impliquent une évolution dans l'usage de ces systèmes tout en rendant obsolètes les méthodes de conception traditionnelles de ces derniers.

Dans le cadre d'une thèse CIFRE avec Siemens, une approche de rétro-ingénierie des spécifications systèmes a été proposée pour les systèmes ferroviaires. Cette approche basée sur le formalisme développé SCOLA intègre des techniques d'analyse de sûreté fonctionnelles qui viennent en support aux approches sécuritaires existantes [IKR17]. L'approche développée a été utilisée pour la formalisation et l'analyse sécuritaire du système d'automatisation du métro de Siemens.

Actuellement, nous travaillons à développer une approche complète permettant la conceptualisation et la modélisation des scénarios décrivant les contextes d'exécution ou d'évolution d'un Véhicule Autonome (VA). Ces travaux de thèse qui s'inscrivent dans le cadre du projet SVA de l'IRT-SystemX, ont pour finalité la génération automatique des scénarios ayant un impact sur les performances et la sûreté de fonctionnement du VA.

Par ailleurs, une collaboration avec l'institut VeDeCoM (Véhicule Décarboné Communicant et sa Mobilité) se concrétise par le co-encadrement de plusieurs thèses sur le véhicule intelligent et/ou autonome. Une première thèse a ainsi été soutenue en mars 2018. Elle portait sur l'aide au stationnement en voirie dans les centres urbains et utilisait pour ce faire des algorithmes d'apprentissage. L'objectif était d'apprendre les probabilités de trouver une place de stationnement disponible dans les différentes artères et de déterminer alors l'itinéraire offrant la meilleure espérance de temps pour se garer. Deux thèses complémentaires viennent par ailleurs de démarrer sur la gestion intelligente d'une flotte de véhicules autonomes électriques (taxis, navettes). L'une, plutôt centralisée, s'appuiera principalement sur des techniques d'optimisation combinatoire et d'algorithmique on-line tandis que l'autre, dans un cadre plus distribué sur les différents VAs, utilisera principalement des méthodes d'apprentissage et de clustering. Dans les deux cas, il s'agira également de gérer la dimension énergétique à travers les

bornes de recharge.

Enfin, le laboratoire DAVID est un des acteurs principaux dans la définition d'un programme de recherche et d'innovation sur la mobilité s'appuyant sur les trois communautés d'agglomération du territoire avec comme objectifs la création d'un jumeau numérique dédié à la mobilité et des solutions numériques pour la logistique du dernier kilomètre.

Analyse et prédiction des structures moléculaires.

La modélisation de structures moléculaires adresse des problèmes à la fois issus de la chimie, organique et inorganique, en particulier liés à la chimie verte et de la biologie où les différentes meta-structures moléculaires (ARN, protéines, etc) sont le matériel d'étude.

En chimie et en biologie, l'équipe traite de la modélisation des différentes entités (atomes, molécules, clusters de molécules, etc), de leurs interactions (liaisons covalentes, liaisons hydrogènes, etc) et surtout de leurs dynamiques moléculaires. Notre approche sur la dynamique moléculaire en chimie est un facteur clé de la compréhension des interactions et en particulier des minima énergétiques, elle nous permet également la prédiction de cages moléculaires et la prédiction de chaînes réactionnelles pertinentes et ayant un impact écologique le plus bas possible; c'est notamment l'objet d'une thèse financée par le Labex CHARMMAT en partenariat avec le LAMBE de l'université d'Evry). Elle permet également la compréhension de certaines fonctions biologiques liées à la structure de l'ARN ou des interactions médicamenteuses en terme de docking de protéines (encadrement d'un doctorat CIFRE avec la startup BIONEXT) et physico-chimiques.

L'approche est une modélisation discrète, ciblant le niveau de granularité le plus pertinent permettant de répondre aux problématiques chimiques et biologiques sur des structures moléculaires de tailles réalistes. La génération d'objets combinatoires complexes est également un outil important de la prédiction et de la modélisation moléculaire [Bar+15]; c'est l'objet d'une thèse financée par le Labex DIGICOSME en lien avec le LRI de l'Université Paris-Sud et l'ILV de l'UVSQ). La complémentarité de plusieurs de nos domaines d'expertises : algorithmique discrète, théorie des graphes, théorie algorithmique des jeux et modèles aléatoires est pleinement à l'œuvre dans cette activité. Dans ce cadre, l'équipe maintient une activité de recherche avec l'équipe de Bio-informatique du LRI de l'Université Paris-Sud sur la prédiction de la structure tridimensionnelle de l'ARN (co-encadrement d'une doctorants financée par le DIM IdF) [Bou+17].

Le travail se fait en étroite collaboration avec des biologistes, bio-informaticiens, chimistes et physico-chimistes de l'eco-système de Paris-Saclay (le LAMBE de l'université d'Evry Val d'Essonne, l'ILV de l'UVSQ, le LRI de l'université Paris-Sud), en particulier au sein du Labex CHARMMAT pour ce qui concerne la chimie, le Labex DIGICOSME et le DIM IdF pour ce qui concerne la bio-informatique

4.3.2 Liste des thèses par domaines d'expertises et/ou domaines d'application (20)

Maîtrise énergétique des infrastructures pour les *smartcities* (9) :

- [10/2012 - 07/2016]
Nom, Prénom : Choutri Amira
Titre : Gestion des ressources et de la consommation énergétique dans les réseaux mobiles hétérogènes.
Etablissement d'inscription, financement : UVSQ, contrat doctoral UVSQ
Encadrement : Dominique Barth (dir.), Leïla Kloul
- [06/2014 - 06/2017]
Nom, Prénom : Ali Mohammed Shabbir
Titre : Apprentissage distribué dans les jeux pour les réseaux sans fil
Etablissement d'inscription, financement : Télécom ParisTech, ANR NETLEARN.
Encadrement : Marceau Coupechoux (dir, TPT), Pierre Coucheney.
- [10/2014 - 11/2017]

Nom, Prénom : Meng Huixing

Titre : Modeling patterns for performance analysis of production and safety systems in process industry.

Etablissement d'inscription, financement : Ecole Polytechnique, bourse CSC (Chinese Scholarship Council).

Encadrement : Antoine Rauzy (dir., NTNU Trondheim), Leïla Kloul.

— [10/2014 - 12/2017]

Nom, Prénom : Ben Hassine Nesrine

Titre : Apprentissage automatique pour la gestion des ressources réseau.

Etablissement d'inscription, financement : UVSQ, ANR NETLEARN et ANR etc

Encadrement : Dominique Barth (dir.), Pascale Minet (INRIA Paris), Dana Marinca

— [01/2015 -]

Nom, Prénom : Ehounou Wilfried

Titre : Algorithmes de découverte de graphes pour la maintenance de réseaux énergétiques.

Etablissement d'inscription, financement : UVSQ, Contrat CIFRE DCBrain

Encadrement : Dominique Barth (dir.), Arnaud de Moissac (DCBrain)

— [10/2015 -]

Nom, Prénom : Bennaceur Walid

Titre : Modèles formels pour l'analyse de la sûreté de fonctionnement d'un Datacenter.

Etablissement d'inscription, financement : UVSQ, contrat doctoral UVSQ.

Encadrement : Leïla Kloul.

— [10/2016 -]

Nom, Prénom : Ait El Mahjoub Youssef

Titre : Evaluation de performances pour des réseaux économes en énergie

Etablissement d'inscription, financement : UVSQ, Labex DIGICOSME (TSP).

Encadrement : Jean-Michel Fourneau (dir.), Hind Castel-Taleb (Télécom Sud Paris).

— [10/2017 -]

Nom, Prénom : Ségard Sébastien

Titre : Machine learning pour le pilotage énergétique intelligent de bâtiments.

Etablissement d'inscription, financement : UVSQ, CDI CESI

Encadrement : Dominique Barth (dir.), Benjamin Cohen-Boulakia (CESI)

— [04/2018 -]

Nom, Prénom : Guiraud Maël

Titre : Algorithmique de routage pour la maîtrise de la latence

Etablissement d'inscription, financement : UVSQ, contrat CIFRE Nokia Bell Labs.

Encadrement : Dominique Barth (dir.), Olivier Marcé (Nokia), Yann Strozecki.

Analyse et prédiction de structures moléculaires (5) :

— [10/2013 - 5/2016]

Nom, Prénom : Boudard Mélanie

Titre : Prédiction de structure tridimensionnelle de molécules d'ARN par minimisation de regret.

Etablissement d'inscription, financement : UPSud, contrat DIM IdF

Encadrement : Alain Denise (dir., UPSud), Dominique Barth, Johanne Cohen (UPSud-CNRS)

— [10/2013 - 06/2017]

Nom, Prénom : Voland Mathieu

Titre : Prédiction in silico d'interactions par similarité entre macromolécules biologiques

Etablissement d'inscription, financement : UVSQ, Contrat CIFRE BIONEXT

Encadrement : Dominique Barth (dir.), Sandrine Vial.

