
CURRICULUM VITAE

de

FATIHA SAÏS

MAÎTRE DE CONFÉRENCES - HDR

LRI - CNRS & UNIVERSITÉ PARIS SACLAY, ORSAY, FRANCE

Laboratoire de Recherche en Informatique, UMR 8623 CNRS
Université Paris Sud, Bâtiment Ada Lovelace 650, F-91405, Orsay Cedex, France
Contact : tél. : +33 (0)1 69 15 66 29 – Web : <http://www.lri.fr>

Table des matières

1	Description de l'activité d'enseignement	6
1.1	Synthèse des enseignements	6
1.2	Tâches et responsabilités liées à la pédagogie	6
2	Activités de Recherche	8
2.1	Contexte scientifique	8
2.2	Description des travaux antérieurs	8
2.2.1	Liage de données	8
2.2.2	Invalidation de liens d'identité	9
2.2.3	Détection de liens d'identité contextuels	9
2.2.4	Découverte automatique de clés	10
2.2.5	Enrichissement de données	10
2.2.6	Génération de méta-faits temporels dans les graphes de connaissances	11
2.2.7	Modélisation et utilisation des informations personnelles	11
2.3	Projet de Recherche	11
3	Encadrement doctoral et scientifique	12
3.1	Encadrement doctoral	12
3.2	Encadrement de post-doctorants et d'ingénieurs	13
3.3	Encadrement d'étudiants en Master recherche	14
3.4	Encadrement d'étudiants en Magistère, Master 1 et d'ingénieurs	15
4	Publications et production scientifique	16
4.1	Revue internationale avec comité de lecture	16
4.2	Revue nationale	16
4.3	Chapitres de livre	16
4.4	Conférences internationales avec comité de lecture	16
4.5	Workshops internationaux avec comité de lecture	18
4.6	Conférences nationales avec comité de lecture	18
4.7	Workshops nationaux avec comité de lecture	19
4.8	Communications scientifiques	20
5	Implication dans les projets de recherche	21
5.1	Responsabilité et coordination de projets scientifiques	21
5.2	Participation à des projets scientifiques	22
6	Animation scientifique	24
6.1	Implication dans les GdRs nationaux	24
6.2	Edition de numéros spéciaux dans des revues	24
6.3	Organisation de conférences et de workshops	24
6.4	Présidence de sessions dans des conférences nationales et internationales	24
7	Evaluation de la recherche	25
7.1	Expertise de projets de recherche	25
7.2	Membre de jurys de thèse	25
7.3	Membre de comités de sélection	25
7.4	Participation à des comités de programme	25

FATIHA SAÏS

MAÎTRE DE CONFÉRENCES – HDR EN INFORMATIQUE
LRI –Laboratoire de Recherche en Informatique
CNRS 8623, Université Paris Sud, Paris Saclay
Equipe LaHDAK – Données et Connaissances Massives et Hétérogènes
Département Informatique de l’IUT d’Orsay (Enseignement)

*Née le 29 mai 1979,
à Ighil-Boukiassa (Algérie)
Nationalité Française
Mariée, 2 enfants*

COORDONNÉES

Recherche :

LRI (Laboratoire de Recherche en Informatique)
CNRS UMR 8623, Université Paris Saclay
Université Paris Sud, bât. 650-Ada Lovelace
91405 Orsay Cedex

E-mail : sais@lri.fr
Tél : 01 69 15 68 42

Page web : www.lri.fr/~sais

Enseignement :

IUT d’Orsay (Université Paris-Sud)
Département informatique
Plateau du Moulon
91405 Orsay Cedex

CURSUS UNIVERSITAIRE

- 2019 : **HDR (Habilitation à Diriger des Recherches) en informatique de l’Université Paris Sud, Orsay.**
 - Intitulé : **Knowledge Graph Refinement : Link Detection, Link Invalidation, Data Enrichment and Key Discovery**
 - Marraine d’HDR : Chantal Reynaud
 - Date de soutenance : 20 Juin 2019
 - Composition du Jury :

Oscar Corcho	– Professeur, Universidad Politécnica de Madrid (Rapporteur)
Fabien Gandon	– Directeur de Recherche, INRIA Sophia-Antipolis (Rapporteur)
Hala Skaf Molli	– Maître de Conférences HDR, Université de Nantes (Rapporteur)
Sarah Cohen-Boulakia	– Professeur, Université Paris Sud (Examinatrice et Présidente du jury)
Juliette Dibie	– Professeur, AgroParisTech (Examinatrice)
Chantal Reynaud-Delaître	– Professeur, Université Paris Sud (Marraine)
- 2007 : **Doctorat en Informatique de l’Université Paris Sud, Orsay.**
 - Intitulé : **Intégration Sémantique de Données guidée par une Ontologie**
 - Encadrantes de thèse : Marie-Christine Rousset et Nathalie Pernelle
 - Date de soutenance : 07 décembre 2007
 - Mention : Très honorable
 - Composition du Jury :

Patrice Buche	– Ingénieur de Recherche HDR, AgroParisTech (Examineur)
Christine Froidevaux	– Professeur, Université Paris Sud (Présidente du Jury)
Hélène Gagliardi	– Maître de Conférences, Université Paris Sud (Examinatrice)
Daniel Kayser	– Professeur, Université Paris Nord (Rapporteur)
Alain Léger	– Directeur de Recherche, France Télécom R&D (Membre invité)
Nathalie Pernelle	– Maître de Conférences, Université Paris Sud (Co-directrice)
Jean-Marc Petit	– Professeur INSA-Lyon, LIRIS Université Lyon (Rapporteur)
Marie-Christine Rousset	– Professeur, Université Joseph Fourier (Directrice)

- 2003-2004 : Master 2 recherche, Parcours I3 (Information, Interaction, Intelligence) de l'Université Paris Sud, Orsay. Mention : Bien
Titre : *Transformation d'informations structurées en documents XML guidée par une ontologie.*
Encadrants : *Olivier Haemmerlé, Nathalie Pernelle et Hélène Gagliardi.*
- 2001-2003 : Maîtrise d'Informatique de l'Université Aix-Marseille 1. Mention : Assez-Bien
- 2001-2002 : Licence d'Informatique de l'Université Aix-Marseille 1. Mention : Assez-Bien
- 1999-2001 : DUT d'Informatique, au département Informatique de l'IUT de Lens à l'Université d'Artois.
- 1998-1999 : Première année de Tronc Commun en Sciences Exactes, Technologie et Informatique à l'Université de Mouloud Mammeri de Tizi-Ouzou (Algérie)

CURSUS PROFESSIONNEL

- 09/2008–présent : **Maître de Conférences** en informatique à l'Université Paris Sud, attachée à l'IUT d'Orsay pour l'enseignement et au Laboratoire de Recherche en Informatique (CNRS – Université Paris Sud) pour la recherche.
- 01/2008–08/2008 : **Post-doctorante** en informatique au Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM), Université Montpellier 2.
- 10/2004–12/2007 : **Doctorante** en informatique dans l'équipe IASI du Laboratoire de Recherche en Informatique de l'Université Paris Sud et membre du projet Gemo de l'INRIA-Saclay.

SYNTHÈSE DE L'ACTIVITÉ D'ENSEIGNEMENT

Thèmes Bases de données, Conception Orientée Objet, Conception des Systèmes d'Informations, Données du Web, Web Sémantique, Intelligence Artificielle, Systèmes et Réseaux.

Responsabilité de formations Co-responsabilité de la licence professionnelle SRSI (depuis septembre 2018), co-responsabilité et montage du DUT Air France (2014-2015) et de la licence professionnelle PRISM Air France (2015-2016), en formation continue.

Charges collectives Pour le département informatique : correspondante des relations internationales (2011-2013) et responsable de l'organisation du forum « Etudiants–Entreprises » (2009–2015).

Présence dans les instances Membre élue au « Conseil d'IUT » (2014–2018), membre élue au « Conseil du Département Informatique » de l'Université Paris Sud (2017–2019), et membre élue au « Comité de Pôle », de l'école doctorale STIC de l'Université Paris Saclay, depuis décembre 2019.

Enseignements

Thèmes	Niveaux	Etablissements
Bases de données	L1, L2, L3	IUT d'Orsay
Conception Orientée Objet	L1, L2	IUT d'Orsay, UFR Sciences Univ. Paris Sud
Conception des Systèmes d'Informations	L1	IUT d'Orsay
Systèmes et Réseaux	L1, L2	IUT d'Orsay
Programmation Web/Données du Web	L3, M2	IUT d'Orsay, Université Paris Saclay
Web Sémantique	M2	Université Paris Saclay, Université Paris Dauphine
Intelligence Artificielle et Raisonnement	Bac+5	Polytech Paris Sud

SYNTHÈSE DE L'ACTIVITÉ DE RECHERCHE

Mots clés Représentation de Connaissances, Raisonnement, Web sémantique, Graphes de Connaissances, Ontologies.

Thèmes Intégration de données RDF, découverte de connaissances, liage de données, fusion de données, gestion de l'identité dans le Web de données, gestion contextuelle des informations personnelles.

Encadrement Co-encadrement de 5 thèses dont 2 en cours (2 thèses à 50% et 3 thèses à 33%), 3 stages post-doctoraux (2 à 50% et 1 à 25 %), 3 ingénieurs (dont un à 100%), 11 stages de M2 recherche et plus d'une dizaine de stage de TER et de magistère M1.

Projets responsabilité scientifique de 6 projets (1 PSPC-BPI, 1 ANR, 1 industriel (avec IBM), 1 PEPS CNRS, 1 projet Labex, 1 projet IRS). Participation à 7 autres projets de recherche (2 ANR, 2 industriels (avec Thales, FT/Orange), 1 EIT europe, 1 IDEX Paris Saclay, 1 CNRS Mastodon).

Publications

J'ai co-écrit plus de 60 articles publiés dans des chapitres de livres, revues, conférences et workshops internationaux et francophones dont certains sont parus dans des journaux et conférences de haut niveau comme : Journal of Web Semantics, Knowledge Based Systems Journal, International Semantic Web Conference (ISWC) et AAAI Conference on Artificial Intelligence. Mes publications sont accessibles sur ma page Web :

<https://www.lri.fr/~sais/publication.html>

	International	National
Revue avec comité de lecture	4, (+1 en révision), (+1 soumise)	3
Chapitres de livres	3	
Conférences avec comités de lecture	17	16
Workshops, posters	9	4
Tutoriels	2	3

Valorisation

J'ai mené différentes actions de valorisation de mes travaux de recherche. Je cite ici :

- Participation au projet de Startup DataForYou soutenu par la SATT Paris Saclay 2018-2020 [Lauréat IncubAlliance 2019 du meilleur partenariat avec un organisme public].
- Participation au projet OMTE Lahdak-Tools de maturation technologique soutenu par la SATT Paris Saclay 2013-2014.
- ADT OD-Map, Action de Développement Technologique (2011– 2012) de l'Inria sur l'optimisation des outils d'alignement d'ontologies (TAXOMAP) et de liage de données LN2R et la construction d'une chaîne de traitements plus intégrée.
- Dépôt logiciel à l'APP de l'outil LN2R linking tool 1.0 en 2012 (IDDN .FR.001.52006.00.S.P.2012.000.10000).
- Participation aux Prix de la valorisation 2012 et 2014, de l'Université Paris Sud pour l'outil LN2R linking tool.

Responsabilités liées à la recherche :

- Responsable adjointe de l'équipe LaHDAK du LRI, depuis décembre 2019.
- Responsable de l'axe "Intégration de données et de connaissances" de l'équipe LaHDAK, depuis septembre 2016.
- Responsable de l'organisation des séminaires de l'équipe LaHDAK depuis octobre 2015.
- Membre de la commission locaux du laboratoire 2009–2015.

1 Description de l'activité d'enseignement

1.1 Synthèse des enseignements

Durant ma carrière de Maître de conférences à l'IUT d'Orsay, j'ai enseigné dans différentes filières, en formation initiale, en apprentissage et en formation continue. J'ai enseigné dans des formations de différents niveaux d'études (du L1 au M2), en DUT et en Licence professionnelle (IUT d'Orsay), en L2 (UFR des sciences de l'Université Paris Sud), en 5ème année d'ingénieurs à PolyTech Paris Sud, en Master 2 Recherche à l'Université Paris Sud et à l'Université Paris Dauphine. Au fil des années, les thèmes de mes enseignements ont bien sûr varié, ainsi que la nature des interventions (TP, TD, Cours-TD et cours, encadrement de stages ou de projets tutorés).

En tant que Maître de Conférences à l'IUT d'Orsay, j'enseigne diverses matières en Informatique allant de la conception orientée objet, à la programmation Web en passant par la sécurité des bases de données et les bases de données réparties. De 2009 à 2019 je suis intervenue dans les modules Web Sémantique et intégration d'informations dans deux masters recherche : le master ISI de l'Université Paris Dauphine et le master IAC de l'Université Paris Sud (jusqu'en 2013). Depuis sept. 2015 j'interviens et je coordonne le cours "*Information Integration*" du master anglophone D&K de l'Université Paris Saclay. Je suis aussi intervenue à l'UFR des Sciences de l'Université Paris Sud pour des besoins ponctuels. Enfin, dans le cadre d'une formation proposée par le CNRS sur le "Panorama IA" je suis intervenue pour un cours sur le Web Sémantique et ses applications. Ma charge d'enseignement est d'en moyenne 250h eq. TD par an. Les volumes indiqués dans le tableau ci-dessous sont en heures équivalent TD par an par matière.

En italique sont indiquées les enseignements dont j'ai eu la responsabilité. Les mentions INI, FC et APP réfèrent à des formations initiales, en formation continue et en apprentissage, respectivement.