- [10/2014 -]
Nom, Prénom : Bricage Marie
Titre : Algorithmique pour la génération de cages moléculaires.
Etablissement d'inscription, financement : UVSQ, contrat doctoral UVSQ
Encadrement : Dominique Barth (dir.), Sandrine Vial
- [10/2014 - 12/2017]
Nom, Prénom : Bougueroua Sana
Titre : Algorithmes pour l'analyse et la prédiction des conformations des systèmes moléculaires en phase gazeuse.
Etablissement d'inscription, financement : UVSQ, Labex CHARMMMAT
Encadrement : Dominique Barth (dir.), Marie-Pierre Gageot (LAMBE, Evry), Franck Quesette
- [10/2016 -]
Nom, Prénom : Nouleho Ilenno Stefi
Titre : Algorithmique de graphe pour la prédiction de chaînes de réactions moléculaires
Etablissement d'inscription, financement : UVSQ, RTRA DIGITEO
Encadrement : Dominique Barth (dir.), Marc-Antoine Weisser (Supelec)

Mobilité et smartcities (6) :

- [01/2014 - 04/2017]
Nom, Prénom : Issad Mélissa
Titre : Reverse engineering of system specifications : application to the railway systems of Siemens.
Etablissement d'inscription, financement : Ecole Centrale, contrat CIFRE Siemens
Encadrement : Antoine Rauzy (dir., NTNU Trondheim), Leïla Kloul.
- [10/2014 - 02/2018]
Nom, Prénom : Houissa Asma
Titre : Algorithmes d'apprentissage pour l'aide au stationnement urbain.
Etablissement d'inscription, financement : UVSQ, CDD Védécom
Encadrement : Dominique Barth (dir.), Thierry Mautor.
- [10/2016 -]
Nom, Prénom : Chen Wei
Titre : Modèles formels pour la conceptualisation et la caractérisation des cas d'usage pour le véhicule autonome.
Etablissement d'inscription, financement : UVSQ, CDD IRT Systemix.
Encadrement : Leïla Kloul.
- [10/2017 -]
Nom, Prénom : Hoche Toussaint
Titre : Optimisation pour la gestion de ressources d'une flotte de taxis autonomes
Etablissement d'inscription, financement : UVSQ, CDD Védécom
Encadrement : Dominique Barth (dir.), Thierry Mautor, Wilco Burghout (Védécom)
- [12/2017 -]
Nom, Prénom : Babicheva Tatiana
Titre : Machine Learning pour la gestion distribuée d'une flotte de taxis autonomes.
Etablissement d'inscription, financement : UVSQ, CDD Védécom
Encadrement : Leïla Kloul, (dir.), Dominique Barth, Wilco Burghout (Védécom)
- [4/2018 -]
Nom, Prénom : Volpi Yoann
Titre : Optimisation d'un réseau logistique multi-flux par apprentissage de son comportement
Etablissement d'inscription, financement : UVSQ, contrat CIFRE DCBrain

Encadrement : Thierry Mautor (dir.), Dominique Barth, Arnaud de Moissac (DCBrain)

4.3.3 Sélection des produits et des activités de recherche

	2015	2016	2017	2018	Total
Doctorats	—	2	5	1	8
Livres, chapitres	—	—	—	1	1
Journaux internationaux	4	10	7	-	21
Conférences et workshop internationaux avec actes publiés	16	11	12	2	41
Autres conférences	—	2	—	—	2
Posters	—	—	1	—	1

Est à présent listé une sélection des publications de l'équipe durant cette période :

Livres, Actes ou Chapitre (1)

- [CCG18] Brigitte CHAUVIN, Julien CLÉMENT et Danièle GARDY. *Arbres pour l'algorithmique*. Springer, 2018.

Revues Internationales (5)

- [LeC+15] B. LECUN, T. MAUTOR, F. QUESSETTE et M.-A. WEISSER. "Bin packing with fragmentable items : presentation and approximations". In : *Theoretical Computer Science* (2015). DOI : <http://dx.doi.org/10.1016/j.tcs.2015.08.005>.
- [Ait+16] Farah AIT-SALAHT, Hind CASTEL-TALEB, Jean-Michel FOURNEAU et Nihal PEKERGIN. "Performance Analysis of a Queue by Combining Stochastic Bounds, Real Traffic Traces and Histograms". In : *Comput. J.* 59.12 (2016), p. 1817–1830.
- [Wat+16] Dimitri WATEL, Marc-Antoine WEISSER, Cédric BENTZ et Dominique BARTH. "Directed Steiner trees with diffusion costs". In : *J. Comb. Optim.* 32.4 (2016), p. 1089–1106. DOI : 10.1007/s10878-015-9925-3.
- [Bou+17] Mélanie BOUDARD, Dominique BARTH, Julie BERNAUER, Alain DENISE et Johanne COHEN. "GARN2 : coarse-grained prediction of 3D structure of large RNA molecules by regret minimization". In : *Bioinformatics* 33.16 (2017), p. 2479–2486. DOI : 10.1093/bioinformatics/btx175.
- [IKR18] Mélissa ISSAD, Leïla KLOUL et Antoine RAUZY. "Scenario-Oriented Reverse Engineering of Complex Railway System Specifications". In : *Journal of Systems Engineering* 21.2 (2018), p. 91–104.

Conférences Internationales avec PC et Actes (10)

- [Bar+15] Dominique BARTH, Olivier DAVID, Franck QUESSETTE, Vincent REINHARD, Yann STROZECKI et Sandrine VIAL. "Efficient generation of stable planar cages for chemistry". In : *14th International Symposium on Experimental Algorithms (SEA)*. 2015.
- [Ben+15] Nesrine BEN HASSINE, Dana MARINCA, Pascale MINET et Barth DOMINIQUE. "Popularity Prediction in Content Delivery Networks". In : *Personal Indoor and Mobile Radio Communication (PIMRC)*. 2015.

- [CKC15] Pierre COUCHENEY, Kinda KHAWAM et Johanne COHEN. “Multi-Armed Bandit for distributed Inter-Cell Interference Coordination”. In : *IEEE International Conference on Communications (ICC)*. IEEE. 2015, p. 3323–3328.
- [FMB16] Jean-Michel FOURNEAU, Andrea MARIN et Simonetta BALSAMO. “Modeling Energy Packets Networks in the Presence of Failures”. In : *24th IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, MAS-COTS 2016, London, United Kingdom, September 19-21, 2016*. IEEE Computer Society, 2016, p. 144–153.
- [Has+16] Nesrine Ben HASSINE, Dana MARINCE, Pascale MINET et Dominique BARTH. “Expert-based On-line Learning and Prediction in Content Delivery Networks”. In : *IWCMC 2016*. 2016.
- [MS16] Arnaud MARY et Yann STROZECKI. “Efficient enumeration of solutions produced by closure operations”. In : *STACS*. 2016.
- [BKR17] Walid Mokhtar BENNACEUR, Leïla KLOUL et Antoine RAUZY. “Safety Analysis of a Data Center’s Electrical System Using Production Trees”. In : *LNCS proceedings of the 5th International Symposium on Model-Based Safety and Assessment IMBSA*. T. 10437. Trento, Italie : Springer-Verlag, nov. 2017, p. 82–96.
- [Fou+17] Jean-Michel FOURNEAU, Lynda MOKDAD, Jalel BEN-OTHTMAN et Abdelkrim ABDELLI. “Admission Control Based on WRR in WiMAX Networks”. In : *2017 IEEE Wireless Communications and Networking Conference, WCNC 2017, San Francisco, CA, USA, March 19-22, 2017*. IEEE, 2017, p. 1–6.
- [IKR17] Mélissa ISSAD, Leïla KLOUL et Antoine RAUZY. “Scenario-based FMEA analysis and evaluation in a railway context”. In : *The Proceedings of the annual IEEE Reliability and Maintainability Symposium (RAMS)*. Miami, USA, 23-26 Janvier 2017.
- [Pan+17] E. de PANAFIEU, G. COLLET, D. GARDY, B. GITTENBERGER et V. RAVELOMANANA. “Threshold functions for small subgraphs : an analytic approach”. In : *Eurocomb*. Sous la dir. de Christian Krattenthaler MICHAEL DRMOTA Mihyun Kang et Jaroslav NESETRIL. T. 61. Wien (Austria) : Electronic Notes in Discrete Mathematics, août 2017, p. 271–277.

Réalisations

En matière de réalisation logicielle et d’application de résultats de recherche, un certain nombre de réalisations de l’équipe ont eu un impact sur leur domaine d’application.

- L’implémentation des algorithmes issus de la thèse de Sana Bougueroua sont mis à la disposition des chercheurs en chimie et physique (<http://hydrochronographe.prism.uvsq.fr>). Ces algorithmes ont déjà permis au laboratoire de chimie LAMBE de l’université d’Evry de montrer que l’eau liquide à l’interface avec l’air est organisée en un réseau 2D étendu (publication dans "The journal of physical chemistry letter").
- Le formalisme SCOLA développé durant la thèse de Mélissa Issad est aujourd’hui utilisé au sein de l’entreprise SIEMENS pour la génération de scénarios tests qui serviront à valider les exigences système du CBTC (système de contrôle automatique) des futurs trains de la ligne 4 du métro parisien.
- Le logiciel XBorné d’analyse numérique de modèles Markoviens a été étendu pour traiter des distributions discrètes de probabilités utilisées en fiabilité et comprend maintenant des moteurs de simulation en temps discret. Certains des outils ont été intégrés dans le logiciel MarmoteCore d’Inria pour fournir un noyau d’analyse Markovienne.

Rayonnement

En termes de rayonnement scientifique, en plus des publications et participations de membres de l'équipe à des conférences internationales, et de leur participation à des comités de programmes de conférences (voir Annexe), les membres de l'équipe sont régulièrement invités à participer à des événements à destination des entreprises, des collectivités locales et du grand public :

- Co-organisation et interventions pour des Pleinière "Ville numérique" de SYSTEM@TIC de mai 2016, novembre 2017 (IA et bigData) et d'avril 2018 (Mobilité urbaine, animation de la table ronde) ; Think Tank IoT des pôles de Compétitivité Cap Digital et Systematic Paris Région, mars 2017.
- Intervention au Think Smart Grids, I.A. meeting, juillet 2017 ; Table ronde Interconnectés Tour, Versailles, I.A. au service de l'Intelligence Collective, octobre 2017 ; Intervention au MeetUp I.A. Startups, Paris, mars 2018 ; Paris-Saclay SPRING 2018.
- Séminaire Campus 2017 du M.E.S.R.I., Smart Building, septembre 2017 ;
- NOKIA Open Days, Table ronde Mobilités, octobre 2017 ; Paris-Saclay SPRING 2018 ; jury du Challenge Open Innovation de Philips, 2018.

L'équipe maintient aussi un séminaire régulier (une à deux fois par mois), invitant très majoritairement des intervenants extérieurs.

4.3.4 Données chiffrées

Titre	Type	Budget ¹	date début	date fin	Statut ¹
NETLEARN : Orchestration d'algorithmes d'apprentissage distribués pour la gestion des ressources dans les réseaux mobiles	ANR INFRA	120224	10/2014	04/2017	Partenaire
N-Green : Nouvelle Génération de Routeurs pour des Réseaux Efficaces en Energie	ANR INFRA	72384	01/2015	06/2018	Partenaire
MARMOTE : MARKovian MOdeling Tools and Environments	ANR	148 651	01/2013	12/2016	Partenaire
METACONC : Méthodes analytiques non conventionnelles en combinatoire	ANR int.		10/2017	12/2019	Membre
AGGREG : Aggregate Queries	ANR	4500	10/2014	9/2019	Membre
CHARMMMAT Chimie des Architecture Moléculaires Multifonctionnelles et des MATériaux	Labex	171183	01/2015	-	Membre
DIGICOSME : Digital worlds : distributed data, programs and architectures / DIGITEO	Labex/RTRA	188656	01/2015	-	Membre
Threshold problems and phase transitions in graph-like structures	AMADEUS	1000	01/2015	12/2019	Membre
ACCA : Analyse de Contraintes par la Combinatoire Analytique	PICS CNRS	2000	01/2017	12/ 2019	Membre
Projet DAVID/Frei University	PROCOPE	6000	01/2013	12/2015	Partenaire

Contrats de recherche de l'équipe depuis septembre 2015

¹ Nb : dans le cas où le laboratoire est partenaire, est indiqué le budget total alloué à l'équipe. Dans le cas où le laboratoire est membre du projet, est indiqué le montant des activités directement financées

par le projet au sein de l'équipe puisque la totalité du budget est géré par un seul partenaire ou une structure dédiée.

Type de financement	nombre de doctorants	budget d'accompagnement ¹
Allocation UVSQ	2	-
CIFRE	5	35 keuros
LABEX, RTRA	3	10 keuros
ANR	2	-
Védécom	3	135 keuros
IRT SystemX	1	45 keuros

Financements de doctorats au sein de l'équipe depuis septembre 2015

¹ Hors contrat doctoral.

Le Labex DIGICOSME a également financé un an de post-doctorat et deux stages de masters 2 au sein de l'équipe.

4.3.5 Activités entre 2013 et 2015

L'équipe ALMOST est une émanation de l'équipe MAGMAT (Models Algorithms and Games for Molecular Analysis and Telecommunications in *smartcities*) du laboratoire PRiSM. Cette équipe se fondait déjà sur la synergie de trois thématiques (modèle stochastique, algorithmique et optimisation, théorie algorithmique des jeux) avec deux domaines d'application dominants, l'analyse moléculaire et les réseaux de télécommunication. L'investissement de l'équipe dans la définition du projet scientifique de DAVID lui a permis d'orienter plus avant une grande part de ses activités vers la ville numérique, en particulier la mobilité urbaine. Le tableau suivant résume les activités de publications de l'équipe MAGMAT durant cette période.

	2013	2014	2015	Total
Doctorats	2	4	—	6
Journaux internationaux	8	9	4	21
Conférences et workshop internationaux avec actes publiés	17	17	15	49
Autres conférences	—	1	—	1

4.4 Produits et activités de l'équipe PETRUS

Nous présentons ici les membres titulaires et temporaires de l'équipe en 2018. Les doctorants sont présentés dans la section suivante.

Membres titulaires

Nicolas Anciaux (CR Inria, HdR, Resp.)
 Luc Bouganim, Inria, (DR Inria HDR)
 Guillaume Scerri, (MdC UVSQ)
 Philippe Pucheral, (PR UVSQ)
 Iulian Sandu Popa (MdC UVSQ)

Membres non titulaires

Emmanuelle Perrot, Assistante Administrative, Inria
 Benjamin Nguyen, PR INSA CVL, membre associé

Ingénieurs contractuels et post-doctorants

Aydogan Ersoz, Inria, depuis 2015
 Laurent Schneider, Inria, depuis 2017

4.4.1 Bilan scientifique

Préambule : PETRUS est une équipe mixte Inria-UVSQ qui a été créée en juillet 2017 suite à un processus d'évaluation de son projet et des activités de l'équipe SMIS dont elle est issue. Le projet et le bilan ayant été dans ce cadre rédigé en anglais, nous avons pris la liberté de rédiger ce présent chapitre en anglais, tout en respectant la structure.

PETRUS (Personal and TRUSTed cloud) is a common team between Inria and UVSQ, member of the DAVID Lab. It replaces to the SMIS team which ended in 2016. We describe here the overall objectives of PETRUS, our research program structured around four axes, and our main results since 2016. The results obtained in the previous period 2013-2015 are presented at the end of Section 4.

Overall Objective. The Personal Cloud paradigm holds the promise of a Privacy-by-Design storage and computing platform, where each individual can gather her complete digital environment in one place and use it under control. However, this paradigm leaves the privacy and security issues in user's hands, which leads to a paradox if we consider the weaknesses of individuals' autonomy in terms of computer security, ability and willingness to administer sharing policies. The objective of PETRUS is to tackle the privacy and security challenges of the personal cloud from an architectural point of view. While many research works tackle the organization of the user's workspace, the semantic unification of personal information, the personal data analytics problems, the objective of the PETRUS project-team is to tackle the privacy and security challenges from an architectural point of view. More precisely, our goals are to propose (1) new architectures (encompassing both software and hardware aspects) for secure personal cloud data management and formally prove important bricks of the architecture, (2) new data administration models reaching the main requirements of a personal cloud (decentralized access and usage control models, data sharing, data collection and retention models, etc.) and study the enforcement of the resulting privacy policies based on secure hardware, (3) new secure distributed database indexing models, privacy preserving query processing strategies and data anonymization techniques for the personal cloud. In addition, we (4) explore the links between economic, legal,

societal and technological aspects, and in particular co-construct a secure personal cloud taking into account the EU privacy regulation called ‘GDPR’.

The four main research axes investigated in PETRUS are associated to each one of the above-mentioned research goals. We present below our recent results for each of them, as well as the main software developed by the team and called PlugDB.

Axis 1 - Personal cloud server architectures. Based on the intuition that user control, security and privacy are key properties in the definition of trusted personal cloud solutions, we explore new trustworthy architectures (encompassing both software and hardware aspects) for secure personal cloud data management. First, we consider the cases where the Secure Personal Cloud architecture relieves the individual security and privacy task by employing a secure token (i.e., a tamper-resistant hardware device) to control all the sensitive information (e.g., encryption keys, metadata, indexes) and operations (e.g., authentication, data encryption/decryption, access control, and query processing). However, secure tokens are usually equipped with extremely low RAM but have significant Flash storage capacity (Gigabytes), which raises important barriers for embedded data management. We proposed a new embedded search engine specifically designed for secure tokens [Lal+17], which applies to the important use-case of managing and securing documents in the Personal Cloud context. Conventional search engines privilege either insertion or query scalability but cannot meet both requirements at the same time. Moreover, very few solutions support data deletions and updates in this context. In this work, we introduce three design principles, namely Write-Once Partitioning, Linear Pipelining and Background Linear Merging, and show how they can be combined to match the hardware constraints and reconcile high insert/delete/update rate and query scalability. The integration of our solution in another important use-case related to performing information retrieval in smart objects has been studied in [Anc+15].