Matière	Nature	Niveau	Formation	Volume	Effectif
<i>Sécurité des bases de données</i>	C/TD/TP	IUT(L3)	APP	43	~29
<i>XML et ses applications</i>	C/TD/TP	IUT(L2)	FC	20	~ 15
<i>Bases de la conception orientée objet</i>	C/TD/TP	IUT(L1)	INI, FC, APP	42~ 48	~ 25
<i>Bases de données réparties</i>	C/TD/TP	IUT(L3)	APP, FC	18~36	~29
<i>Analyse et conception des SIs</i>	C/TD/TP	IUT(L2)	FC	32	17
<i>Projet Java / UML</i>	TP	IUT(L2)	APP	19	~ 25
Bases de la conception orientée objet	TD/TP	IUT(L1)	INI	70	~17
Introduction aux bases de données	TD/TP	IUT(L1)	INI	88	~17
Bases de données avancées	TD/TP	IUT(L2)	INI	42	~27
Systèmes et Réseaux	TD/TP	IUT(L2)	INI	28	~ 27
Programmation Web	C/TD/TP	IUT(L2)	INI	45	~ 27
Conception et programmation objet avancée	TD/TP	IUT(L2)	INI	22.5	~27
Introduction à UML	C/TD/TP	UFR(L2)	INI	10	~40
Web Sémantique	C/TD	M2(ISI)	INI	9	~15
Intégration d'informations hétérogènes	C/TD	M2(IAC)	INI	9	~12
<i>Information Integration</i>	C/TD	M2 (D&K)	INI	15	~15
<i>Contraintes et Résolution en IA</i>	C/TD/TP	M2(Polytech)	APP	32	28
<i>Panorama IA - Web sémantique</i>	C	au CNRS	FC	4.5	12

1.2 Tâches et responsabilités liées à la pédagogie

En plus des tâches de préparation des supports de cours, de TD et de TP pour les cours dont j'ai été responsable, j'ai assuré diverses tâches, telles que l'encadrement de projets tutorés en première année de DUT, le suivi des stages de 2ème année de DUT, en année spéciale et dans deux licences professionnelles de l'IUT d'Orsay (PRISM et SRSI), l'encadrement de projets personnels et professionnel (PPP) en première année de DUT ainsi que la participation aux jurys de fin de semestre et de fin d'année.

Par ailleurs, j'ai pris en charge plusieurs responsabilités, présentées ci-dessous, essentielles pour la formation des étudiants et la vie de l'établissement.

Responsabilité de formations

- Co-responsable pédagogique de la Licence Professionnelle SRSI – Sécurité des Systèmes et des Réseaux, « Formation par alternance et labélisée par l'ANSSI », depuis septembre 2018, à l'IUT d'Orsay.

Tâches : recrutement des étudiants, aide à la recherche de missions d'alternance, gestion de la relation avec les entreprises et des missions en entreprise, relation avec le CFA, jurys de validation et d'attribution des diplômes, gestion du renouvellement de l'habilitation et de l'évaluation (HCERES) du diplôme, relation avec l'agence ANSSI (Agence nationale de la sécurité des systèmes d'information), participation aux jurys de VAE et communication autour du diplôme (forums des licences professionnelles et de poursuite d'études, par exemple)

- Co-responsable pédagogique de la formation DUT Air-France en « Formation Continue » (pour la reconversion professionnelle des employés chez Air France) en 2014-2015, à l'IUT d'Orsay.

Tâches : négociation et montage de la formation avec l'entreprise Air France, gestion du recrutement des étudiants, gestion de l'équipe enseignante, suivi pédagogique de la formation au travers de bilans intermédiaires, la gestion des stages chez Air-France, traçabilité des informations pour les audits de formation continue et jurys de validation et d'attribution des diplômes de DUT.

- Co-responsable pédagogique de la licence PRISM – Programmation Internet et Systèmes Mobiles en « Formation Continue » (pour les employés chez Air France) en 2015-2016, à l'IUT d'Orsay.

Tâches : gestion de l'équipe enseignante, suivi pédagogique de la formation au travers de bilans intermédiaires, la gestion des stages chez Air-France, traçabilité des informations pour les audits de formation continue et jurys de validation et d'attribution des diplômes de licence PRISM.

Accompagnement pour la mobilité internationale et l'insertion professionnelle

- Correspondante des relations internationales pour le département informatique de l'IUT Orsay, 2011– 2013.

Tâches : gestion et étude des demandes des étudiants pour des départs pour des semestres à l'étranger, gestion et aide au montage de dossiers pour l'AMIE (Aide à la Mobilité Internationale des Etudiants), aide et conseil pour les demandes de visa et de permis de travail demandés par certains pays tels que le Canada et le Cambodge, relations avec les universités étrangères sous convention telles que l'Université de Chicoutimi (Canada), l'Université Trois Rivières au Québec (Canada) et l'Université de Cordoue en Espagne ou sans convention telles que Glyndwr University au Royaume Uni, aide et mise en contact pour la recherche de stages à l'étranger, organisation des soutenances de stages et bilans pédagogiques en relation avec les universités d'accueil.

- Responsable de l'organisation du forum « Etudiants–Entreprises » pour le département Informatique de l'IUT d'Orsay, 2009–2015.

Tâches : communication autour de l'événement auprès des entreprises, communication auprès des étudiants de 2ème année de DUT informatique, de DUT en un an et des étudiants en Licences professionnelles, planification des présentations des entreprises, organisation des stands pour les rencontres étudiants-entreprises et animation du forum.

- Tutorat d'activité en entreprises pour des étudiants en licences professionnelles et suivi de stages de 2ème années de DUT.

Tâches : accompagnement des étudiants durant leur stage en entreprise, contact avec les encadrants professionnels, une ou deux visites en entreprise (selon les exigences de la formation), suivi de la mission, aide à la rédaction du rapport et à la préparation de la soutenance, participation au jury de soutenances des stages

Présence dans les instances

- Membre élue au « Conseil d'IUT », 2014–2018, à l'IUT d'Orsay.
- Membre élue au « Conseil du Département Informatique » de l'Université Paris Sud, 2017–2019.
- Membre élue au « Comité de Pôle », de l'école doctorale STIC de l'Université Paris Saclay, depuis décembre 2019.

2 Activités de Recherche

2.1 Contexte scientifique

Avec l'initiative du Web de données (Linked Open Data -LOD, en anglais), défini par Tim-Berners Lee en 2006, comme *“Le Web sémantique ne consiste pas seulement à mettre des données sur le Web. Il s'agit de créer des liens pour qu'une personne ou une machine puisse explorer le réseau de données. Avec les données liées, quand vous en avez une partie, vous pouvez trouver d'autres données connexes”*, nous assistons aujourd'hui à une production sans précédent de ressources publiées sous la forme de données liées. Cela a permis la création de graphes de connaissances (GC) contenant des milliards de triplets RDF (Resource Description Framework), tels que DBpedia, YAGO et Wikidata du côté académique, et Google Knowledge Graph ou eBay Knowledge Graph du côté commercial. Ces graphes de connaissances sont construits en appliquant différentes approches, telles que l'extraction automatique d'informations à partir de textes, en réutilisant des sources existantes ou de façon collaborative en s'appuyant sur les méthodes d'extraction d'informations à partir de textes. Cependant, créer des graphes de connaissances tout en garantissant leur complétude et leur correction est un défi scientifique majeur en raison de l'imperfection des méthodes utilisées pour leur construction et de l'évolution des entités qui y sont décrites. Nous référons au problème d'enrichissement et de correction des graphes de connaissances par le terme *“raffinement de graphes de connaissances”*.

Pour ce problème complexe, j'ai travaillé sur plusieurs problématiques relevant soit de l'enrichissement de graphes de connaissances ou de la validation de leur contenu. J'ai notamment étudié les problèmes de liage de données, de l'invalidation de liens, de la découverte de clés, de la fusion de données et de la génération de méta-faits temporels dans les graphes de connaissances. J'ai par ailleurs travaillé sur la modélisation et l'exploitation contextuelle d'informations personnelles, qui s'inscrit dans un tout autre cadre, celui des systèmes de gestion d'informations personnelles (PIMS).

Ces travaux ont été réalisés en collaboration avec plusieurs collègues chercheurs et enseignants chercheurs, dans le cadre de plusieurs thèses, post-docs et stages de masters ; et pour certains dans le cadre de projets de recherche ANR, CNRS et industriels, impliquant différents organismes et entreprises, tels que, l'INRIA, le CNRS, l'INRA, l'INA, l'ABES, l'IGN et Thalès.

2.2 Description des travaux antérieurs

2.2.1 Liage de données

Le problème de liage de données concerne la détection de liens d'identité exprimant que deux descriptions différentes réfèrent au même objet du monde réel (e.g. même personne, même gène, même livre). Ce problème a été étudié depuis les années 60 dans divers domaines tels que les bases de données et les statistiques. Dans le cadre du LOD, le liage de données est devenu un thème de recherche central dans la communauté du Web sémantique. Le liage de données est, en effet, une tâche essentielle pour la création d'un espace de données où celles-ci sont liées par des liens sémantiques permettant aux robots logiciels d'explorer le Web de données en navigant via les liens. Ainsi, une réelle exploitation et une capitalisation des données et des connaissances peuvent être réalisées. Cependant, dans un tel cadre, la tâche de liage de données est rendue plus complexe du fait de la nature hétérogène, distribuée et de la qualité variable des données (e.g. fraîcheur, réputation, précision).

Dans ce contexte, ma recherche m'a tout d'abord menée vers l'étude et le développement d'approches automatiques pour le liage de données exploitant des ontologies. Les ontologies offrent un modèle commun de représentation des connaissances et des mécanismes de raisonnement logiques associés. Les outils développés sont fondés sur les langages standards du W3C pour le Web sémantique tel que OWL pour la représentation des ontologies et des liens et RDF pour la description des données. Tout d'abord dans le cadre de ma thèse soutenue en 2007, j'ai développé LN2R [6,30], une approche de liage de données combinant deux méthodes : (i) une méthode logique (L2R) qui traduit les axiomes déclarés dans l'ontologie en un ensemble de règles d'inférence en logique du premier ordre, et qui applique

un raisonnement utilisant la règle de résolution unitaire pour inférer des liens “*owl:sameAs*” (resp. “*owl:differentFrom*”) pour des paires de données référant au même objet (resp. à des objets différents) du monde réel; et (ii) une méthode numérique (N2R) qui exploite les axiomes de l’ontologie pour construire des fonctions de calcul de similarité entre données, qui sont ensuite représentées dans un système d’équations que l’on résout en utilisant une méthode itérative inspirée de la méthode de Jacobi. J’ai défini et développé cette approche dans sa première version pendant ma thèse. Ensuite, en collaboration avec Nathalie Pernelle (Maître de conférences à l’Université Paris Sud), j’ai contribué à l’amélioration et au développement de plusieurs extensions [11,35] de LN2R, en particulier dans le cadre du projet ANR GeOnto, de l’ADT INRIA OD-MAP et de l’OMTE LaHDAK-Tools. Nous avons par exemple étendu LN2R de façon à pouvoir considérer les propriétés issues de différentes ontologies non alignées pour le calcul de similarité entre deux instances [35]. Nous avons également travaillé sur l’explication des résultats de la méthode numérique N2R en utilisant les réseaux de Pétri [11,56]. En 2010 en collaboration avec Nobal Niraula, j’ai participé à la compétition internationale OAEI dans laquelle LN2R [40] a été classé 2ème sur 7 outils dans le track Person-Restaurant. Enfin, en collaboration avec Patrice Buche (Ingénieur de Recherche HDR à l’INRA Montpellier) et Juliette Dibie (Professeur à AgroParisTech), j’ai défini une méthode de détection de redondances dans un entrepôt thématique de données [27,55]. Cet entrepôt est construit à l’issue d’une chaîne de traitements appliquée sur des tableaux représentant des résultats d’expérimentations dans le domaine du risque alimentaire. Ces tableaux, une fois extraits à partir de documents PDF, sont annotés sémantiquement par des concepts, des termes et des relations sémantiques d’une ontologie de domaine. Ces annotations sont représentées par des ensembles flous afin de capturer l’imprécision des données. La méthode de détection de redondances que nous avons développée permet de détecter les instances de relations sémantiques redondantes en utilisant des mesures de similarité adaptées à la comparaison d’ensembles flous.

2.2.2 Invalidation de liens d’identité

Les liens d’identité représentés par le prédicat *owl:sameAs* représentent actuellement plus de 558 millions de liens sur le LOD. Ils sont pour la plupart produits par des outils automatiques de liage de données qui ne garantissent pas une précision de 100%. Cela a conduit à l’introduction d’un nombre important de liens erronés sur le LOD. Des études ont montré que lorsque la sémantique du prédicat *owl:sameAs* (i.e. réflexivité, symétrie, transitivité et partage de propriétés) est appliquée, de nombreuses erreurs sont générées et propagées. C’est la raison pour laquelle je me suis intéressée au problème d’invalidation de liens *owl:sameAs*. Dans le cadre du projet de post-doc de Laura Papaleo co-encadré avec Nathalie Pernelle et financé par le projet ANR Qualinca, j’ai développé une méthode [22] qui exploite la sémantique logique des axiomes (e.g. fonctionnalité des propriétés) déclarés dans l’ontologie, pour détecter des incohérences dans les données qui peuvent conduire à l’invalidation de liens. Par des expérimentations sur des données de OAEI 2010, nous avons montré que cette approche permettait de détecter des liens d’identité erronés et d’améliorer d’environ 25% les résultats des méthodes de liage de données, en terme de précision. Une des limites de cette méthode est le passage à l’échelle et l’hypothèse trop forte sur l’homogénéité des ontologies décrivant les données. Plus récemment, dans le cadre de la thèse de Joe Raad, nous avons développé en collaboration avec Frank van Harmelen et de Wouter Beek (Vrije Universiteit Amsterdam), une méthode d’invalidation de liens [3,14] qui passe à l’échelle du LOD, c’est-à-dire, capable de traiter les 28 milliards de triplets RDF et les 558 millions de liens *owl:sameAs* sans hypothèse sur les données et les ontologies. L’idée consiste à exploiter uniquement la topologie du graphe formé à partir des liens d’identité, pour associer un degré d’erreur à chaque lien du graphe. La méthode Louvain connue pour la détection de communautés est utilisée pour structurer le graphe en plusieurs communautés dont la densité est exploitée pour le calcul du degré d’erreur. Les premières expérimentations ont donné des résultats prometteurs, l’invalidation de plus de 1 million de liens et un rappel de 93% sur un échantillon sélectionné aléatoirement.

2.2.3 Détection de liens d’identité contextuels

Plusieurs études ont montré que la relation d’identité peut être subjective, dépendante du contexte et peut varier dans le temps. Cette caractéristique ne peut pas être prise en compte avec l’utilisation du prédicat *owl:sameAs* qui assure la règle de partage de propriétés ($owl:sameAs(X, Y) \wedge p(X, Z) \Rightarrow p(Y, Z)$). Aussi, des liens alternatifs au *owl:sameAs* sont nécessaires pour capturer le contexte dans lequel la relation d’identité est valide. Par exemple, l’identité entre deux médicaments m_1, m_2 décrits par (*nom, prix, fabricant, composition-chimique, maladies*) dépend du contexte dans lequel on se situe. Dans un contexte commercial, m_1 et m_2 seront identiques s’ils ont le même nom, même prix et même fabricant et dans un contexte scientifique m_1, m_2 seraient considérés comme identiques s’ils ont

la même composition chimique et concernent les mêmes maladies. Pour répondre à ce problème, j’ai développé dans le cadre de la thèse de Joe Raad une approche de détection de liens d’identité contextuels [8,16,45] dans les graphes de connaissances. Le contexte a été défini comme une sous-partie de l’ontologie décrivant les données et organisées dans un treillis. Dans le cadre du projet LIONES du CDS (Center of Data Science) Paris Saclay, des expériences sur des graphes de connaissances représentant des processus de transformation scientifiques ont pu montrer l’intérêt des contextes pour la tâche de prédiction de valeurs manquantes d’observation. Ces travaux sont poursuivis dans la thèse de Lucas Simonne qui s’inscrit dans le cadre du projet WarmRules (2019-2022) financé par l’IRS DATA-IA. Il s’agit d’exploiter les contextes d’identité lors de la découverte de règles graduelles causales dans des données phénotypiques sur les plantes et des données environnementales.

2.2.4 Découverte automatique de clés

Certaines approches de liage de données s’appuyant sur les connaissances telles que LN2R, exploitent des clés (ensemble de propriétés permettant d’identifier de façon unique un objet) pour inférer des liens d’identité entre instances. Dans ces approches, les clés sont supposées être fournies par un expert du domaine. Néanmoins, dans un contexte tel que le LOD contenant actuellement des centaines de milliers de sources de données et près de 60 milliards de triplets RDF, il n’est pas envisageable de considérer une telle hypothèse. Aussi, dans le cadre de la thèse de Danai Symeonidou (soutenue en 2014) en co-encadrement avec Nathalie Pernelle, nous avons développé deux méthodes efficaces (KD2R et SAKey) de découverte automatique de clés à partir de données RDF. L’outil KD2R [5] permet de découvrir des clés sûres (valides pour 100% des données) à partir de plusieurs sources de données. SAKey [23] permet de découvrir des clés avec des exceptions (clés qui ne sont pas valides pour 100% des données). Ainsi, il est possible de tenir compte de la qualité variable des données et de découvrir le plus de clés possible. Par des expérimentations, sur près d’une dizaine de jeux de données dont certains proviennent du LOD (DBPedia) et d’autres de compétitions internationales (OAEI), nous avons validé l’efficacité et la pertinence de ces approches. Cependant, dans certains jeux de données particuliers et contenant beaucoup de redondances, parfois aucune clé ne peut être découverte. C’est la raison pour laquelle j’ai développé une approche de découverte de clés conditionnelles [18] qui ne sont valides que dans une partie des données, caractérisées par des conditions. Enfin, en collaboration avec des collègues du LIG et du LIRMM dans le cadre du projet ANR Qualinca, j’ai travaillé sur la comparaison théorique [24,52] et expérimentale [11] des différentes sémantiques des clés, i.e., celle définie dans OWL2 et celle considérant l’égalité des ensembles de valeurs au lieu de l’existence d’une valeur commune dans la sémantique OWL2. Dans le continuum de la découverte de clés, je travaille en ce moment sur le développement d’approches de détection d’expressions référentielles permettant d’identifier de façon unique une instance (premiers résultats publiés dans [34,43]).

2.2.5 Enrichissement de données

Un moyen d’enrichir les graphes de connaissances est de fusionner les descriptions des instances liées par des liens d’identité. La tâche de *fusion de données* permet d’agréger les données liées et de produire une représentation unique et homogène qui soit la plus complète, la plus concise et la plus correcte possible des données liées. Dans cette direction, en collaboration avec Rallou Thomopoulos (Chargé de Recherche INRA Montpellier) j’ai développé une approche originale de fusion de données multi-critères qui s’appuie sur des indicateurs de qualité (la fréquence et l’homogénéité des valeurs, leur niveau de précision, la fraîcheur et la fiabilité des sources). La méthode permet de calculer pour chaque valeur de propriété, un degré de confiance puis de proposer un classement pour les valeurs des propriétés. Par ailleurs, en collaboration avec Sébastien Destercke (Chargé de Recherche CNRS au HEUDIASYC à Compiègne) et de Rallou Thomopoulos, nous avons travaillé sur le problème de modélisation de l’incertitude dans les résultats de la fusion. Nous avons proposé trois modélisations : la première utilise une représentation par des ensembles flous [29], la deuxième est fondée sur la théorie des possibilités [28] et la dernière s’appuie sur la théorie des croyances [57]. Plus récemment, nous avons proposé une approche pouvant expliquer le résultat de la fusion de données en utilisant la théorie de l’argumentation. L’approche proposée [15,48], génère des arguments en faveur ou en défaveur d’une valeur en exploitant les informations sur les critères, les règles logiques déclenchées et les relations de spécificité entre valeurs, détectées par la méthode de fusion de données et conservées grâce à l’ontologie de fusion qui a été définie dans les précédents travaux.

J’ai aussi travaillé avec Rallou Thomopoulos (CR INRA Montpellier) sur le problème de la prédiction de valeurs manquantes dans une base de règles construite à partir de résultats scientifiques tirés de la littérature mais aussi à partir de connaissances expertes. Ces règles expriment la notion de causalité entre un ensemble de conditions expérimentales

et les résultats observables dans les expériences scientifiques sur des produits céréaliers. Nous avons développé une approche [4,10] qui s'appuie sur des méthodes de liage de données pour créer une partition des règles. A partir de cette partition, un algorithme de prédiction que nous avons développé peut prédire les informations manquantes. Une comparaison expérimentale a été effectuée avec une approche fondée sur les arbres de décision et a montré l'efficacité et la pertinence de l'approche que nous avons développée.

2.2.6 Génération de méta-faits temporels dans les graphes de connaissances

Les entités (e.g. les personnes, les lieux, les organisations, les lois) décrites dans les graphes de connaissances sont en évolution permanente, et les informations les concernant évoluent également dans le temps (e.g., l'adresse ou la fonction d'une personne). Malgré l'importance de l'information temporelle, les graphes de connaissances actuels ne fournissent que peu d'informations sur la validité temporelle des faits. Quelques initiatives sont proposées dans des graphes de connaissances comme Yago2 et Wikidata, mais cela ne concerne que très peu de prédicats (seulement 15 prédicats dans Yago2).

En collaboration avec Gianluca Quercini (Maître de conférence à Centrale-Supelec), dans le cadre du post-doc de Joana Malaverri financé par le projet VASTE (Labex Digicosme), j'ai développé une approche [13,63], qui en deux étapes génère deux types de méta-faits temporels, de type timestamp ou intervalle, pour un ensemble de prédicats défini en entrée. Par exemple, étant donné le fait $f : \langle \text{Barack_Obama}, \text{presidentOf}, \text{USA} \rangle$, deux métafaits temporels sont générés, à savoir $\text{startOnDate}(f, '2009-01-20')$ et $\text{endOnDate}(f, '2017-01-20')$. L'algorithme applique tout d'abord une première étape qui, à partir d'un ensemble de motifs de graphes de connaissances génère un ensemble initial de méta-faits temporels. Ensuite, dans une deuxième étape, cet ensemble initial est propagé et amplifié par l'utilisation de règles de Horn exprimant des dépendances sémantiques entre les prédicats. Ces règles sont découvertes par un outil de fouille de règles tel que AMIE+. Une première évaluation expérimentale a été réalisée sur le graphe de connaissances DBpedia en comparaison avec Yago. Bien que perfectibles pour notamment les méta-faits intervalles, les résultats obtenus sont prometteurs et ont montré la pertinence d'une telle approche pour la génération de méta-faits temporels dans les graphes de connaissances.

2.2.7 Modélisation et utilisation des informations personnelles

Dans un tout autre contexte dans le cadre du projet ANR PIMI (2010-2013) et de la thèse de Rania Khefifi, co-encadrée avec Pascal Poizat (Professeur à l'Université Paris Ouest Nanterre la Défense et membre du LIP6), j'ai travaillé sur le problème de la gestion des informations personnelles. Nous avons tout d'abord proposé une modélisation multi-points de vue des informations personnelles en utilisant des ontologies existantes ou créées par l'utilisateur. En effet, lorsque l'on utilise des informations personnelles, il est très important de pouvoir prendre en compte le contexte (social, géographique et temporel) de leur utilisabilité. C'est ainsi que nous avons proposé une représentation contextuelle des informations personnelles [9,25,54]. Ensuite, pour assister l'utilisateur dans la réalisation de procédures administratives en ligne (e.g. inscription universitaire, demande de visa), nous avons développé un système d'aide au remplissage automatique de formulaires [9] en générant automatiquement des requêtes contextualisées sur l'espace d'informations personnelles. Pour l'évaluation de ces requêtes, nous avons développé deux algorithmes (un strict et un flexible). Enfin, pour permettre la réalisation automatique de procédures abstraites représentées par des workflows, nous avons développé une approche par composition automatique de services [25]. Pour prendre en compte les contextes d'exécution des procédures (e.g. professionnel, personnel, en france, en espagne) mais aussi les politiques de sécurité éventuelles spécifiées par l'utilisateur, nous avons étendu l'algorithme Graphplan pour calculer les compositions de services possibles.

2.3 Projet de Recherche

Dans la continuité de mes travaux sur le raffinement de graphes de connaissances où l'on s'intéresse à l'enrichissement et à la validité des données et connaissances, je souhaite suivre deux directions de recherche complémentaires. La première direction vise à augmenter le niveau d'expressivité des graphes de connaissances de façon à pouvoir appliquer des techniques de raisonnement symbolique dans différentes tâches telles que le liage de données, l'enrichissement par l'inférence de nouvelles connaissances, l'évaluation de requêtes sur un ou éventuellement plusieurs graphes de connaissances et la détection d'inconsistances. Pour ce dernier point, je souhaite notamment étudier et développer des approches qui permettent d'enrichir les graphes de connaissances par des informations négatives telles que, $\neg$

<Obama, bornIn, Kenya> ou encore <Musée-Louvre, differentFrom, Musée-Branly> de façon à rendre possible la détection de déclarations contradictoires pouvant être ensuite soumises à une phase de réparation pour ainsi améliorer la validité du contenu des graphes de connaissances. Une des pistes que j'envisage d'explorer est la découverte d'axiomes de disjonction entre classes (e.g. la classe Personne est disjointe de la classe Ville) à partir des données. Ces classes de disjonctions peuvent ensuite être exploitées pour générer des liens de différence entre instances et ensuite générer des informations négatives tout en considérant l'hypothèse du monde ouvert dans laquelle se positionnent les graphes de connaissances RDF.

Par ailleurs, je souhaite davantage tenir compte de l'aspect évolutif des graphes de connaissances qui est une de leur caractéristiques intrinsèques. En effet, ils sont fréquemment mis à jour, car les entités du monde réel évoluent et de nouvelles connaissances et données à leur sujet sont générées en permanence : "chaque jour, nous créons 2,5 quintillions d'octets de données". L'évolution des graphes de connaissances peut être reflétée par des changements au niveau conceptuel, c'est-à-dire, la définition de classes, de propriétés et d'axiomes déclarés dans l'ontologie. Les changements peuvent également concerner le niveau instances (données) où les changements peuvent affecter les types des instances, les propriétés les décrivant ou encore les liens d'identité entre instances. Dans un tel cadre, je souhaiterais me focaliser dans un premier temps sur l'impact de l'évolution au niveau des instances sur l'ensemble des tâches impliquées dans le processus de raffinement des graphes de connaissances, notamment, le liage de données, la découverte de connaissances et la fusion de données. En effet, lorsque les données changent, les axiomes (par exemple, les clés, les propriétés fonctionnelles) découverts peuvent devenir invalides et de nouveaux axiomes peuvent émerger (un travail préliminaire que j'ai réalisé sur cette question a été publié dans [17]); les liens d'identité peuvent devenir invalides et de nouveaux peuvent être détectés; et enfin des valeurs de propriétés peuvent être ajoutées/supprimées ou modifiées, par conséquent le résultat de la fusion de données doit être aussi mis à jour. L'une des questions majeures à étudier est de comment adapter les algorithmes de raffinement de graphe de connaissances déjà développés pour gérer cette évolution de données et de connaissances de façon incrémentale. Un autre aspect découlant de l'évolution des graphes de connaissances concerne la validité temporelle des faits, problème que j'ai commencé à aborder dans [13,63] en proposant une approche d'enrichissement de graphes de connaissances par des méta-faits temporels. Ce travail préliminaire devra être complété par une approche d'évaluation de la véracité des faits en combinant les informations sur la validité temporelle des faits avec d'autres critères liés à la véracité tels que la fiabilité des sources, la fréquence des valeurs dans les sources, etc.