Second, managing huge quantities of data with constrained resources led us to focus on Solid State Disks (SSD) controllers based on flash memory. Open-Channel SSDs have emerged as a means to obtain predictably high performance, based on a clean break from the block device abstraction. In the uFLIP-OC study [citeTSZC16], we revisited our initial work on flash devices (uFLIP) in the context of Open-Channel SSDs. We showed how uFLIP-OC can be used to guide the design of host-based data systems on open-channel SSDs.

Axis 2 - Privacy preserving administration models and enforcement. The recent EU General Data Protection Regulation (GDPR) and the emergence of Personal Cloud platforms hold the promise of giving the control back to the individual on her data. However, this shift leaves the privacy and security issues in user’s hands, a role that few people can properly endorse. This research axis is devoted to the definition of privacy preserving administration models that are easily manageable for the individual and enforced by default (i.e., secure implementation).

In 2016, we proposed a data sharing algebra called DatSha. It builds upon the concept of UCON, and enables a user to specify privacy-preserving workflows on their data such that it gives non-expert users a means to exert usage control through example, reusing algebraic plans proposed by expert users.

More recently, we proposed a new data administration model, called SWYSWYK, to regulate data sharing and enforcement aspects based on the use of secure -trusted- execution environments. In this study, we first defined of a Privacy-by-Design sharing architecture, which allows each user to physically visualize the net effects of sharing rules on her Personal Cloud and automatically provides tangible guarantees about the enforcement of the defined sharing policies. The architecture relies on a secure reference monitor, a set of user defined functions only interacting with the secure monitor and isolated from the unsecure environment, and an unsecure personal cloud platform managing encrypted personal data. This architecture is presented in [TAP17]. Second, we explored the semantics of the models which could be supported using the SWYSWYK architecture. Our proposal builds on the fact that

the personal cloud content intrinsically describes the individual's acquaintances under different forms (e.g., contact files, agendas, identity pictures, address book entries, etc.). We have characterized new sharing models which should be able to map personal data to acquaintances (or subjects) and exploit their links with the stored documents (or objects) to produce authorizations satisfying users' sharing desires. We have demonstrated the semantics of such a model in a real personal cloud platform (namely Cozy) and apply it to a smart surrounding scenario, with the goal to assess its practical interest.

Axis 3 - Global query evaluation. Crossing data belonging to multiple individuals (e.g., performing statistical queries over personal data, computing queries on social graphs or organizing participatory data collection) in a fully decentralized setting while providing strong and personalized privacy guarantees, is at the heart of the expectations of a personal cloud, as it is the mean for a population of individuals to collectively contribute with their personal data to explore a societal problem (e.g., participate to an epidemiological study, recommendation systems, feed IA or learning processes, etc.). This means proposing new secure distributed database indexing models, privacy preserving query processing strategies and data anonymization techniques for the personal cloud. PETRUS explores several research directions for the secure execution of various algorithms through distributed protocols on personal data stored in an unstructured network of Personal Clouds (i.e., personal trusted device) so that each user can keep control over her data. The data could be stored locally in a personal cloud or encrypted on some external untrusted cloud. Execution takes place on a hybrid infrastructure called the Asymmetric Architecture (AA) : the network of personal clouds, supplemented by an untrusted cloud supporting IaaS or PaaS. Our objective is to show that many different algorithms and computing paradigms can be executed on AA, thus achieving secure and private computation. Among the main problems tackled in this context we indicate : (a) distributed SQL query processing, for queries without joins, but including Group By and aggregates (e.g., TODS'16 in [TNP16]); (b) computing MapReduce processes on the AA to maintain the flexibility and efficiency of MapReduce, while adding security into the mix. (e.g., Coopis'16 in [TBP15b]); (c) privacy-aware mobile distributed systems for efficient data aggregation in mobile participatory sensing (e.g., SSDBM'16); (d) distributed vehicular re-routing system for congestion avoidance (e.g., TMC'17 in [PSB17]); and (e) pushing personalized privacy guarantees in the processing of database queries so that individuals can disclose different amounts of information (i.e., data at different levels of accuracy) depending on their own perception of the risk (e.g., DATA'17).

Axis 4 - Economic, legal and societal issues. More transversal and multidisciplinary research perspectives are needed to address the links between economic, legal, societal and technological aspects in the personal cloud.

We especially work around the Data Portability and Users' Empowerment principles as a Privacy Incentive. 'Data Portability' was recently introduced in regulations (smart disclosure in the US, GDPR in EU), and is tightly coupled with the notion of Personal Cloud. We conduct a study of this principle in common with the DANTE Lab at UVSQ, in particular with Prof. Celia Zolynski (jurist, member of the *Conseil National du Numérique*), within the DATAIA convergence institute at Inria and in the SIHS CNRS federation at UVSQ. Our recent contributions analyze the technical conditions under which individuals can get their data back from service providers according to this data portability principle, and examine its technical feasibility and legal opportunity. We also explain how data portability favors a form of Users' Empowerment, which can be viewed as a potential privacy incentive. Our recent results are presented in multi-disciplinary papers appeared in prestigious French journals like DALLOZ and 'Revue Contrats, Concurrence, Consommation. In the context of these studies, we can also expect an institutional impact. As example, we served as scientific experts for France at the European Council regarding the draft proposal of "Directive of the European Parliament and of the Council on certain aspects concerning contracts for the online and other distance sales of goods".

This regulation proposal was opening the possibility to customers to pay using their personal data in exchange for digital content (e.g., music, books, etc), which was raising technical questions around data privacy from the European countries. We received around 70 questions from the different delegations, and spent a full day at the EU commission to present a technical view of the data management and privacy issues and answer the questions. Today, the draft proposal has evolved and the idea of formalizing the payment of digital content using personal data has been discarded.

Software and applications. The Petrus research strategy aims at materializing its scientific contributions in an single advanced hardware/software platform, called PlugDB. PlugDB is the cornerstone to help validating our research results through accurate performance measurements on a real platform, a common practice in the DB community, and target the best conferences. It is also a strong vector to federate the team, simplify the bootstrapping of new PhD or master students, conduct multi-disciplinary research and open the way to industrial collaborations and technological transfers. During the evaluation period, several important contributions of the team have been integrated and validated on PlugDB, in particular the secure search engine proposed in [Anc+15] (see Axis 1) and the secure GroupBy evaluation protocol presented in (e.g., TODS'16 in [TNP16]).

On the applications side, Petrus has developed a strong expertise in the management and protection of social-medical data, building on our past DMSP (Dossier Medico-Social Partagé) experiment in the field conducted in 2012. This expertise is being exploited to develop a dedicated healthcare and well-being personal cloud platform. During the evaluation period, we have developed close links with the Hippocad company, which is specialized in social-health care coordination for dependent people. In 2016, we answered together a call for tender 'DomYcile' launched by the Yvelines District to equip 10000 dependent people with a personal home data box to improve social-health care coordination. Our solution is based on the integration of PlugDB in the 'Humanbox' developed by the Hippocad. We won the call for tender and the DomYcile project has started in June 2017 whose goal is to design and deploy the solution in 2018. In line with this project, we are currently setting up an Inria Innovation Lab with the Hippocad company, which goal is first to successfully conduct the DomYcile project, but also industrialize this platform and deploy it at even larger scale.