Dans le cadre des graphes de connaissances scientifiques caractérisés par leur complexité et leur incomplétude due à la rareté des données d'expérience (coût de réalisation élevé), je souhaiterais étudier le problème de la découverte de règles causales pour l'explication de certains phénomènes biologiques et naturels, notamment, l'impact du changement climatique sur les phénotypes des plantes. Ceci est l'objectif du projet WarmRules (2019-2022) financé par DATA-IA dans lequel j'envisage de me focaliser sur la découverte de règles de causalité graduelles afin d'exprimer les effets de variations climatiques (augmentation ou diminution de certains facteurs) sur les phénotypes des plantes. Découvrir par exemple, l'effet de l'augmentation de la température au sol sur la date de floraison du maïs. Dans ce travail, des méthodes que j'ai développées dans de précédents travaux peuvent être exploitées ou adaptées telles que la méthode de détection de liens d'identité contextuels [8,16,45] qui devraient permettre d'isoler les facteurs invariants des facteurs variables pouvant contenir la cause et/ou expliquer un phénomène donné. Un premier travail que j'ai réalisé [44] sur la découverte de séquences graduelles partiellement ordonnées peut servir de base pour le calcul d'atomes graduels, pouvant être impliqués dans les règles causales graduelles.

Plus généralement et à plus long terme je souhaiterais m'intéresser plus à l'utilisation de modèles d'incertitude et de méthodes d'apprentissage numériques pour capturer davantage l'incertitude et l'imprécision des données et des connaissances déclarées dans les graphes de connaissances évolutifs et incertains.

3 Encadrement doctoral et scientifique

3.1 Encadrement doctoral

[Total : 5]

1. **Lucas Simonne (Décembre 2019– Décembre 2022)**, direction et co-encadrement avec Nathalie Pernelle (Professeur à l'Université Paris 13 et membre du LIPN) et Juliette Dibie (Professeur à AgroParisTech).

Sujet de la thèse : *Découverte de règles causales dans les graphes de connaissances scientifiques.*

Financement : Projet WarmRules lauréat de l'appel à projets de l'institut convergence DATA-IA en 2019.

2. **Armita Khajeh Nassiri (Décembre 2019– Décembre 2022)**, direction et co-encadrement avec Nathalie Pernelle (Professeur à l'Université Paris 13 et membre du LIPN) et Gianluca Quercini (Maître de conférence à CentraleSupélec et membre du LRI).

Sujet de la thèse : *Découverte d'expressions référentielles dans les graphes de connaissances*.

Financement : Projet BPI-PSPC AIDA en collaboration avec IBM France, SOFTEAM, Stet, DecisionBrain, CEA, CentraleSupélec et l'ENS Saclay.

3. **Joe Raad (Octobre 2015– Octobre 2018)**, co-encadrement à 33% avec Juliette Dibie (Professeur à AgroParisTech), Liliana Ibanescu (Maître de Conférences à AgroParisTech) et Nathalie Pernelle (Maître de Conférences à l'Université Paris Sud et membre du LRI).

Titre de la thèse : *Gestion d'identité dans des graphes de connaissances*.

Financement : Projet LIONES lauréat de l'appel à projets 2015 du Center of Data Science (CDS) de Paris Saclay.

Prix : **Premier Accessit au prix** de la meilleure thèse en Intelligence Artificielle de l'AFIA en 2019.

Position actuelle : Post-Doc à Vrije Universiteit, Amsterdam, Netherlands avec Prof. Frank V. Harmelen.

4. **Danai Symeonidou (Octobre 2011– Octobre 2014)**, co-encadrement à 50% avec Nathalie Pernelle (Maître de Conférences à l'Université Paris Sud et membre du LRI).

Titre de la thèse : *Découverte automatique de clés pour le liage de données*.

Financement : une bourse doctorale du ministère.

Situation actuelle : Chargée de Recherche à l'INRA.

5. **Rania Khelifi (Décembre 2010– Septembre 2014)**, co-encadrement à 50% avec Pascal Poizat (Professeur à l'Université Paris Ouest Nanterre la Défense et membre du LIP6).

Titre de la thèse : *Informations personnelles sensibles aux contextes : modélisation, interrogation et composition*.

Financement : Projet ANR PIMI (ANR VERSO 2010–2013) puis ATER en 2013-2014.

Situation actuelle : Cadre dans une entreprise en Recherche et Développement.

3.2 Encadrement de post-doctorants et d'ingénieurs

[Total : 6]

1. **Joana Gonzales Malaverri (mai 2018- mai 2019)** co-encadrement 60% avec Gianluca Quercini (Maître de conférences à Centrale-Supélec), du projet post-doctoral dans le cadre du projet VASTE financé par le labex Digicosme Paris Saclay.
Titre du projet post-doctoral : Time-aware veracity assessment in knowledge graphs.
2. **Vincent Henry (décembre 2016-décembre 2017)**, co-encadrement à 33% avec Anne Goelzer (Ingénieur de Recherche INRA), Vincent Fromion (Directeur de Recherche INRA) et Juliette Dibie (Professeur AgroParisTech) d'un projet post-doctoral dans le cadre du projet IMSV Paris Saclay.
Projet intitulé : Conception d'une ontologie systémique multi-échelles de la plante.
3. **Laura Papaleo (Janvier 2013 à Juin 2015)**, co-Encadrement à 50% avec Nathalie Pernelle d'un projet post-doctoral dans le cadre du projet ANR Qualinca.
Projet intitulé : SameAs link Invalidation and Requalification in the Linked Open Data.
4. **Cyril Dumont (2014)**, participation à l'encadrement du projet d'ingénieur dans le cadre du projet Digiteo OMTE (Opération de Maturation Technico-Economique, LaHDAK-Tools (2014) en collaboration avec le pôle développement du LRI.
Projet intitulé : Intégration et Maturation de l'outil LN2R-LT dans la plateforme LaHDAK-Tools.
5. **Tushar Ghosh (2011-2012)**, co-encadrement à 25% avec Nathalie Pernelle (Maître de conférences à l'Université Paris Sud), Chantal Reynaud (Professeur à l'Université Paris Sud) et Brigitte Safar (Maître de conférences à l'Université Paris Sud) du projet d'ingénieur dans le cadre de l'ADT (Aide au Développement Technologique) OD-Map (2011-2012) financée par l'INRIA Saclay.
Intitulé du projet : Linking of RDF data conforming to different ontologies.
6. **Nobal Niraula (2010)** encadrement à 100% du projet d'ingénieur de Nobal Niraula dans le cadre du projet ANR GeOnto project.
Intitulé du projet : Software development of LN2R – a logical and a numerical approach for reference reconciliation.

3.3 Encadrement d'étudiants en Master recherche

[Total : 12]

1. **Yuru He (2020)** : encadrement à 60% avec Marcos Da Silveira (Chercheur au LIST, Luxembourg) et Cédric Pruski (Chercheur au LIST, Luxembourg). Stage dans le cadre du Master 2 Informatique – Data and Knowledge – de l'Université Paris Saclay, promotion 2020.
Sujet de stage : Découverte d'axiomes de disjonctions dans les graphes de connaissances.
2. **Sergey Kononov (2019)**, encadrement à 100% du stage de Master 2 Recherche "data and knowledge" de l'Université Paris Saclay, promotion 2019.
Stage intitulé : Génération de méta-faits temporels dans les graphes de connaissance RDF en utilisant les réseaux probabilistes marcoviens.
3. **Sixiao Zhu (2019)**, co-encadrement à 33% avec Nathalie Pernelle (Maître de Conférence Université Paris Sud), Dominique Quadri (Maître de conférences, Université Paris Sud) et Joe Raad (post-doc à Vrije Universiteit, Amsterdam) du stage de Master 2 Recherche "data and knowledge" de l'Université Paris Saclay, promotion 2019.
Stage intitulé : Détection de liens d'identité invalides en utilisant la programmation mathématiques.
4. **Aziza Filali Rothi (2019)**, co-encadrement à 33% avec Juliette Dibie (Professeur AgroParisTech) et Nathalie Pernelle (Maître de Conférence Université Paris Sud) du stage de Master 2 Recherche Informatique de l'Université de Caen, promotion 2019.
Stage intitulé : Sélection de contextes d'identité pertinents pour la découverte de règles causales graduelles.
5. **Anderson Carlos Ferreira Da Silva (2018)** co-encadrement à 40% avec Frédéric Andrès (Professeur NII, Tokyo) et Emmanuel Waller (Maître de Conférence Université Paris Sud) du stage de Master 2 Recherche "data and knowledge" de l'Université Paris Saclay, promotion 2018
Stage intitulé : Approche exploitant le contenu des descriptions des ressources et de leur évolution pour la détection de liens d'identité invalides.
6. **Sara Oulkif (2017)**, co-encadrement à 50% avec Juliette Dibie (Professeur AgroParisTech) du stage de Master 2 Recherche Systèmes d'information et Business Intelligence CNAM et université Paris 1, promotion 2017.
Stage intitulé : Data Quality Assesment in RDF Scientific Knowledge Bases. Sujet financé par DigiCosme du Labex Paris Saclay.
7. **Ioanna Giannopoulos (2014)**, co-encadrement à 50% avec Rallou Thoumoupolos (Chargée de Recherche INRA) du stage de Master 2 Recherche Informatique de l'université Paris Dauphine, promotion 2014.
Stage intitulé : Fusion de données liées exploitant la sémantique de l'ontologie et fournissant une évaluation de la qualité des données.
8. **Hiba Hajri (2014)**, co-encadrement à 50% avec Lilian Ibanescu (Maître de Conférences, AgroParisTech) du stage Master 2 Recherche Systèmes d'information et Business Intelligence CNAM et Université Paris 1, promotion 2014.
Stage intitulé : Définition d'un langage pour représenter les évolutions d'une ontologie.
9. **Maria Koutraki (2012)** co-encadrement à 50% avec Nathalie Pernelle (Maître de Conférences à l'université Paris Sud) du stage de *Master 2 research in Computer Science of University of Crete* (promotion 2012).
Stage intitulé : Identity link discovery by considering not-aligned properties.
10. **Danai Symeonidou (2011)** co-encadrement à 50% avec Nathalie Pernelle (Maître de Conférences, Université Paris Sud) du stage du *master Franco-Hellenic, University of Crete et Université Paris Sud* (promotion 2011).
Stage intitulé : *Automatic discovery of key constraints for Data linking*.
11. **Rania Khefifi (2010)** co-encadrement à 33% avec Patrice Buche (Ingénieur de Recherche INRA) et Juliette Dibie-Barthélemy (Maître de Conférences HDR, AgroParisTech) du stage de *Master 2 Recherche Informatique de l'université Paris Dauphine* (promotion 2010).
Stage intitulé : *Détection de redondances dans les données tabulaires annotées par une ontologie de domaine*.
12. **Souhir Gahbiche** co-encadrement à 50% avec Nathalie Pernelle (Maître de Conférences, Université Paris Sud) du stage de *Master 2 Recherche Informatique de l'université Paris Sud* (promotion 2009).
Stage intitulé : *Explication de décision de réconciliation, décisions obtenues par une méthode numérique de liage de données*.

3.4 Encadrement d'étudiants en Magistère, Master 1 et d'ingénieurs

[(Total : 13)]

1. **Yuru He** (Mai – Août 2019) : encadrement à 100%. Stage dans le cadre du Master 1 Informatique de l'Université Paris Sud, promotion 2017.
Sujet de stage : Découverte d'axiomes de disjonctions dans les graphes de connaissances.
2. **Simon Ser** (Mai – Août 2017) : co-encadrement à 50% avec Maguelonne Teisseire (DR IRSTEA, Montpellier). Stage dans le cadre du Magistère – Licence 3 Informatique de l'Université Paris Sud, promotion 2017.
Sujet de stage : Découverte de séquences graduelles partiellement ordonnées dans les données scientifiques.
3. **Viktor Frelikh et Thomas Perraut** (Janvier – Mai 2017) : encadrement à 100%. Stage dans le cadre du TER –Travaux d'Etude et de Recherche – du Master 1 Informatique de l'Université Paris Sud, promotion 2017.
Sujet de stage : Explication des résultats de Fusion de données RDF.
4. **Anderson Carlos Ferreira Da Silva** (Janvier – Juillet 2017) : encadrement à 100%. Stage dans le cadre du TER –Travaux d'Etude et de Recherche – du Master 1 Informatique de l'Université Paris Sud, promotion 2017 et d'un stage d'été (Mai – Juillet 2017).
Sujet de stage : Fusion efficace de données RDF en exploitant la qualité des données et les connaissances du domaine.
5. **Dimitrios Christaras Papageorgiou** (Janvier – Mai 2017) : co-encadrement à 50% avec Nathalie Pernelle (MCF HDR Université Paris Sud). Stage dans le cadre du TER –Travaux d'Etude et de Recherche – du Magistère/Master 1 Informatique de l'Université Paris Sud, promotion 2017. Ce TER a été poursuivi par un stage de Magistère (Mai –Août 2017).
Sujet de stage : Invalidation numérique de liens d'identité dans les données RDF .
6. **Daniel Mercier** (Mai – Septembre 2016) : co-encadrement à 50% avec Nathalie Pernelle (MCF HDR Université Paris Sud). Stage dans le cadre du Magistère /Master 1 Informatique de l'Université Paris Sud, promotion 2016.
Sujet de stage : Invalidation numérique de liens d'identité dans les données RDF.
7. **Adrien Pavao** (Mai – Juillet 2016) : encadrement à 100 %. Stage dans le cadre du Magistère/Licence Informatique de l'Université Paris Sud, promotion 2016.
Sujet de stage : Optimisation et développement d'une interface graphique pour l'outil de fusion de données.
8. **Daniel Mercier et de Sujeeban Thuraisamy** (Janvier –Mai 2016) : co-encadrement à 50% avec Nathalie Pernelle (MCF Université Paris Sud, LRI). Stage dans le cadre du TER – Travaux d'Etude et de Recherche – du Magistère/Master 1 Informatique de l'Université Paris Sud, promotion 2016.
Sujet de stage : Détection et la représentation des changements dans les sources de données RDF.
9. **Anthony Pecriaux** (Juin – Septembre 2013) : Encadrement à 100%. Stage dans le cadre du DUT Informatique à l'IUT d'Orsay, Université Paris Sud, promotion 2013. Sujet de stage : « raffinement de mises en correspondance : Capitalisation d'expérience pour améliorer la réconciliation de références.
10. **Charles Jonard** (Mai –Août 2012) : co-encadrement à 50% avec Jean-Christophe Souplet (Responsable du pôle développement du LRI). Stage d'ingénieur 5ème année de Polytech Paris Sud, promotion 2012. Sujet de stage : Développement d'une interface graphique pour l'outil LN2R.
11. **Vincent EULOGE** (Janvier – Mai 2012). Encadrement à 100%. Stage dans le cadre du TER – Travaux d'Etude et de Recherche – du Master 1 Informatique de l'Université Paris Sud, promotion 2012. Sujet de stage : Développement d'un navigateur pour les Données Liées dans le LOD.
12. **Ren Jie et Pei Li** (Janvier –Mars 2010) : co-encadrement à 50% avec Nathalie Pernelle (MCF Université Paris Sud, LRI). Stage de 3 mois, Licence 3 Informatique dans le cadre du programme international HUST entre l'Université Paris Sud et le gouvernement Chinois, promotion 2010. Sujet de stage : Apprentissage de règles de classification pour le liage de données : application aux descriptions de produits électroniques.
13. **Demba Ndiaye** (Janvier – Mai 2010) : Encadrement à 100%. Stage dans le cadre du TER – Travaux d'Etude et de Recherche – du Master 1 Informatique de l'Université Paris Sud, promotion 2010. Sujet de stage : Développement d'un outil flexible de Fusion de Références.

4 Publications et production scientifique

Dans ce dossier, je présente mes publications par catégorie par ordre chronologique (du plus récent au plus ancien). Les différents classements internationaux que j'ai utilisés pour qualifier le rang de mes publications sont :

- Core : <http://portal.core.edu.au/conf-ranks/>
- Qualis : <http://www.conferenceranks.com/>
- LRI-AERES : <http://www.lix.polytechnique.fr/~lengrand/Events/2008-06-AERES/Conferences.htm>

4.1 Revues internationales avec comité de lecture

- [1] The sameAs Problem : A Survey on Identity Management in the Web of Data. *Joe Raad, Nathalie Pernelle, Fatiha Saïs, Wouter Beek, Frank van Harmelen*. International Semantic Web Journal, 2020 [under review <http://www.semantic-web-journal.net/system/files/swj2430.pdf>].
- [2] BiPOm : a rule-based ontology to represent and infer molecule knowledge from a biological process-centered viewpoint. *Vincent Henry, Fatiha Saïs, Elodie Marchadier, Juliette Dibie, Anne Goelzer and Vincent Fromion*. International BMC Bioinformatics Journal, 2020 [en révision].
- [3] Constructing and Cleaning Identity Graphs in the LOD Cloud. *Joe Raad, Wouter Beek, Frank van Harmelen, Nathalie Pernelle and Fatiha Saïs*. International Journal on Data Intelligence, 2019.
- [4] Ontology-aware prediction from rules : A reconciliation-based approach. *Fatiha Saïs, Rallou Thomopoulos*. In *Journal of Knowledge Based Systems* (2014), Elsevier. Volume 67, pages 117-130. April 2014. [Impact Factor : 4.62]
- [5] An Automatic Key Discovery Approach for Data Linking. *Nathalie Pernelle, Fatiha Saïs, Danai Symeonidou*. In *Journal of Web Semantics* (2013), Elsevier. Volume 23, pages 16-30. December 2013. [Impact Factor : 2.75]
- [6] Combining a Logical and a Numerical Method for Data Reconciliation. *Fatiha Saïs, Nathalie Pernelle, Marie-Christine Rousset*. In *Journal on Data Semantics XII* (2009), LNCS 5480, Springer. Volume 12, pages 16-30. [RG¹ Impact Factor : 2.44]

4.2 Revues nationales

- [7] Détection de liens d'identité erronés en utilisant la détection de communautés dans les graphes d'identité. *Joe Raad, Wouter Beek, Nathalie Pernelle, Fatiha Saïs, and Frank van Harmelen*. In *Revue des Sciences et Technologies de l'Information-Série ISI : Ingénierie des Systèmes d'Information*, 23(3-4) :95-118, 2018.
- [8] Comment représenter et découvrir des liens d'identités contextuels dans une base de connaissances : applications à des données expérimentales en science du vivant (2018). *Joe Raad, Nathalie Pernelle, Fatiha Saïs, Juliette Dibie-Barthélemy, Liliana Ibanescu and Stéphane Dervaux*. In *Revue d'Intelligence Artificielle - Numéro Spécial d'Ingénierie des Connaissances* 32(3) :345-372, 2018.
- [9] Vers une utilisation automatique des Informations Personnelles pour la réalisation de e-procédures (2014). *Rania Khefifi, Pascal Poizat and Fatiha Saïs*. In *Revue des Nouvelles Technologies de l'Information* vol. RNTI-W-2 pages : 75-89, 2014.

4.3 Chapitres de livre

- [10] A reconciliation-driven approach of case-based prediction : State of the art, method overview and application in food science. *Fatiha Saïs and Rallou Thomopoulos*. Chapitre du livre *Case-Based Reasoning : Strategies, Developments and Applications*, in Ed. Nova Science Publishers, Tao LIN (Ed.), 2015.
- [11] Explaining Reference Reconciliation Decisions : a Coloured Petri Nets based Approach. *Souhir Gahbiche, Nathalie Pernelle, Fatiha Saïs*. Chapitre du livre *Advances in Knowledge Discovery and Management, Studies in Computational Intelligence*, paru aux éditions Springer Verlag, H. Briand, F. Guillet, G. Ritschard and D. Zighed (Ed.) 2011.
- [12] Data Extraction, Transformation and Integration guided by an Ontology. *Chantal Reynaud, Nathalie Pernelle, Marie-Christine Rousset, Brigitte Safar, Fatiha Saïs*. Chapitre du livre *Data Warehousing Design and Advanced Engineering Applications : Methods for Complex Construction, Advances in Data Warehousing and Mining Book Series*, paru aux éditions IGI Publishing, Ladjel Bellatreche (Ed.), 2009.

4.4 Conférences internationales avec comité de lecture

- [13] MOMENT : Temporal Meta-Fact Generation and Propagation in Knowledge Graphs. *Fatiha Saïs, Joana E. Gonzales Malaverri and Gianluca Quercini*. In proceedings of the 35th ACM/SIGAPP Symposium On Applied Computing - Semantic Web and Applications track (SWA), Brno, Czech Republic March 30-April 3, 2020 (to appear) [Core B].

1. Research gate Impact Factor en 2014.

- [14] Detecting Erroneous Identity Links on the Web Using Network Metrics. *Joe Raad, Wouter Beek, Frank van Harmelen, Nathalie Pernelle and Fatiha Saïs*. In proceedings of the 16th International Semantic Web Conference (ISWC), pages : 391-407. Monterey, CA, USA, October, 8-12 2018. [Short-listed best paper] [Core A] .
- [15] Argumentation-based Explanation of Linked Data Fusion. *Fatiha Saïs, Rallou Thomopoulos and Andreson Carlos Ferreira*. In proceedings of the International Conference on Enabling Technologies : Infrastructure for Collaborative Entreprises (WETICE2018), pages : 275-280, Paris, France, June, 27-29 2018. [Core B]
- [16] Detection of Contextual Identity Links in a Knowledge Base. *Joe Raad, Nathalie Pernelle and Fatiha Saïs*. In proceedings of the Ninth International Conference on Knowledge Capture (K-Cap) 2017, pages : 8 :1-8 :8. December 4th-6th, 2017 Austin, Texas, United States [Core A].
- [17] Inferring the evolution of ontology axioms from linked data dynamics. *Fatiha Saïs, Cédric Pruski, Marcos Da Silveira*. In proceedings of the Ninth International Conference on Knowledge Capture (K- Cap) 2017, pages : 43 :1-43 :4. December 4th-6th, 2017 Austin, Texas, United States [Short paper] [Core A].
- [18] VICKEY : Mining Conditional Keys on RDF datasets. *Danai Symeonidou, Luis Galarraga, Nathalie Pernelle, Fatiha Saïs and Fabian Suchanek*. In proceedings of the 16th International Semantic Web Conference (ISWC) 2017, pages : 661-677. Vienna, Austria, October 21-25, 2017. [Core A]
- [19] BiPOM : Biological interlocked Process Ontology for metabolism. How to infer molecule knowledge from biological process ? *Vincent Henry, Fatiha Saïs, Elodie Marchadier, Juliette Dibie, Anne Goelzer and Vincent Fromion*. In proceedings of the 8th International Conference on Biomedical Ontologies (ICBO) 2017, Newcastle, United Kingdom, September 13-15, 2017. [Best paper].
- [20] RDF Data Evolution : Efficient Detection and Semantic Representation of changes. *Nathalie Pernelle, Fatiha Saïs , Daniel Mercier, Sujeeban Thiraisamy*. In proceedings of the 12th edition of the SEMANTiCS 2016, 12-15 September, 2016, Leipzig (Germany) [Short paper].
- [21] Key Discovery for Numerical Data : Application to Oenological Practices. *Danai Symeonidou, Isabelle Sanchez, Madalina Croitoru, Pascal Neveu, Nathalie Pernelle, Fatiha Saïs, Aurelie Roland-Vialaret, Patrice Buche, Aunur-Rofiq Muljarto, Remi Schneider*. In proceedings of the 22nd International Conference on Conceptual Structures (ICCS), pages : 222-236, LNCS 9717, Springer 2016, Annecy (France) [Short-listed best paper]. [Core B]
- [22] Logical Detection of Invalid SameAs Statements in RDF Data (2014). *Laura Papaleo, Nathalie Pernelle, Fatiha Saïs, Cyril Dumont*. In proceedings of 19th Knowledge Acquisition, Modeling and Management (EKAW), pages : 33-49, LNCS 8876, Springer 2014, pages 373-384, December, 2014, Linköping (Sweden). [Core B]
- [23] SAKey : Scalable Almost Key Discovery in RDF Data (2014). *Danai Symeonidou, Vincent Armant, Nathalie Pernelle, Fatiha Saïs*. In proceedings of the 13th International Semantic Web Conference (ISWC), pages : 33-49, LNCS 8796, Springer 2014, 19-23 October, Riva del Garda (Italy). [Core A]
- [24] Defining Key Semantics for the RDF Datasets : Experiments and Evaluations (2014). *Manuel Atencia, Michel Chein, Madalina Croitoru, Jérôme David, Michel Leclère, Nathalie Pernelle, Fatiha Saïs, François Scharffe, Danai Symeonidou*. In proceedings of the 21th International Conference on Conceptual Structures (ICCS), pages : 65-78, LNCS 8577, Springer 2014, Iasi (Romania). [Core B]
- [25] Automatic Composition of Form-Based Services in a Context-Aware Personal Information Space. *Rania Khefifi, Pascal Poizat and Fatiha Saïs*. In proceedings of 11th International Conference on Service Oriented Computing (ICSOC 2013), pages : 575-583, LNCS, Springer, 1-5 December, 2013, Berlin (Germany). [Core A]
- [26] Modelling and Querying Context-Aware Personal Information Spaces. *Rania Khefifi, Pascal Poizat and Fatiha Saïs*. In proceedings of the 23th International Conference on Database and Expert Systems Applications (DEXA 2012). Pages 103-110. September 3 - 7, 2012, Vienna (Austria) [Core B]
- [27] An Ontology-based method for Duplicate Detection in Web Data Tables. *Patrice Buche, Juliette Dibie-Barthelemy, Rania Khefifi and Fatiha Saïs*. In proceedings of the 22nd International Conference on Database and Expert Systems Applications (DEXA 2011). Part I, LNCS 6860, pages : 511-525, August 2011, Toulouse (France). [Core B]
- [28] Ontology-Driven Possibilistic Reference Fusion. *Fatiha Saïs, Rallou Thomopoulos and Sébastien Destercke*. In the 9th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 2010), Springer LNCS 6427, pages : 1079-1096. October 2010, Crete (Greece).
- [29] Reference Fusion and Flexible Querying. *Fatiha Saïs, Rallou Thomopoulos*. In the 7th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE-OTM 2008), pages 1541-1549, Springer LNCS 5332, November 2008, Monterrey (Mexico).
- [30] L2R : a Logical method for Reference Reconciliation. *Fatiha Saïs, Nathalie Pernelle, Marie-Christine Rousset*. Proceedings of the twenty-second Conference AAAI– on Artificial Intelligence, pages 329–334. July 2007, Vancouver – British Columbia (Canada). [Core A*]
- [31] A semantic enrichment of data tables applied to food risk assessment, *Hélène Gagliardi, Ollivier Haemmerlé, Nathalie Pernelle, Fatiha Saïs*. Proceedings of the 8th International Conference –DS on Discovery Science. Lecture Notes LNAI-3735 Springer Verlag, pages 374-376, October 2005, Singapour.