4.4.2 Liste des thèses au sein de l'équipe (7)

- [10/2012 – 01/2016]
Nom, Prénom : Lallali Saliha
Titre : A scalable search engine for the Personal Cloud
Etablissement d'inscription, financement : UVSQ, ANR KISS
Encadrement : Nicolas Anciaux, Philippe Pucheral (dir.), Iulian Sandu Popa
- [09/2013 – 12/2016]
Nom, Prénom : Katsouraki Athanasia
Titre : Sharing and Usage Control of Personal Information
Etablissement d'inscription, financement : UVSQ, Bourse Inria CORDIV
Encadrement : Luc Bouganim (dir.), Benjamin Nguyen
- [11/2014 – 04/2018]
Nom, Prénom : Tran-Van Paul
Titre : Sharing Documents in the Personal Cloud Paradigm
Etablissement d'inscription, financement : UVSQ, CIFRE (CozyCloud)
Encadrement : Benjamin André (CozyCloud), Nicolas Anciaux, Philippe Pucheral (dir.)
- [10/2015 -]
Nom, Prénom : Michel Axel
Titre : Management of distributed queries under personalized privacy constraints
Etablissement d'inscription, financement : INSA CVL, Bourse Région Centre Val de Loire

- Encadrement** : Benjamin Nguyen (dir., INSA CVL), Philippe Pucheral
- [04/2016 -]
- Nom, Prénom** : Loudet Julien
- Titre** : Highly Distributed Queries on Personal Data Management Systems with strong Privacy Guarantees
- Etablissement d'inscription, financement** : UVSQ, CIFRE (CozyCloud)
- Encadrement** : Luc Bouganim (dir.), Iulian Sandu Popa, Benjamin André (CozyCloud)
- [10/2016 -]
- Nom, Prénom** : Ladjel Riad
- Titre** : Secure Distributed Computation for the Personal Cloud
- Etablissement d'inscription, financement** : UVSQ, IDEX Paris-Saclay programme IDI
- Encadrement** : Nicolas Anciaux (dir.), Philippe Pucheral, G. Scerri
- [04/2017 -]
- Nom, Prénom** : Tsolovos Dimitrios
- Titre** : Privacy-by-Design Middleware for Urban-scale Mobile Crowd-sensing
- Etablissement d'inscription, financement** : UVSQ, Bourse CORDI Inria
- Encadrement** : Nicolas Anciaux (co-dir.), Valérie Issarny (co-dir)

4.4.3 Sélection des produits et des activités de recherche

	2015	2016	2017	2018	Total
Doctorats, HDR	3	3	0	1	7
Livres, chapitres	—	1	2	—	3
Journaux internationaux	3	3	2	—	8
Journaux nationaux	1	4	3	—	8
Conférences et workshop internationaux avec actes publiés	5	3	6	—	14
Autres conférences	7	2	1	—	10
Conférences invitées	2	—	—	—	2

We give below a selection of the major publications of the team since 2015. They appeared in international conferences and journals (ranked A or A* in Core (<http://www.core.edu.au/conference-portal>) except [TSZ15c] and [JKS17] which are also prestigious but too young to be indexed in Core).

Revues Internationales (4)

- [TSZ15a] Dai Hai TON THAT, Iulian SANDU POPA et Karine ZEITOUNI. “TRIFL : A Generic Trajectory Index for Flash Storage”. In : *ACM Transactions on Spatial Algorithms and Systems* 1.2 (juil. 2015), 44 pages. DOI : 10.1145/2786758. URL : <https://hal.inria.fr/hal-01176563>.
- [TNP16] Cuong Quoc TO, Benjamin NGUYEN et Philippe PUCHERAL. “Private and Scalable Execution of SQL Aggregates on a Secure Decentralized Architecture”. In : *ACM Transactions on Database Systems* 41.3 (2016), 16 :1–16 :43. URL : <https://hal.archives-ouvertes.fr/hal-01296432>.
- [Lal+17] Saliha LALLALI, Nicolas ANCIAUX, Iulian Sandu POPA et Philippe PUCHERAL. “Supporting secure keyword search in the personal cloud”. In : *Inf. Syst.* 72 (2017), p. 1–26. DOI : 10.1016/j.is.2017.09.003. URL : <https://doi.org/10.1016/j.is.2017.09.003>.

- [PPB17] Susan Juan PAN, Iulian Sandu POPA et Cristian BORCEA. “DIVERT : A Distributed Vehicular Traffic Re-Routing System for Congestion Avoidance”. In : *IEEE Trans. Mob. Comput.* 16.1 (2017), p. 58–72. DOI : 10.1109/TMC.2016.2538226. URL : <https://doi.org/10.1109/TMC.2016.2538226>.

Conférences Internationales avec PC et Actes (5)

- [Anc+15] Nicolas ANCIAUX, Saliha LALLALI, Iulian SANDU POPA et Philippe PUCHERAL. “A Scalable Search Engine for Mass Storage Smart Objects”. In : *41th International Conference on Very Large Data Bases (VLDB)*. T. 8. 9. Sept. 2015, p. 910–921.
- [TNP15] Quoc-Cuong TO, Benjamin NGUYEN et Philippe PUCHERAL. “TrustedMR : A Trusted MapReduce System based on Tamper Resistance Hardware”. In : *Proceedings of the 23rd International Conference on Cooperative Information Systems (COOPIS)*. Août 2015. URL : <https://hal.inria.fr/hal-01185484>.
- [Ton+16] Dai Hai TON THAT, Iulian SANDU POPA, Karine ZEITOUNI et Cristian BORCEA. “PAM-PAS : Privacy-Aware Mobile Participatory Sensing Using Secure Probes”. In : *International Conference on Scientific and Statistical Database Management (SSDBM '16)*. Proceedings of the 28th International Conference on Scientific and Statistical Database Management. Budapest, Hungary, juil. 2016. DOI : 10.1145/1235. URL : <https://hal.inria.fr/hal-01426375>.
- [JKS17] Charlie JACOMME, Steve KREMER et Guillaume SCERRI. “Symbolic Models for Isolated Execution Environments”. In : *2nd IEEE European Symposium on Security and Privacy (EuroS&P'17)*. Sous la dir. de Cătălin HRIȚCU. Proceedings of the 2nd IEEE European Symposium on Security and Privacy. Paris, France : Springer, avr. 2017. URL : <https://hal.inria.fr/hal-01396291>.
- [TAP17] Paul TRAN-VAN, Nicolas ANCIAUX et Philippe PUCHERAL. “SWYSWYK : a Privacy-by-Design Paradigm for Personal Information Management Systems”. In : *Proc. of the 26th International Conference on Information Systems Development (ISD)*. Cyprus, 2017.

4.4.4 Données chiffrées

Titre	Type	Budget	date début	date fin	Statut
PerSoCloud	ANR	170 000	01/01/2017	01/01/2020	partenaire
KISS	ANR	230 000	01/12/2011	01/12/2016	partenaire
SECSI	PIA	149 000	01/05/2016	31/12/2017	partenaire
VALDO	COMUE	50 000	01/05/2015	01/09/2016	partenaire

The projects KISS, PersoCloud et SECSI have funded several years of engineers in the team since 2015.

Type de financement	nombre de doctorants	budget d'accompagnement
Bourse Inria CORDI	2	-
Bourse Région centre VDL INSA CVL	1	-
IDEX	1	-
ANR	1	-
CIFRE	2	75 000

4.4.5 Activités entre 2013 et 2015

The PETRUS team is the follow-up of the SMIS team. SMIS had three research axis : (1) Embedded Data Management, (2) Access and Usage Control Models and (3) Tamper-resistant Data Management. The SMIS initial research agenda was thus structured around embedding data management into highly constrained and tamper resistant hardware, as the current research of PETRUS is on the design of secure data management architectures based on trusted execution environment (highly constrained but also more extensible hardware components) with a clear focus on the personal cloud context. The work conducted by SMIS during the 2013-2015 period undeniably helped to define the PETRUS research program on a solid basis. PETRUS subsequently became an official Inria project team in common with UVSQ in 2017.

	2013	2014	2015	Total
HDR	1	1	0	2
Doctorats	0	0	3	3
Journaux internationaux	1	3	1	5
Conférences et workshop internationaux avec actes publiés	4	4	5	13

5 Organisation et vie de l'unité

Le laboratoire a depuis la définition de son projet scientifique, la volonté d'une organisation impliquant tous ses membres, ce que sa taille permet. Les informations et événements organisés par chaque équipe sont largement diffusés, et l'incitation à l'interaction entre les équipes est un élément essentiel de la vie de DAVID.

5.1 Pilotage, animation, organisation de l'unité

Comme cela a été indiqué dans la première section de ce document, le laboratoire est organisé en trois équipes collaborant notamment par leur implication dans des actions de recherche transverses en lien avec le projet scientifique du laboratoire et les thèmes développés au sein de la fédération de recherche SIHS. Chaque équipe est dirigée par un membre du laboratoire qui a été désigné lors de la création de DAVID en septembre 2015.

Selon le règlement intérieur dont le laboratoire s'est immédiatement doté, le directeur de DAVID est nommé par l'université après consultation de l'ensemble des membres, pour une période de cinq années renouvelable une fois. La gouvernance s'appuie sur un bureau paritaire de six membres du laboratoire, trois de rang A et trois de rang B, bureau élu par l'ensemble des membres sur proposition du directeur. Un responsable des doctorants, élu par ses pairs, assure un relais entre direction du laboratoire et doctorant(e)s. Afin d'assurer une gouvernance collégiale dans l'esprit dans lequel le projet du laboratoire a été défini, et considérant le nombre de membres titulaires, le choix a été fait d'appuyer la gouvernance par des assemblées générales des membres (avec un rythme d'au moins une tenue de cette assemblée tous les un mois et demi).

Enfin, chaque équipe organise de façon plus ou moins régulière des séminaires de recherche avec des participants invités, séminaires auxquels sont conviés tous les membres de DAVID.