4.5 Workshops internationaux avec comité de lecture

- [32] Fact Validity Time Generation and Expansion in Knowledge Graphs. **Fatiha Saïs**, *Joana E. Gonzales Malaverri and Gianluca Quercini*. In proceedings of the second International Workshop on Hybrid Question Answering with Structured and Unstructured Knowledge, colocated with K-CAP 2019, Marina del Rey, LA (USA).
- [33] Discovering expressive rules for complex ontology matching and data interlinking *Manuel Atencia, Jérôme David, Jérôme Euzenat, Liliana Ibanescu, Nathalie Pernelle, Fatiha Saïs, Élodie Thiéblin, Cássia Trojahn*. In proceedings of the 14th International Workshop on Ontology Matching, colocated with the 18th International Semantic Web Conference ISWC-2019 October 26th, 2019, Auckland (New Zealand). [Poster]
- [34] Generating referring expressions from knowledge graphs. *Armita Khajeh Nassiri, Nathalie Pernelle, Fatiha Saïs*. In proceedings of the 14th International Workshop on Ontology Matching, colocated with the 18th International Semantic Web Conference ISWC-2019, October 26th, 2019, Auckland (New Zealand) [Poster]
- [35] N2R-Part : Identity Link Discovery using Partially Aligned Ontologies. *Nathalie Pernelle, Fatiha Saïs, Brigitte Safar, Maria Koutraki and Tushar Ghosh*. In proceedings of the 2nd International workshop on Open Data (WOD 2013), June 03, 2013 - Paris (France).
- [36] Discovering Keys in RDF/OWL Dataset with KD2R. *Danai Symeonidou and Nathalie Pernelle and Fatiha Saïs* In proceedings of the 2nd International workshop on Open Data (WOD 2013), Demo paper , June 03, 2013 - Paris (France).
- [37] Classification Rule Learning for Data Linking. *Nathalie Pernelle and Fatiha Saïs*. In proceedings of the 2nd International workshop on Linked Web Data Managment (LWDM 2012), in conjunction with the international conference EDBT 2012, Pages 136-139 , March 30, 2012, Berlin (Germany)
- [38] KD2R : a Key Discovery method for semantic Reference Reconciliation. *Danai Symeonidou, Nathalie Pernelle and Fatiha Saïs*. In proceedings of the 7th International IFIP Workshop on Semantic Web & Web Semantics (SWWS 2011) OTM Workshops 2011, pages : 392-401, Springer LNCS, October, 2011 : Heraklion (Greece).
- [39] LDM : Link Discovery Method for new Resource Integration. *Nathalie Pernelle and Fatiha Saïs*. In proceedings of Fourth International Workshop on Resource Discovery colocated with the 8th Extended Semantic Web Conference (ESWC 2011). Springer LNCS. May 29th, 2011 : Heraklion (Greece).
- [40] LN2R - a knowledge based reference reconciliation system : OAEI 2010 Results. **Fatiha Saïs**, *Nobal Niraula, Nathalie Pernelle and Marie-Christine Rousset*. In proceedings of the Fifth International Workshop on Ontology Matching, colocated with the 9th International Semantic Web Conference ISWC, CEUR-WS, Vol-689 pages : 172-179. November 2010, Shanghai (China).
- [41] An automatic ontology-based approach to enrich tables semantically. *Hélène Gagliardi, Ollivier Haemmerlé, Nathalie Pernelle, Fatiha Saïs*. In proceedings of the first International Workshop AAAI- C&O on Context and Ontologies : Theory, Practice and Application. AAAI Press, July 2005, Pittsburgh (USA).
- [42] Enriching a Relational Data Warehouse by Integrating XML Data : report on the e.dot project applied to Microbiology. *Hélène Gagliardi, Ollivier Haemmerlé, Damiano Migliori, Nathalie Pernelle, Marie-Christine Rousset, Fatiha Saïs*. In proceedings of the second Franco-Japanese Workshop ISIP – on Information Search, Integration and Personalization, May 2005, Lyon (France).

4.6 Conférences nationales avec comité de lecture

- [43] Génération d'expressions référentielles à partir de graphes de connaissances RDF. *Armita Khajeh Nassiri, Nathalie Pernelle, Fatiha Saïs*. Dans les actes de la 20ème édition de la conférence Extraction et Gestion de Connaissances (EGC 2020) 27-31 Janvier, 2020, Bruxelles (Belgique).
- [44] Découverte de séquences graduelles partiellement ordonnées à partir de données RDF : application aux données d'expériences scientifiques. *Simon Ser, Fatiha Saïs, Maguelonne Teisseire*. Dans les actes de la 18ème édition de la conférence Extraction et Gestion de Connaissances (EGC 2018), pages : 227-238, Paris (France) 22-26 Janvier 2018.
- [45] Détection de liens d'identité contextuels dans une base de connaissances. *Joe Raad, Nathalie Pernelle, Fatiha Saïs*. Dans actes des 28es Journées francophones d'Ingénierie des Connaissances (IC 2017) associée à la plateforme PFIA, Caen (France), 3-7 Juillet, 2017. [Best paper].
- [46] Approche numérique pour l'invalidation de liens d'identité (owl :SameAs). *Dimitrios Christaras Papageorgiou, Nathalie Pernelle, Fatiha Saïs*. Dans actes des 28es Journées francophones d'Ingénierie des Connaissances (IC 2017) associées à la plateforme PFIA, Caen (France) 3 -7 Juillet, 2017.
- [47] Détection et Représentation des changements dans les sources de données RDF. *Daniel Mercier, Nathalie Pernelle, Fatiha Saïs, Sujeeban Thuraisamy* In proceedings of the 26th Journées francophones d'Ingénierie des Connaissances (IC), pages : 259-262, Juin 2016, Montpellier (France).
- [48] Fusion de données redondantes : une approche explicative. **Fatiha Saïs** and Rallou Thomopoulos. In Actes of the 16th days of "Extraction et Gestion des Connaissances", (EGC'2016), Revue des Technologies de l'Information - RNTI, pages 363-368, pages 257-262, du 20 au 22 Janvier 2016, Reims (France).

- [49] C-SAKey : une approche de découverte de clés conditionnelles dans des données RDF. *Nathalie Pernelle, Danai Symeonidou, Fatiha Saïs*. In proceedings of the 26th Journées francophones d'Ingénierie des Connaissances (IC), Mai 2015, Rennes (France).
- [50] Linked Data Annotation and Fusion driven by Data Quality Evaluation. *Ioanna Giannopoulou, Fatiha Saïs and Rallou Thomopoulos*. In Actes of the 15th days of "Extraction et Gestion des Connaissances", (EGC'2015), Revue des Technologies de l'Information - RNTI, du 27 au 30 Janvier 2015, Luxembourg.
- [51] SAKey : Scalable Almost Key Discovery in RDF Data (2014). *Danai Symeonidou, Vincent Armant, Nathalie Pernelle, Fatiha Saïs*. In acte of the 30th days of "Gestion de Données –Principes, Technologies et Applications". 14-17 October, Grenoble-Autrans (France).
- [52] Définition de la sémantique des clés dans le Web sémantique : un point de vue théorique. *Michel Chein, Madalina Croitoru, Michel Leclerc, Nathalie Pernelle, Fatiha Saïs, Danai Symeonidou*. In proceedings of the 25th Journées francophones d'Ingénierie des Connaissances (IC), 12-16 Mai 2014, Clermont Ferrand (France)
- [53] Découverte de liens d'identité entre instances décrites par des ontologies partiellement alignées. *Nathalie Pernelle, Fatiha Saïs, Brigitte Safar, Maria Koutraki and Tushar Ghosh*. In proceedings of IC conference in PFIA2013 the French Symposium of AI , 1-5 July 2013, Lille (France).
- [54] Modélisation et interrogation d'espaces d'informations personnelles sensibles au contexte. *Rania Khefifi, Pascal Poizat and Fatiha Saïs*. In Actes of the 12th days of "Extraction et Gestion des Connaissances", (EGC'2012), Revue des Technologies de l'Information - RNTI, Pages 573-574, 31 janvier - 3 février, Bordeaux (France).
- [55] Détection de redondances dans les tableaux guidée par une ontologie. *Rania Khefifi, Patrice Buche, Juliette Dibie-Barthelemy and Fatiha Saïs*. In Actes of the eleventh days of "Extraction et Gestion des Connaissances", (EGC'2011), Revue des Technologies de l'Information - RNTI, Pages 563-568. January 2011, Brest (France).
- [56] Explication de décisions de réconciliation de références : approche fondée sur les réseaux de Petri colorés. *Souhir Gahbiche, Nathalie Pernelle, Fatiha Saïs*. Pages 327-338, EGC (Extraction et Gestion de Connaissances). January 2010, Hammamet (Tunisia).
- [57] Evidential reference fusion and flexible querying. *Sebastien Destercke, Fatiha Saïs and Rallou Thomopoulos*. Rencontres francophones sur la Logique Floue et ses Applications. November 2009, Annecy (France).
- [58] Une méthode flexible de fusion de références. *Fatiha Saïs, Rallou Thomopoulos* Paru dans les 4èmes Journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA). Juin 2008, Toulouse (France) .
- [59] Approche logique pour la réconciliation de références. *Fatiha Saïs, Nathalie Pernelle, Marie-Christine Rousset*. Dans les actes des septièmes journées de la conférence EGC – Extraction et Gestion des Connaissances, Revue des Technologies de l'Information – RNTI, pages 623-634. January 2007, Namur (Belgique).
- [60] Réconciliation de références : une approche logique adaptée aux grands volumes de données. *Fatiha Saïs, Nathalie Pernelle and Marie-Christine Rousset*. Dans les actes de la quatrième conférence COSI – l'Optimisation et les Systèmes d'Informations, pages 521 –532, June 2007, Oran (Algeria).
- [61] Découverte de relations candidates à l'enrichissement d'un entrepôt thématique de données. *Hélène Gagliardi, Olivier Haemmerlé, Nathalie Pernelle and Fatiha Saïs*. Dans les actes de la quatrième conférence COSI – l'Optimisation et les Systèmes d'Informations, pages 142 –153, June 2006, Algiers (Algeria).
- [62] Enrichissement sémantique de documents XML représentant des tableaux. *Fatiha Saïs, Hélène Gagliardi, Olivier Haemmerlé and Nathalie Pernelle*. Dans les actes des cinquièmes journées de la conférence EGC – Extraction et Gestion des Connaissances, Revue des Technologies de l'Information – RNTI, pages 407-419. January 2005, Paris (France).

4.7 Workshops nationaux avec comité de lecture

- [63] Enriching Knowledge Bases with Temporal Meta-facts. *Joana Gonzales Malaverri, Fatiha Saïs, Gianluca Quercini*. In proceedings of the 1st Workshop "VERITA : Veracity in Data" 2019. January 22d, 2019, Metz (France).
- [64] On Evaluating the Quality of RDF Identity Links in the LOD. *Laura Papaleo, Nathalie Pernelle, Fatiha Saïs*. In proceedings of the 1st Workshop "From Open Sources to Web of Data" (SoWeDo 2014). 11 mai, 2014, Clermont Ferrand (France).
- [65] Data-Flow Oriented Service Composition : AI-Planning or Petri Nets ?. *Rania Khefifi, Pascal Poizat and Fatiha Saïs*. In proceedings of the 2nd Workshop "From Open Sources to Web of Data" (SOS-DLWD 2013). July 1st, 2013, Lille (France).
- [66] Modelling and Querying Context-Aware Personal Information Spaces. *Rania Khefifi, Pascal Poizat and Fatiha Saïs*. In proceedings of the 2nd Workshop "From Open Sources to Web of Data" (SOS-DLWD 2012). January 2012, Bordeaux (France).
- [67] Passage à l'échelle de la réconciliation de concepts et de la réconciliation de références : quelques points de comparaison. *Nathalie Pernelle, Fatiha Saïs*. Dans les actes de l'atelier DECOR –Passage à l'échelle des techniques de découverte de correspondances – des septièmes journées de la conférence EGC – Extraction et Gestion des Connaissances. January 2007, Namur (Belgique).
- [68] Découverte de relations candidates à l'enrichissement d'un entrepôt thématique de données. *Hélène Gagliardi, Olivier*

4.8 Communications scientifiques

1. Exposé invité (Décembre 2019) à DATAIA Days «Life Sciences & AI», organisée par DATAIA, Orsay. Titre : « Data Linking and Knowledge Discovery in RDF Data : Methods and Some Feedback from Agronomic Applications ».
2. Exposé invité (Décembre 2019) à la journée "Intelligence artificielle et Recherche d'Information" soutenue par les associations AFIA et ARIA, Paris. Titre : «Raffinement de Graphes de Connaissances : découverte de clés et invalidation de liens d'identité».
3. Tutoriel invité (Janvier 2019) à l'école d'hiver E-EGC – Privacy preserving, Reasonning, Explaining, à Metz. Titre : « Identity Management in the Web of Data ».
4. Tutoriel invité (Novembre 2018) au workshop "Symbolic methods for data-interlinking" de la conférence Internationale EKAW 2018, Nancy, France. Ce tutoriel a été réalisé en collaboration avec Nathalie Pernelle et Joe Raad. Titre : « Detection of Invalid Identity Links Statements in RDF Knowledge Graphs ».
5. Tutoriel sélectionné à WWW 2018 (The Web Conference). Tutoriel réalisé en collaboration avec Nathalie Pernelle et Danai Symeonidou. Titre : «Knowledge Graph Completion».
6. Exposé invité (Mars 2018) au métaséminaire du département CEPIA@INRA à AgroParisTech, Massy, France. Titre : « Data Linking and Key Discovery in Knowledge Graphs ».
7. Tutoriel invité (Novembre 2017) à la conférence BDA – Data management : Principles, Technologies and Applications, à Nancy. Titre : «Knowledge Graph Expansion and Enrichment».
8. Séminaire (Mars 2017) au LIST (Luxembourg Institute of Science and Technology), Luxembourg. Title : « Rule Discovery from RDF data »
9. Séminaire (Février 2017) au laboratoire TETIS (CIRAD, Irstea, AgroParisTech, CNRS) à Montpellier. Titre : « From data to Knowledge »
10. Tutoriel sélectionné (Janvier 2014) en conjonction avec la conférence EGC (Extraction et Gestion de Connaissances). Il a été réalisé en collaboration avec Nathalie Pernelle et Fayçal Hamdi. Titre : « Data linking for a Web of Data »
11. Séminaire joint I3 (Avril 2014) au laboratoire GREYC de l'université Caen Basse Normandie. Titre : «Problems and approaches around data linking»
12. Exposé invité (Novembre 2013) aux journées INRA sur les méthodes alternatives pour la réduction des expériences sur les animaux, organisées par INRA Clermont Ferrand. Titre : « Data linking for the reduction of the use of animals in scientific experiments »
13. Exposé invité (Juillet 2013) à l'atelier QetR (Qualité et Robustesse dans le Web de Données), en conjonction avec la conférence EGC 2013, Lille. Titre : « from linked data to the Web of Data »
14. Exposé invité (Septembre 2012) aux journées "Franco- Taiwanese Days" à Paris, organisées par le LRI lab and four Universities from China. Title : « Ontology-Based Data Reconciliation ».
15. Exposé invité (Juin 2010) aux journées MIA-INRA sur l'intégration de données, Montpellier. Titre : « Logical and Numerical Approach for reference reconciliation ».

5 Implication dans les projets de recherche

5.1 Responsabilité et coordination de projets scientifiques

[Total : 8 projets]

[8] Projet PSPC AIDA (AI for Digital Automation) (2019-2023), [budget : 1.7 ME]

Directrice scientifique et chef de fil pour les partenaires académiques de l'Université Paris Saclay (LRI, LIMSI, ENS, Centrale-Supélec et CEA) du projet de plateforme AIDA qui a pour principal objet de permettre d'infuser des capacités d'Intelligence Artificielle dans des systèmes ou applications visant à l'automatisation des opérations dans les entreprises. Cette « infusion » est simplifiée par une abstraction appelée « Agent Opérationnel Intelligent » (AOI) qui s'entoure de composants logiciels permettant à des non spécialistes des technologies d'IA d'intégrer une fonction à base d'IA dans une application de type business content management, business process management, etc.

Le projet est financé pour 42 mois par BPI France et réunit plusieurs partenaires industriels : *IBM, Softeam, Stet et DecisionBrain* et académiques de l'université Paris Saclay : *les laboratoires LRI et LIMSI, CentraleSupélec, l'ENS Paris Saclay et le CEA.*

[7] Projet WarmRules – Institut Convergence DataIA (2019-2022), [budget : 220KE]

Co-Responsable scientifique du projet WarmRules dont l'objectif est de proposer une nouvelle approche de détection de règles causales graduelles dans des bases de connaissances en s'appuyant sur les relations temporelles, les liens d'identité contextuels et des méthodes statistiques. Ce projet sera appliqué à deux bases de connaissances issues de deux études de cas différentes dans le domaine du réchauffement climatique. Il traitera d'une part de l'épineuse question de l'adaptation des organismes au changement climatique à travers l'étude du maïs. D'autre part, la généralité de l'approche proposée sera testée sur un cas d'étude concernant des données sur le riz. Le projet réunit plusieurs partenaires académiques : *le laboratoire LRI de l'université Paris Sud, AgroParisTech, l'unité GQE (INRA et Université Paris Sud), Graduate School of Agricultural and Life Sciences University of Tokyo et Graduate School of Agricultural Science, Kobe University, Japan.*

[6] Projet interne financé par le LRI – KeyGraph (2020), [budget : 5KE]

Co-responsable scientifique du projet GraphKey en collaboration entre les équipes LaHDAK et Galac du LRI. Ce projet vise d'une part à concevoir des algorithmes pour la découverte de clés-graphes correspondant à des patterns de graphes permettant d'identifier de façon unique une classe d'individus (e.g. des personnes, des lieux, des oeuvres) dans les graphes de connaissances. L'étude des propriétés théoriques des structures de graphes de ces patterns permettra de définir des stratégies efficaces pour leur découverte.

[5] Projet VASTE – Digicosme Paris Saclay (2017-2018), [budget : 50KE]

Responsable scientifique du projet Vaste qui vise à développer des méthodes d'évaluation de la véracité des faits et des événements en utilisant des techniques fondées sur l'utilisation d'ontologie. Les données du domaine de l'épidémiologie animale sont collectées et peuvent être exploitées comme cas d'utilisation pour évaluer les méthodes développées. Ce projet a été financé par le labex Digicosme Paris Saclay et regroupe plusieurs partenaires de : AgroParisTech, Université de Versailles Saint Quentin, CIRAD/Irstea (Montpellier) et Telecom ParisTech.

[4] Projet PEPS CNRS RD2C (2016-2018), [budget : 15KE] Le but du projet RD2C est de découvrir des règles à partir de données RDF pour améliorer la qualité des bases de connaissances scientifiques (e.g. FMA, EFO). Nous visons à combiner voire étendre les travaux issus de la programmation logique inductive (ILP) et des approches numériques pour développer une nouvelle approche de découverte de règles en logique du premier ordre avec des contraintes. Ces méthodes devront exploiter toute la richesse des données en tenant compte de leur incertitude, de leur complexité intrinsèque et de leur imprécision, particulièrement pour les valeurs numériques. Les approches développées permettront notamment aux scientifiques de mieux comprendre leurs données, d'enrichir les bases de connaissances ou de détecter des données erronées. Enfin, la démarche interdisciplinaire adoptée s'appuiera sur l'extension de méthodes numériques et des approches sémantiques s'appuyant sur des connaissances expertes pour la découverte de règles dans des données scientifiques.

Le projet réunit plusieurs partenaires académiques : *le laboratoire LRI de l'université Paris Sud, le laboratoire Heudiasyc de l'UTC de Compiègne, l'équipe Link d'AgroParisTech et le laboratoire LIPADE de l'université Paris 5.*

[3] Projet interne financé par le LRI – DEVIEW (2017-2019), [budget : 10KE]

Responsable scientifique du projet DEVIEW qui vise à concevoir des modèles et des méthodes pour l'évaluation de la véracité tout en fournissant des explications interprétables. Collaboration entre le LRI et Centrale-Supélec.

[2] Responsable scientifique pour le LRI du projet ANR Qualinca (2012-2016), [budget : 120KE]. Qualinca (Qualité et interopérabilité de grands catalogues documentaires) est un projet de recherche fondamentale labellisé par l'ANR-Contint 2011 qui vise à développer des mécanismes permettant de qualifier le niveau de qualité d'une base documentaire existante, d'améliorer le niveau de qualité d'une base, de maintenir un niveau de qualité donné en contrôlant les opérations de mises à jour de ces bases, et d'exploiter ces bases en prenant en compte leur divers niveaux de qualité.

Le projet réunit plusieurs partenaires académiques : *le laboratoire LRI de l'université Paris Sud, le laboratoire LIG de l'Université Joseph-Fourrier, le laboratoire LIRMM de l'université Montpellier 2, l'INA (Institut National de l'Audio-visuel) et l'ABES (Agence Bibliographique de l'Enseignement Supérieur).*

[1] Collaboration project – HEDI (2009-2010), [budget : 25KE]

Co-responsable scientifique avec Nathalie Pernelle (Maître de Conférences à l'Université Paris Sud) du projet de collaboration HEDI – Heterogeneous Electronic Data Integration Integration entre l'équipe LEO de l'INRIA Saclay Île-de-France et Thales Corporate Services (2009-2010). L'objectif de ce contrat était de fournir à Thales une aide experte pour le développement d'un outil interne dédié à la réconciliation de descriptions de produits électroniques.

5.2 Participation à des projets scientifiques

[Total : 10 projets]

[10] Projet LIONES du Center of Data Science de l'IDEX Paris Saclay (2015-2018).

Le but du projet pluridisciplinaire LIONES est de proposer une représentation ontologique des processus de transformation des microorganismes : représentation des opérations élémentaires, appelées étapes du processus, la succession des étapes, les objets impliqués ainsi que leurs observations à différentes échelles et leur évolution durant le processus. La succession des étapes au cours du processus et les changements ou transformations des objets pendant ces étapes sont représentées par des liens sémantiques. Cette représentation ontologique permet d'une part de modéliser le processus de transformation afin d'en étudier les impacts environnementaux et socio-économiques et d'autre part la détection et la découverte de nouvelles liaisons sémantiques et intra-échelles et inter-échelles entre les objets mais également durant leur changement à travers les différentes étapes du processus.

Le projet réunit plusieurs partenaires dans deux disciplines : *l'informatique (le laboratoire LRI de l'Université Paris Sud et l'équipe Link d'AgroParisTech) et la biologie (unité GMPA JRU, INRA).* Le projet LIONES a été financé par le Center for Data Science (CDS) de l'IDEX Paris Saclay suite à un appel projets en 2015.

[9] SATT Paris Saclay project – DataForYou (2018-2020).

Membre du projet SATT DataForYou visant à accompagner la création de la start-up DataForYou qui a pour objectif de construire des outils pour soutenir les collectivités territoriales (par exemple, les mairies, les administrations départementales) en France, d'une part, dans leur processus de mise en conformité avec le RGPD, amis aussi dans l'optimisation des services rendus au citoyens en s'appuyant sur des outils d'analyse comportementale des citoyens. Dans ce projet, j'ai été impliquée dans l'encadrement de deux ingénieurs sur le lot RGPD.

[8] CNRS Mastodons – QCM-BioChem (2018-2019).

Le projet QCM-BioChem résulte de la fusion de trois projets Mastodons, QualiBioConsensus, HyQual et DECADE et propose de s'attaquer aux défis posés par la très grande variété des données biologiques et chimiques et des méthodes d'analyse et de fouille sur ces données. L'originalité des approches développées dans QCM-BioChem repose sur la prise en compte et l'explicitation systématique de critères de qualité, au niveau des données, des méthodes d'analyse et de fouille, et jusqu'aux connaissances produites.

[7] CNRS Mastodons – HyQualiBio (2017-2018).

Le projet HyQualiBio est un projet du défi Mastodons (Mission Interdisciplinarité du CNRS) issu de la fusion des projets Mastodons QualiBioConsensus et HyQual en 2017. Ce projet proposait de s'attaquer aux défis posés par la très grande variété des données biologiques et des méthodes d'analyse et de fouille de données biologiques. L'originalité des approches développées dans HyQualiBio repose sur la prise en compte et l'explicitation systématique de critères de qualité, au niveau des données, des méthodes d'analyse et de fouille, et jusqu'aux connaissances produites.

[6] Projet ANR PIMI (2010-2013).

Le projet ANR PIMI a pour objectif la définition d'un environnement de conception et d'une plate-forme de déploiement pour les systèmes de gestion d'informations Personnelles (PIMS). Les futurs PIMS devront permettre aux utilisateurs de gérer et d'accéder aux données personnelles avec des services pertinents par rapport à leurs besoins et qui garantissent la confidentialité de leurs informations. Les ontologies sont utilisées pour décrire sémantiquement les informations personnelles, les services ainsi que les besoins des utilisateurs.

Le projet réunit plusieurs partenaires académiques et industriels : comme *le LRI de l'université Paris-Sud, l'IRIT de l'université Paul Sabatier, le Lamsade de l'Université Paris -Dauphine et la société GeniGraph.*

[5] DataBridges activity of European EIT ICT Labs (2011-2012).

Participation à l'activité DataBridges "Data Integration in Digital Cities" de l'EIT (European Institute of Innovation and Technology) ICT Labs dont le but est la construction de systèmes ouverts et extensibles d'intégration de sources de données hétérogènes concernant les villes numériques - 2011 et 2012, dirigée par Ioana Manolescu (Directrice de Recherche Inria-Saclay).

[4] Projet industriel PossiblePasts avec la start-up ThinkStraight (2010).

La startup ThinkStraight a identifié un problème de « forensic » de données dans le contexte spécialisé des données financières (comptables). Le projet nous a permis d'analyser les questions scientifiques soulevées par leur problème, que nous avons trouvé novateur et de réaliser l'état de l'art des deux domaines les plus proches, à savoir le nettoyage des données et l'analyse de la qualité des données. Nous avons également analysé l'architecture du logiciel que ThinkStraight envisageait de construire et de la comparer aux technologies existantes développées au sein de l'équipe, notamment les prototypes LN2R et EdiFlow ; Nous avons constaté que le problème de ThinkStraight nécessitait des fonctionnalités radicalement différentes et, compte tenu des tout premiers stades du développement de l'entreprise, la collaboration n'a pas été poursuivie. Ce projet a été dirigé par Ioana Manolescu (Directrice de Recherche Inria-Saclay).

[3] Projet ANR GeOnto (2008-2011).

Le projet GeOnto a pour objectif la constitution, l'alignement, la comparaison et l'exploitation d'ontologies géographiques hétérogènes. Il porte sur l'interopérabilité de données diverses relatives à l'information géographique. Le projet réunit plusieurs partenaires académiques : l'équipe IASI/LEO de l'université Paris-Sud, IGN-Cogit, IRIT de l'université Paul Sabatier et le LIUPPA de l'université de Pau et des pays de l'Adour. Dans ce projet je suis intervenue dans le lot concernant la réconciliation d'instances d'ontologies géographiques combinée avec la réconciliation de schémas (taxonomies).

[2] Projet ANR/RNTL WebContent (2006-2010).

L'objectif du projet est de produire une plate-forme flexible et générique pour la gestion de contenus et l'intégration des technologies du Web Sémantique dans le but de montrer leur utilité sur des applications réelles à fort impact économique ou sociétal. Il a impliqué une quinzaine de partenaires industriels et académiques dont l'INRIA LEO, CEA, Exalead, EADS, INRA, etc. Dans ce projet la méthode d'annotation sémantique de tableaux XTab2SML a été complétée par le partenaire INRA, elle est en cours d'intégration dans la plateforme logicielle du projet.

[1] Projet industriel PICSEL3 en collaboration avec France Télécom R&D (2004-2007)

Le projet PICSEL3 – *Production d'Interfaces à base de Connaissances pour des Services En Ligne* – est le troisième d'une série de projets de collaboration entre l'équipe IASI du LRI et France Télécom R&D durant dix ans. Le but est de développer un système d'intégration mixte de données pouvant jouer le rôle de médiateur entre des sources distantes et un entrepôt local de données. C'est dans le cadre de ce projet que se sont inscrits mes travaux sur la réconciliation de références guidée par une ontologie.

6 Animation scientifique

6.1 Implication dans les GdRs nationaux

Co-présidente de l'action RoD – Raisonner sur les Données – soutenue par le GdR Madics (Masses de Données, Informations et Connaissances en Sciences) et le GdR IA, depuis 2019. Deux demi-journées les 25 et 26 mai 2020, sont en cours d'organisation dans le cadre des journées plénières du GdR IA 2020, UTC de Compiègne.

6.2 Edition de numéros spéciaux dans des revues

Revues internationales

Co-éditrice du Springer Nature Journal - Computer Science Topical Issue on : Web for Information and Knowledge Exploration, Sharing and Security. Editeur : Springer Nature Switzerland AG. [à paraître en 2020]

Revues nationales

- Co-éditrice du numéro spécial de la revue ISI (Ingénierie des Systèmes d'Informations) : *Web des données : publications, liage et capitalisation*, Hermann Editions 2016.
- Co-éditrice du numéro spécial de la revue RNTI (Revue des Nouvelles Technologies de l'Information) : *Des Sources Ouvertes au Web de Données*, Hermann Editions 2014.
- Co-éditrice du numéro spécial de la revue RNTI (Revue des Nouvelles Technologies de l'Information) : *Mesurer et évaluer la Qualité des Données et des Connaissances*, Hermann Editions 2012.

6.3 Organisation de conférences et de workshops

- Co-présidente du comité de programme de la conférence internationale Web2Touch Conference Track, on Semantic Technologies for *Smart Information Sharing and Web Collaboration*, @ 29th IEEE WETICE 2020, Bayonne, France, June 20-12, 2020.
- Co-présidente du comité de programme de la conférence internationale Web2Touch Conference Track, on Semantic Technologies for *Smart Information Sharing and Web Collaboration*, @ 28th IEEE WETICE 2019, Capri (Napoli), Italy, June 12-14, 2019.
- Co-présidente des comités scientifique et d'organisation de la journée EGC-IA 2019 soutenue par les associations EGC et AFIA sur le thème de *Découverte de connaissance dans le Web des données*, 10 Mai 2019, Orsay, France.
- Co-organisatrice de l'atelier "Véracité dans les données" (VERITA) organisé dans le cadre de la conférence EGC (Extraction et Gestion de Connaissances) 2019.
- Présidente du comité de programme et membre du comité d'organisation de "Paris-Saclay Junior Conference on Data Science and Engineering (JDSE) en 2018.
- Co-présidente scientifique de l'école d'hiver e-EGC sur le thème "Social Networks" co-organisée avec la conférence Francophone EGC (Extraction et Gestion de Connaissances) 2018, January 22-26, Paris.
- Membre du comité d'organisation de la conférence Francophone EGC (Extraction et Gestion de Connaissances) 2018, Janvier 22-26, 2018, Paris, France.
- Co-organisatrice de l'atelier "From Open Sources to Web of Data" (SoWeDo) organisé dans le cadre de la conférence IC (Ingénierie des Connaissances) 2014 et 2016.
- Co-organisatrice des éditions 2012 et 2013 de l'atelier "From Open Sources to Web of Data" (SOS-DLWD) organisé dans le cadre de la conférence EGC (Extraction et Gestion de Connaissances) 2012 et dans le cadre de la conférence IC 2013.
- Co-organisatrice du workshop interne à l'Action Line Digital Cities de EIT –European Institute of Innovation and Technology–ICT Labs, projet co-financé par la communauté européenne, en 2012.
- Co-chair des éditions 2009, 2010 et 2011 de l'atelier EvalECD sur l'Évaluation des méthodes d'Extraction de Connaissances dans les Données – associé à la conférence EGC (Extraction et Gestion de Connaissances).

6.4 Présidence de sessions dans des conférences nationales et internationales

- Présidence de session dans la conférence internationale ACM SAC 2020.
- Présidence de session dans la conférence internationale K-CAP 2019.
- Présidence de plusieurs sessions de la conférence internationale Web2Touch-Wetice en 2018 et en 2019.
- Présidence de session dans la conférence internationale ICCS en 2016.
- Présidence de sessions dans les éditions de la conférence EGC en 2011, 2016, 2018, 2019, 2020.

7 Evaluation de la recherche

7.1 Expertise de projets de recherche

- Expertise de projets pour le programme ANR CONTINT en 2016.
- Expertise de projets pour le programme ANR Jeunes Chercheuses et Jeunes Chercheurs en 2016.
- Expertise de projets pour l'appel à projets Chercheur(se)s d'avenir pour la région Languedoc-Roussillon en 2013 et en 2015.
- Expertise de projets pour le programme ANR Jeunes Chercheuses et Jeunes Chercheurs en 2012.

7.2 Membre de jurys de thèse

1. Rapporteur de la thèse de Justine Reynaud, soutenue le 10 décembre 2019, à l'Université de Lorraine. Titre : "Découverte de définitions dans le web des données".
2. Rapporteur de la thèse de Mussab Zneika, soutenue le 20 septembre 2019, à l'Université Paris-Seine, Université Cergy-Pontoise. Titre : "Querying Semantic Web/Linked Data Graphs Using Summarization".
3. Examinatrice pour la thèse de Maria Koutraki, soutenue le 27 septembre 2016, à l'Université de Versailles Saint-Quentin-en-Yvelines (UVSQ). Titre : "Schema Discovery for Web Data".
4. Examinatrice pour la thèse de Soumia Lilia Berrahou, soutenue le 29 septembre 2015, à l'Université Montpellier 2. Titre : "Extraction of arguments of n-ary relations in the texts guided by a RTO of domain".
5. Examinatrice pour la thèse de Laurie Serrano, soutenue le 24 janvier 2014, à l'Université Caen Basse-Normandie. Titre : "Towards user-oriented knowledge capitalization : automatic extraction and structuring of information from open sources".

7.3 Membre de comités de sélection

1. Membre extérieur du comité de sélection à Centrale-Supélec, Université Paris Saclay pour un recrutement d'un Enseignant Chercheur en Informatique, en 2020.
2. Membre interne du comité de sélection (COS) à Université Paris Sud (IUT d'Orsay) pour un recrutement d'un maître de conférences, section CNU 27 en 2019.
3. Membre extérieur du comité de sélection (COS) à l'Université de Caen Basse-Normandie pour un recrutement d'un maître de conférences, section CNU 27 en 2017.
4. Membre extérieur du comité de sélection (COS) à l'Université de Cergy-Pontoise pour un recrutement d'un maître de conférences, section CNU 27, en 2016.
5. Membre extérieur du comité de sélection (section 27) de l'Université Lyon 1, pour un recrutement d'un maître de conférences, section CNU 27 en 2010.
6. Membre du jury de sélection des candidats au concours 2011 de Chargé de Recherche Cemagref sur *les Entre-pôts de données spatialisées : conception et extraction de connaissances*.

7.4 Participation à des comités de programme

Comités de programme de Conférences internationales

- European Conference on Artificial Intelligence ECAI 2016 et 2020.
- International Conference on Conceptual Structures ICCS 2014–2020
- ACM/SIGAPP International Symposium On Applied Computing - Track : Semantic Web and Applications, 2020.
- International conference on Knowledge Capture, K-Cap 2019.
- ACS/IEEE International Conference on Computer Systems and Applications (AICCSA). Track : Knowledge Management & Ontology engineering, 2017 – 2019.
- Annual International Symposium on Information Management and Big Data (SIMBig), 2016–2019.
- International Conference on Information Systems and Economic Intelligence in 2012
- International Conference on Advances in Databases, Knowledge, and Data Applications 2013 and 2014
- International Database Engineering and Applications Symposium (IDEAS) en 2008.
- International Conference IEEE, IRI Information Reuse and Integration 2007 et 2009.

Comités de programme de Workshops internationaux

- International Workshop HQA (2018 and 2019) : Second International Workshop on Hybrid Question Answering with Structured and Unstructured Knowledge, co-located with WWW in 2018 and with K-CAP in 2019.
- International Workshop on Graph Structures for Knowledge Representation and Reasoning in 2015-2020
- International Workshop on Resource Discovery co-located with the international conference ESWC 2011, 2012, 2013 and 2014.
- OTM Academy PhD workshop from 2009 to 2014.
- Student Workshop of The International ACM Conference on Management of Emergent Digital EcoSystems (MEDES 2009).

Comités de programme de Conférences nationales et ateliers nationaux

- Conférence Francophone sur l'Extraction et la Gestion de Connaissances (EGC) 2011–2020.
- Conférence Francophone sur l'ingénierie des Connaissances (IC) 2015– 2020.
- Conférence sur la Gestion de Données – Principes, Technologies et Applications (BDA), 2019.
- Paris-Saclay Junior Conference on Data Science and Engineering (jDSParis16) en 2016–2019.
- Conférence maghrébine sur l'Extraction et la Gestion de Connaissances EGC-M en 2010 et en 2011.
- Atelier Qualité des Données du Web à la conférence EGC 2016.
- Workshop "Constitution, enrichissement et exploitation de ressources GEOgraphiques" pour l'analyse de données (CerGEO) 2014.
- Atelier IN-OVIVE associé à la conférence IC 2013–2019.
- Atelier Gestion et Analyse des données Spatiales et Temporelles à la conférence EGC 2014 et 2015.
- Atelier Gestion et Analyse des données Spatiales et Temporelles (GAST) 2015.
- Atelier "Constitution, enrichissement et exploitation de ressources GEOgraphiques" pour l'analyse de données à la conférence EGC 2014.
- Atelier Ontologies et jeux de données pour évaluer le Web Sémantique, à IC 2012.
- Atelier Qualité et Robustesse dans le Web de Données associé à la conférence IC 2011, 2012 et 2013.
- Atelier COTA 2010 et 2011– Cotation des Informations : Théorie et Applications– associé à la conférence IC 2010 et 2011.
- Atelier Web Social associé à la conférence EGC 2010 et 2011.

Relectures pour des journaux internationaux

- International Journal of Web Semantics (JWS) from 2012, 2017, 2018 et 2019.
- International Journal of Discrete Applied Mathematics 2018.
- International Journal of Knowledge Based Systems 2015.
- International Journal on Data Semantics (JoDS) from 2013 to 2015.
- International Journal of Metadata, Semantics and Ontologies (IJMSO) 2012
- Reviewer for *KDE Journal –Data and Knowledge Engineering – Special Issue on Contribution of Ontologies in Designing Advanced Information Systems* 2009.

Relectures pour des conférences et workshops internationaux. IEEE International Conference on Data Engineering ICDE 2014, SIGMOD 2015, The 8th Extended Semantic Web Conference (ESWC) in 2011, International Conference on Conceptual Structures in ICCS 2014, IJCAI International Workshop on Graph Structures for Knowledge Representation and Reasoning in 2015, International Conference on Information Systems and Economic Intelligence in 2012, International Conference on Advances in Databases, Knowledge, and Data Applications 2013 and 2014, International Workshop on Resource Discovery co-located with the international conference ESWC 2011, 2012, 2013 and 2014, International Database Engineering and Applications Symposium (IDEAS) en 2008, International Conference IEEE, IRI Information Reuse and Integration 2007 et 2009.