5.2 Parité

La question de la parité est pleinement prise en compte dans l'animation et la gouvernance de DAVID. On l'a vu, le bureau du laboratoire est paritaire. De plus, la proportion d'un tiers de femmes et de deux tiers d'hommes se vérifie dans la population étudiante de nos masters, des doctorant(e)s du laboratoire, des maîtres de conférences et chargés de recherche et des professeurs et directeurs de recherche. Enfin, la direction du laboratoire informe régulièrement les membres de DAVID de toute manifestation organisée par l'université concernant la parité et les incite à y participer.

5.3 Intégrité scientifique

En matière d'intégrité scientifique, le laboratoire s'appuie pleinement sur les instances de l'UVSQ. En effet, la Direction des Affaires Juridiques et Institutionnelles (DAJI) peut intervenir en matière d'intégrité scientifique et d'éthique selon deux modalités : d'une part, par le biais de la participation à la rédaction et par la validation de dispositions applicables en la matière ; d'autre part, en cas de poursuites et de sanctions pour non-respect des règles visant à assurer le respect de l'intégrité scientifique et de l'éthique. En particulier, la sensibilisation s'appuie :

- sur la mise en place de structures internes dont la mission consiste notamment à garantir l'intégrité scientifique et l'éthique en matière de recherche. Ainsi, les statuts de l'UVSQ prévoient l'existence d'un comité d'éthique de l'établissement.
- sur l'adoption de textes réglementaires, publiés sur le site web de l'UVSQ, à savoir notamment charte anti-plagiat, règlement des études de l'université Paris Saclay (référence au plagiat). De façon générale, les laboratoires disposent d'un règlement intérieur prévoyant des dispositions relatives, a minima, à la propriété intellectuelle et à la diffusion des résultats.

- sur l'intégration de stipulations relatives à la propriété intellectuelle et au partage des connaissances dans les conventions signées par le Président de l'université.
- sur la nomination, depuis 2012, d'un Correspondant Informatique et Libertés (la responsable de la Direction des Affaires Juridiques) qui a notamment pour mission de connaître les traitements informatiques de données à caractère personnel (registre) et de faire connaître, au sein de l'établissement, les règles relatives à l'utilisation de celles-ci (régime déclaratoire ou régime d'autorisation, selon la nature sensible ou non des données traitées; transfert de données; vidéosurveillance, etc).
- sur l'existence d'une procédure relative à l'autorisation de cumul d'activités (problématiques notamment de conflits d'intérêts).

5.4 Protection et sécurité

Chaque année, la direction du laboratoire établit le document unique des risques professionnels avec les services compétents de l'université, document qui peut être consulté par tous les membres de DAVID. Par ailleurs, la protection des données est un souci constant du laboratoire, ne serait-ce que parce qu'elle est inhérente aux activités de recherche des équipes ADAM et PETRUS.

6 Analyse SWOT

6.1 Forces

- La cohérence et le rayonnement d'un projet scientifique interdisciplinaire impliquant tous les membres du laboratoire dont la taille est bien adaptée au projet.
- La complémentarité et l'interaction entre Algorithmes et Données.
- Le positionnement des activités de recherche du laboratoire sur plusieurs thèmes de recherche liés à des défis sociétaux actuellement majeurs : la mobilité urbaine, la protection des données personnelles, l'énergie et l'environnement.
- La participation collégiale de tous les membres de DAVID à la vie du laboratoire.
- La qualité scientifique des activités menées et la capacité du laboratoire à des partenariats et contrats de recherche, notamment destinés au financement de contrats doctoraux.

6.2 Faiblesses

- Le peu de moyens et l'absence de soutien politique de la part de la tutelle UVSQ.
- Le morcellement des origines de financement des activités de recherche et les efforts que cela implique pour tous les membres de DAVID.
- Le taux d'encadrement très élevé au sein des départements d'informatique de l'UFR des Sciences et de l'IUT de Vélizy qui impliquent des services d'enseignement très lourds aux membres de DAVID.
- La difficulté actuelle dans ce contexte pour les membres de DAVID d'investir dans la participation de notre laboratoire à des projets de recherche contractuels au niveau européen, qui plus est dans le contexte d'un laboratoire qui n'a que deux ans et qui a du depuis 2015 assoir son activité, sans un appui suffisant de la tutelle dans le suivi administratif des projets.

6.3 Opportunités

- La structuration des activités de recherche de l'université Paris-Saclay et des moyens qu'elle offre, écosystème au sein duquel DAVID positionne son identité et la qualité de ses activités.

- Les perspectives de collaboration offertes par l'interaction croissante de DAVID avec des acteurs territoriaux au niveau des communautés d'agglomération, du département des Yvelines et de la région IdF.

6.4 Menaces

- La baisse de motivation des membres de DAVID dans l'investissement pour l'obtention des moyens pour la recherche, dans l'éclatement des appels à projets et structures de financement. L'énergie dépensée l'est au détriment des activités de recherche elles mêmes.
- L'absence de perspectives à court et moyens termes d'obtention de supports de postes pour l'accroissement, le renouvellement et le rajeunissement du laboratoire.

7 Projet scientifique à cinq ans

Le projet scientifique du laboratoire DAVID, exposé en section 1.2.1, définit la feuille de route du laboratoire de 2016 à 2021. Le projet scientifique à horizon de cinq ans que nous présentons ici consiste donc en une mise à jour de cette feuille de route, principalement due à l'implication du laboratoire dans la nouvelle fédération SIHS, et sa projection jusqu'en 2024. Le Laboratoire maintiendra également sa stratégie de recherche, en particulier par le renforcement de son partenariat avec les territoires proches et les *startups innovantes*. L'ambition de DAVID est enfin d'asseoir au sein de l'écosystème de l'université Paris-Saclay, et en particulier au sein de son département STIC, sa position d'acteur incontournable sur les thèmes liés à la ville numérique.

Dans ce but, le projet scientifique du laboratoire se structurera notamment dès 2019 autour de trois axes de recherche transverses, chacun impliquant au moins deux équipes de DAVID.

Axe transverse "Mobilités et *smartcities*". Cet axe de recherche transverse s'inscrit pleinement dans l'action des trois équipes de DAVID dans le thème "Mobilité" de la fédération, thème co dirigé par Leïla Kloul (ALMOST) et Yoan Demoly (UMR PRINTEMPS). La ville numérique doit permettre à un territoire urbain ou péri-urbain de s'adapter dynamiquement, à différentes échelles de temps, aux besoins de mobilités et de transport des citoyens et des acteurs publics et économiques, sans priver les élus de leur capacité de gouvernance, en anticipant l'évolution de l'organisation du territoire et de ses activités. Il s'agira donc de développer des méthodes de compréhension de la mobilité d'un territoire et de ses usages à partir de l'intégration des données quantitatives et qualitatives disponibles, de pouvoir dimensionner et piloter au mieux les moyens, infrastructures et équipements de mobilité, équipements éventuellement partageables, autonomes et adaptables à la logistique urbaine (équipement dont on devra garantir la fiabilité), et de proposer des outils permettant d'optimiser, au sens de la mobilité, l'organisation d'un territoire (dimension spatiale) et de ses activités (dimension temporelle). Il est donc essentiel de pouvoir disposer des solutions numériques permettant de modéliser, simuler et prédire la mobilité urbaine, en intégrant les particularités sociologiques propres à chaque territoire.

Axe transverse "Données, justice et société". En lien avec le thème "Société réelle / société virtuelle" de la fédération, co-animé par Mélanie Clément-Fontaine (EA DANTE) et Philippe Pucheral (PETRUS), cet axe transverse auquel participera les trois équipes de DAVID vise à la fois les aspects juridiques et sociaux du foisonnement des données, notamment personnelles et privées, et l'impact de la numérisation et de l'ouverture des données de justice et sociologiques. En particulier, le contrôle de l'utilisation et du stockage et de la durée des données privées par leur propriétaire, et le statut juridique de la donnée sont deux sujets d'études très actuels.

Par ailleurs, les initiatives d'utilisation d'algorithmes et de modèles prédictifs sur des corpus de données juridiques pour supporter la prise de décision se multiplient, tant dans un contexte académique qu'industriel. L'aide à la décision juridique pose un certain nombre de problèmes, parmi lesquels d'une part l'analyse et la classification de diverses sources de données juridiques, et d'autre part l'extraction de connaissances par utilisation de modèles prédictifs, par recherche de similarités ou encore par extraction de règles d'association.

Enfin, la recherche de cohérence à partir de données sociologiques individuelles, notamment dans un contexte professionnel ou d'enseignement, est un enjeu de recherche majeur afin de mieux comprendre les usages et pratiques professionnelles et de proposer des approches capables d'établir des grilles de lecture, voire des nomenclatures, les mieux adaptés à ces pratiques.

Axe transverse "Energie et ville durable". En lien avec le thème "Territoire et environnement" de la fédération, co-animé par Jacques de Maillard (UMR CESDIP) et Karine Zeitouni (ADAM), cet axe transverse vise à la fois la gestion optimisée de la production et de la distribution d'énergie en

zone urbaine et l'exposition des habitants aux conséquences environnementales de l'utilisation de cette énergie. L'objectif est de définir et d'évaluer des outils numériques permettant un pilotage optimisée de *smartgrids* (production et distribution d'électricité), en intégrant différentes données permettant l'anticipation des usages, *smartgrids* dont devra aussi être assurée une maintenance prédictive. Il est donc essentiel de pouvoir comprendre et anticiper et le comportement énergétique de l'usage (à son domicile, dans sa mobilité, dans son activité professionnelle) par une analyse de données quantitatives et qualitatives, afin notamment de définir des méthodes d'apprentissage permettant aux infrastructure de s'adapter aux usages en minimisant le coût énergétique.

L'exposition de la population aux conséquences environnementales de la consommation énergétique est une préoccupation centrale de la gouvernance des zones urbaines et périurbaines où les risques sanitaires de la pollution, notamment atmosphérique, sont extrêmement alarmants. Néanmoins, l'exposition individuelle réelle à cette pollution est encore mal connue, et l'objectif sera de pouvoir intégrer les différentes données disponibles pour analyser, comprendre et prédire cette exposition.

Sont maintenant précisées les évolutions du projet scientifique de chaque équipe, principalement en lien avec ces axes transverses.

7.1 Equipe ADAM

Dans la continuité des travaux menés autour de l'intégration de données, du traitement des requêtes et de l'extraction de connaissances, le projet scientifique de l'équipe ADAM s'oriente principalement vers (i) l'extraction de schéma pour des données massives, (ii) la modélisation et la fouille de données de capteurs, et (iii) l'intégration et la fouille de données pour la prédiction dans le domaine du Droit. Ce projet scientifique prend ses racines dans les travaux en cours au sein de projets collaboratifs de l'ANR, du programme Européen H2020 ou encore de la fédération SIHS. Nous illustrons par la suite ces volets du projet scientifique.

L'absence de schéma d'une source de données est un frein réel à son utilisation. Nous avons proposé une approche de découverte automatique du schéma d'une source de données faiblement structurée ou non structurée, nous nous proposons de traiter le problème du passage à l'échelle et de proposer des algorithmes de regroupement d'entités en classes adaptés aux données massives, et également de proposer des algorithmes permettant de maintenir le schéma cohérent au fil des évolutions survenues dans les données.

"Energie et ville durable". Dans le cadre du projet ANR Polluscope, nous nous intéressons à la modélisation et la fouille de données de capteurs. L'objectif de ce projet est d'étudier l'exposition individuelle à la pollution atmosphérique au moyen de collecte participatives par le biais de capteurs nomades à bas cout. Pour ce faire, une plateforme de collecte et d'analyse des données de capteurs mobiles est mise en place. L'avènement des capteurs à bas coût génère de nouvelles opportunités pour observer et analyser l'état ou l'environnement d'un lieu ou d'objets mobiles. Leur exploitation aujourd'hui est souvent limitée à la détection de seuils d'alertes. Mais ils offrent un réel potentiel pour comprendre et inférer, voire prédire des événements complexes. Cela nécessite de tenir compte de leur caractéristiques très particulières. Il s'agit de séries de mesures horodatées en flux continu, souvent bruitées et comportant des valeurs manquantes dues à des pertes de signal. Elles sont généralement auto-corrélées dans le temps et dans l'espace, si l'on exclut le bruit. Par conséquent, nous cherchons à capturer ces spécificités dans le modèle de données et leur manipulation. Pour cela, nous proposons de substituer les données brutes de capteurs par un modèle fonctionnel qui capture mieux leur évolution, élimine le bruit et permet leur compression. Toutefois, la question de la qualité des données depuis le capteur jusqu' à la restitution des résultats reste posée. Nous nous intéressons aux facteurs de qualité de l'ensemble du processus d'acquisition incluant, outre la mesure, le traitement embarqué, la transmission, l'intégration, l'anonymisation et l'exploitation de ces données. L'objectif est l'évaluation,

l'amélioration et la restitution de la qualité afin de fournir des indicateurs pertinents du phénomène mesuré à une granularité spatiale et temporelle fine.

"Mobilités et smart cities". Dans le projet Polluscope, nous visons également le développement de modèles analytiques permettant d'extraire des règles ou des régularités des observations et mesures collectées en lien avec le contexte géographique. Ceci comprend la découverte de profils types d'exposition, de caractérisations pour des micro-environnements, et des corrélations avec des activités ou des risques sanitaires. D'un point de vue méthodologique, il s'agit de proposer des méthodes de fouille de données temporelles (tendances, apprentissage incrémental sur des flux), spatiales (micro-environnements) et spatiotemporels (détection d'activités). Les algorithmes devront s'adapter à notre modèle fonctionnel de données et passer à l'échelle.

Par ailleurs, nous entendons étendre nos travaux sur la fouille de trajectoires dans trois directions : la prédiction de mobilité; les trajectoires sémantiques; et les trajectoires en milieux intérieur. Nous devons notamment prendre en compte des granularités différentes et une imprécision de données, des dimensions multiples et l'aspect massif de ces données. Cet axe fait écho à différents projets qui ont été initiés récemment, dont une thèse en collaboration avec le Brésil, le projet PATRIMA Trajectoire et le projet H2020-RISE Marie-Curie MASTER.

"Données, justice et société". Nous avons initié une collaboration avec le laboratoire DANTE sur le thème de la justice prédictive. L'aide à la décision juridique pose un certain nombre de problèmes, parmi lesquels l'analyse de diverses sources, et l'extraction de connaissances par des modèles prédictifs, par recherche de similarités ou encore par extraction de règles d'association. Les approches existantes ne prennent pas en compte la caractérisation de la provenance, qui est importante dans ce contexte. Nous cherchons à intégrer les données dont découle la prédiction ainsi que la qualité intrinsèque de ces données et son influence sur l'algorithme ou le modèle utilisé pour la prédiction. Il s'agit de (i) construire un espace de données virtuel par l'analyse de sources de données juridiques, leur intégration et leur enrichissement, (ii) d'extraire des connaissances à partir de cet espace pour supporter l'aide à la décision juridique et la prédiction de ces décisions et (iii) de fournir un support pour permettre l'explication de cette connaissance et pour caractériser le lien entre son degré de confiance et la qualité des données utilisées pour la produire. Ces travaux seront menés en collaboration avec l'équipe ALMOST.

7.2 Equipe ALMOST

Le projet scientifique de l'équipe ALMOST continuera à s'appuyer sur les activités de recherche dans ses domaines d'expertises (modélisation et algorithmique), en particulier l'approximabilité de problèmes de graphes, la complexité du calcul des équilibres de jeux stochastiques, les techniques formelles pour l'analyse de la sûreté de fonctionnement et l'analyse de réseaux de files d'attente. Il s'inscrira aussi en grande partie en cohérence avec les trois axes transverses indiqués plus haut sur lesquels elle est déjà active.

"Mobilités et smartcities" : mobilité fluide et sûre, logistique. L'équipe mène, en collaboration avec l'équipe ADAM, des activités pluri-disciplinaires avec les SHS au sein de la fédération de recherche SIHS du CNRS pour la compréhension et la prédiction des usages et trajectoires de mobilité en fonction des spécificités des populations. En lien avec l'institut Védécom et l'IRT Systemix, l'équipe vise le développement d'algorithmes et de méthodes d'apprentissage pour la gestion dynamique optimisée de flottes de véhicules autonomes, et pour la sûreté de fonctionnement de leur conception. Ce projet de recherche a aussi pour objectif la définition et l'évaluation de modèles et d'algorithmes de simulation et de prédiction de la mobilité, en particulier pour l'organisation d'un territoire et de ses activités autour de la mobilité des personnes et la logistique urbaine. Avec ces mêmes partenaires, l'équipe va initier

une activité de recherche visant l'utilisation de méthodes d'apprentissage automatique distribuées pour la régulation du trafic routier urbain. Enfin, en lien avec la *startup* DCBrain, seront développées des méthodes d'optimisation pour des problèmes de logistique alliant production, transport, stockage et distribution de produits manufacturés.

"Données, justice et société" : algorithmique pour l'analyse de données et prédiction. En lien avec cet axe, "Données, justice et société", l'équipe a commencé une étude prospective sur l'analyse algorithmique de données sociologiques. Elle a initié un groupe de travail dans le cadre du CNIS (Conseil National de l'Information Statistique) qui vise à la rénovation de la PCS (Nomenclature des Professions et Catégories Socio-professionnelles) pour la rendre plus lisible, d'étendre ces usages et de faciliter sa production et son utilisation. L'équipe participera aussi à terme avec l'équipe ADAM à l'étude d'algorithme de prédiction dans le cadre de la jurisprudence prédictive.

"Energie et ville durable" : maîtrise de la distribution et de la consommation d'énergie. Concernant cet axe, l'équipe a initié en partenariat avec le LINEACT du CESI une étude d'algorithmes d'apprentissage pour l'adaptation de la politique énergétique d'un bâtiment ou d'un groupe de bâtiments aux comportements de ses usagers. L'équipe continuera aussi, en particulier en partenariat avec l'entreprise DCBrain, ses activités en optimisation et en sûreté de fonctionnement sur la configuration et la maintenance prédictive de réseau de distribution d'énergie, en particulier par l'étude de variantes associées du problème de Steiner.

Enfin, l'équipe maintiendra son activité de recherche et son partenariat avec d'autres laboratoires de l'éco-système Paris-Saclay en bio-informatique, notamment sur la prédiction de la structure tridimensionnelle de molécules d'ARN, et en chémo-informatique en lien avec la physico-chimie des matériaux et la conception de cages moléculaires.

7.3 Equipe PETRUS

Le programme scientifique de l'équipe PETRUS s'articule autour de 4 axes de recherche, chacun intrinsèquement liés aux défis du cloud personnel sécurisé. Ce programme, traitant des données et des algorithmes sous l'angle architectural, s'inscrit naturellement dans celui de DAVID tel que décrit en Section 1.2.1 et ci-dessus. Dans cette section, nous ne revenons pas sur les axes de recherche de l'équipe PETRUS (nous renvoyons à la Section 4.4 pour une description plus détaillée des axes de recherche), ni sur l'ensemble des perspectives scientifiques associées à chacun de ces axes. En revanche, nous détaillons plus particulièrement deux de nos perspectives de recherche qui nous semblent particulièrement bien s'inscrire dans DAVID et la fédération SIHS. La première perspective s'inscrit dans les problématiques multidisciplinaires investiguées par PETRUS (Axe 4) au sein du thème "Société réelle et société virtuelle" de la fédération SIHS et de l'axe "Données, justice et société" de DAVID. La seconde perspective, purement scientifique, s'inscrit dans l'axe "Mobilités et smartcities" de DAVID.

"Données, justice et société" : responsabilités, cloud personnel et RGPD. L'équipe PETRUS a développé une collaboration dans la durée avec les juristes du laboratoire DANTE, qui débouche aujourd'hui sur des contributions communes (voir Section 4.4). Les enjeux juridiques et technologiques associés au cloud personnel, méritent en effet d'être étudiés à la lumière des deux disciplines. Nos études ont jusqu'ici porté sur le droit à la portabilité des données personnelles, édicté dans le RGPD, qui consacre un droit de l'individu à la récupération de ses données personnelles. Ce droit ouvre des perspectives d'empowerment et de développement de nouveaux usages tels que le personal-big-data (croisement de données personnelles de nature diverses d'un même individu) et le big-personal-data (croisement de données personnelles de plusieurs individus afin d'en tirer un bénéfice communautaire ou sociétal), réalisés sous le contrôle de l'individu. Le cadre juridique se limite pour l'instant à prescrire

ce droit à la portabilité tout en reconnaissant qu'il s'accompagne de nouvelles formes de responsabilités, sans toutefois en préciser l'articulation entre individus, fournisseurs de plateformes et fournisseurs de services et sans prendre en compte la variété des solutions techniques de cloud personnels. Nous envisageons d'analyser ce double mouvement, juridique et technique, afin d'établir plus précisément les responsabilités inhérentes à l'empowerment, en conformité avec les notions juridiques existantes de responsable de traitement, de sous-traitants et de tiers, et les exemptions du RGPD, ainsi qu'à envisager la préconisation de plateformes offrant un niveau de responsabilité graduel aux individus, de manière adaptée à la technologie. A notre connaissance, aucune étude scientifique ni aucune réflexion de fond n'ont porté sur ces questions (fondamentales mais certes très récentes), ni dans la communauté des juristes ni dans celle des informaticiens ni encore moins de façon croisée entre disciplines.

"Mobilités et smartcities" : architecture privacy-by-design pour le crowd-sensing. Etudier le problème du cloud personnel impose de s'intéresser à la tension existante entre la sécurité de la gestion de données personnelles d'une part, et l'extensibilité des traitements à réaliser sur ces données d'autre part. PETRUS analyse notamment l'influence des composants de l'informatique de confiance (trusted computing) sur cette tension, et leur intégration dans une architecture de cloud personnel. La première étape cette analyse, déjà en cours dans l'équipe, consiste à définir une architecture logique permettant de réguler les grandes étapes du cycle de vie des données dans un cloud personnel, à savoir la collecte, le stockage, le partage, les calculs et la destruction. Une seconde étape, qui fait l'objet d'une perspective de recherche importante, consiste à envisager une architecture adaptée plus particulièrement à la collecte participative (crowd-sensing). Les applications développées dans ce contexte ciblent les domaines variés de la ville intelligente, de la santé au transport en passant par la démocratie participative et l'implication citoyenne, où le respect de la vie privées constituent un enjeu important pour les villes et le citoyen urbain. Dans ce cadre, l'équipe PETRUS collabore avec l'équipe ADAM sur la collecte participative de données spatiales à grande échelle avec des garanties fortes de protection de la vie privée, ainsi qu'au sein du CityLab (Inria Project Lab regroupant plusieurs équipes Inria) sur une nouvelle architecture "privacy-by-design" de crowd-sensing reposant sur des composants de l'informatique de confiance. La continuation et l'approfondissement de ces travaux, à travers l'étude d'architectures distribuées de cloud personnel pour la collecte participative de données de mobilité, constitue pour l'équipe une perspective particulièrement attrayante.

8 Annexe : participation à des comités de programmes

Voici la liste des des conférences internationales pour lesquelles des membres de DAVID participent aux comités de programmes ou de pilotage (Steering comité).

1. ADBIS 2016, 2017 : East-European Conference on Advances in Databases and Information Systems.
2. AICCSA 2015 : ACS/IEEE International Conference on Computer Systems and Applications.
3. AKBC 2017 : Workshop on Automated Knowledge Base Construction.
4. APWEB-WAIM 2017 : APWeb-WAIM Joint Conference on Web and Big Data
5. ASMTA 2016, 2017 : KES Conference on Agent and Multi-agent Systems : Technologies and Applications.
6. Astroinformatics 2017 Astroinformatics Symposium @ EWASS.
7. BigSkyEarth 2016 : BigSkyEarth Conference on Education in Big Data Era @ AstroInfo.
8. CIKM 2017 : International Conference on Information and Knowledge Management.
9. CLOSER 2017 : International Conference on Cloud Computing and Services Science.
10. CoopIS 2017, 2018 : International Conference on Cooperative Information Systems.
11. DATA 2016, 2017 : International Conference on Data Science, Technology and Applications.
12. DAWAK 2015 : International Conference on Big Data Analytics and Knowledge Discovery.
13. DBKDA 2016 : International Conference on Advances in Databases, Knowledge, and Data Applications.
14. EDBT 2015, 2016, 2017 : International Conference on Extending Database Technology.
15. EPEW 2015, 2016, 2017 : European Performance Engineering Workshop.
16. ICDE 2017 : IEEE International Conference on Data Engineering.
17. ICIT 2017 : International Conference on Industrial Computing.
18. ICQT 2015, 2017 : International Workshop on Internet Charging and QoS Technologies.
19. ICSOC 2017 : International Conference on Service Oriented Computing.
20. ICT4AWE'16 : International Conference and Communication Technologies for Ageing Well and e-Health.
21. IMBSA 2017 : The 5th International Symposium on Model based Safety Analysis. (Steering Committee).
22. ISCIS 2016 : International Conference on Information Systems.
23. MASCOTS 201, 2017 : IEEE International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems.
24. MCSMS 2017 : IEEE International Workshop on Mobile Cloud Computing systems, Management, and Security.
25. MDM 2016, 2017 : IEEE International Conference on Mobile Data Management.
26. MEDI 2017 : International Conference on Model and Data Engineering.
27. Mobile Cloud 2016, 2017 : IEEE International Conference on Mobile Cloud Computing, Services, and Engineering.
28. MobilWare 2015, 2016 : International Conference on MOBILE Wireless MiddleWARE, Operating Systems, and Applications.

29. MSIS 2015 : International Workshop on Mobile Spatial Information Systems.
30. NLDB 2016, 2017 : International Conference on Applications of Natural Language to Information Systems.
31. PASM 2016, 2017 : International Workshop on Practical Applications of Stochastic Modeling.
32. SIGMOD 2017, 2018 : ACM International Conference on Management of Data.
33. SIGSPATIAL 2015, 2016 : ACM International Conference on Advances in Geographic Information Systems.
34. SSTD 2017 : International Symposium on Spatial and Temporal Databases.
35. TDLSG 2017 : International Workshop on Time-Dependent Large Scale Graphs @ ECML-PKDD (co-organisation)
36. Valuetools 2015, 2016, 2017 : International Conference on Performance Evaluation Methodologies and Tools.
37. VLDB 2017 : International Conference on Very Large DataBases.
38. WebDB 2015 (Mention "Outstanding Reviewer"), 2016 : International Workshop on the Web and Databases (WebDB) @ACM SIGMOD.
39. Wiopt 2017 : International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks.