

Curriculum Vitæ

Adrien Coulet

16 décembre 2020

Table des matières

1	Identification et cursus	2
2	Activités d'enseignement	4
3	Activités scientifiques	6
3.1	Publications	6
3.2	Autres productions scientifiques	12
3.3	Responsabilités scientifiques	13
3.4	Encadrement doctoral et scientifique	14
3.5	Diffusion des travaux (rayonnement et vulgarisation)	15

1 Identification et cursus

Généralités

Prénom et nom : Adrien Coulet
Date et lieu de naissance : 14/11/1979 à Montpellier, France
Nationalité : Française

Situation familiale : marié, trois enfants

Email : adrien.coulet@inria.fr
Téléphone : +33 6 48 69 61 52
Adresse postale : Centre de Recherche des Cordeliers
Équipe 22 “HeKA” , Escalier D, 1er étage
15 rue de l’école de médecine
75016 Paris, France

Situation actuelle

Chercheur à Inria

depuis le 1^{er} septembre 2020, je suis en détachement à Inria où je co-pilote avec Sarah Zohar la création d’une équipe de recherche appelée HeKA, qui sera commune à l’Inserm, Inria et l’Université de Paris.

HeKA fait suite à l’équipe 22 dirigée par Anita Burgun et s’intéresse aux approches à partir de données et de modèles qui permettent l’acquisition de connaissances biomédicales. De façon générale, je m’intéresse au développement de modèles prédictifs en santé, avec l’objectif qu’ils soient utilisés et aient un impact en pratique clinique. De façon plus spécifique, je m’intéresse d’un point de vue méthodologique à l’apport des connaissances de domaine dans les modèles d’apprentissage, et d’un point de vue applicatif à la variabilité inter-individuelle de la réponse aux médicaments.

depuis février 2020, je suis qualifié pour les postes de professeurs des universités, par la section 27 du CNU.

Situations précédentes

2010-20 **Maître de conférences** en informatique à l’Université de Lorraine à Nancy
chercheur dans l’équipe Orpailleur, du laboratoire **Loria** (Université de Lorraine, CNRS, Inria) et enseignant à **TELECOM Nancy**, école d’ingénieurs en informatique de l’Université de Lorraine.

2017-19 **Chercheur invité** à l’Université Stanford, Californie
en délégation au CNRS (2017-18), puis à Inria (2018-19), en mobilité géographique dans l’équipe de Nigam H. Shah, au **Stanford Center for Biomedical Informatics Research**.

2008-10 **Post-doctorant** à l’Université Stanford, Californie
sur *l’extraction de relations à partir de textes et l’annotation automatique de bases de données biomédicales*, sous la direction de Russ B. Altman et Mark A. Musen au **Stanford Center for Biomedical Informatics Research**.

2004-08 **Doctorant** en informatique à l’Université Henri Poincaré, Nancy
sur *l’intégration de données et la découverte de connaissances biomédicales guidées par les connaissances de domaines*, sous la direction de Marie-Dominique Devignes et Malika Smail-Tabbone dans l’équipe Orpailleur du **Loria**.

Diplômes

- 2019 HDR** en informatique de l'Université de Lorraine, Nancy
intitulée *Mises en correspondances de données, textes et connaissances pour la découverte de connaissances biomédicales*
parrainée par Amedeo Napoli
rapportée par Sarah Cohen-Boulakia, Nathalie Aussenac-Gilles et Olivier Curé,
examinée par Olivier Dameron et Marie-Dominique Devignes
- 2008 Doctorat** en informatique à l'Université Henri Poincaré, Nancy
intitulé *Construction et utilisation d'une base de connaissances pharmacogénomique pour l'intégration de données et la découverte de connaissances*
sous la direction de Marie-Dominique Devignes et Malika Smaïl-Tabbone
rapporté par Alain Viari et Mohand-Saïd Hacid,
examiné par Chantal Raynaud et Nacer Boudjlida
- 2004 Master** Ingénierie pour la Santé et le Médicament, Université Joseph Fourier, Grenoble
spécialité "Approche spatio-temporelle des systèmes vivants"
Stage au Centre d'Innovation Biologique du CHU de Grenoble, sur le clustering de séquences d'expression génique sous la direction de Michel Sève et Jacques Demongeot
- 2003 Diplôme d'ingénieur** en informatique, Polytech'Grenoble, Université Joseph Fourier, Grenoble
spécialité "Technologies de l'Information pour la Santé"
Deux stages marquants : le 1^{er} à Inria Rhône-Alpes sur la conception d'une base de données pour la protéomique avec Alain Viari et François Rechenmann, le 2nd à l'UMR 5094 (CNRS, Université de Montpellier 1) autour d'un criblage moléculaire pour la recherche de nouveaux antibiotiques
- 2000 DEUG en sciences de la vie**, Université Montpellier 2, Montpellier

2 Activités d'enseignement

En bref

J'ai enseigné de 2010 à 2020 à TELECOM Nancy, une école d'ingénieurs en informatique de l'Université de Lorraine, et suis intervenu régulièrement dans les Masters d'Informatique et de Mathématiques à l'Université de Lorraine. Les thèmes de mes enseignements sont essentiellement la gestion de données, la gestion de connaissances et les méthodes d'intelligence artificielle symboliques et numériques. Cette année (2020-21) je vais enseigner une trentaine d'heures sur l'apprentissage automatique et l'inférence causale, en M2 d'informatique médicale à l'Université de Paris.

Responsabilités d'enseignements

J'ai été responsable des modules d'enseignement suivants :

- *Systèmes à base de connaissances**, TELECOM Nancy, 3^{ème} année, 50 étudiants, 30h, pendant 8 ans
- *Intelligence artificielle***, TELECOM Nancy, 3^{ème} année et M2 Mathématiques, 60 étudiants, 30h, 1 an
- *Web sémantique et raisonnement**, M2 Informatique, 15 étudiants, 20h, pendant 7 ans
- *Gestion de masses de données**, TELECOM Nancy, 2^{ème} année, 50 étudiants, 46h, pendant 7 ans
- *Bases de données***, TELECOM Nancy, 1^{ère} année, 130 étudiants, 40h par étudiant, pendant 6 ans

Autres enseignements

J'ai participé aux modules suivants :

- *Ingénierie des connaissances***, M2 Informatique, 15 étudiants, 20h, pendant 4 ans
- *Introduction à l'intelligence artificielle*, TELECOM Nancy, 2^{ème} année, 40 étudiants, 30h, pendant 1 an
- *Méthodes et outils de conception en informatique***, TELECOM Nancy, 2^{ème} année, 130 étu., 60h, pdt 4 ans
- *Projet de conception et de développement (PHP, Java, Python)*, TELECOM Nancy, 2^{ème} année, 130 étudiants, 14h, pendant 4 ans

* : J'ai développé la totalité des supports et matériaux pédagogiques des modules marqués avec une seule astérisque.

** : J'ai développé environ la moitié des supports et matériaux pédagogiques des modules marqués avec deux astérisques.

Responsabilités administratives liées à l'enseignement

A TELECOM Nancy, j'ai piloté en 2014 la création de la spécialité *Intelligence Artificielle et Masses de Données* qui forme les étudiants aux outils et méthodes associées à la science des données, avant d'être responsable de 2014 à 2017 de cette spécialité, suite à sa création. Cette responsabilité a consisté tout d'abord en l'établissement d'un contenu pédagogique cohérent et en lien avec les besoins industriels, puis en la mise en place et la coordination des modules d'enseignement et séminaires, au suivi des étudiants et à la mise au point de l'emploi du temps. Environ 50 étudiants (25 en 1^{ère} année et 25 en 2^{ème}) suivent cet approfondissement chaque année. J'ai organisé annuellement un hackathon d'une semaine pendant lequel les étudiants s'attachaient à valoriser des jeux de données fournis par des partenaires industriels ou académiques. La création de cette spécialité, assez précurseur à l'époque

(2014), a été récompensée par le prix de la formation innovante en 2015 par l'association *Pasc@line* (association d'établissement de formation au numérique et d'entreprises du secteur, renommée depuis *Talents du numérique*).

De 2011 à 2014, j'ai été responsable de la spécialité *Système d'Information des Entreprises* de TELECOM Nancy, ce qui impliquait le même type de tâches de coordination que celles décrites précédemment. En 2013, j'ai participé activement à la mise en place de la *formation ingénieur par apprentissage* de TELECOM Nancy. J'ai participé en 2017 à la création de la nouvelle maquette de l'offre de Master en informatique de l'Université de Lorraine en proposant et décrivant des modules d'enseignements.

J'ai également participé très régulièrement aux oraux des concours et à l'examen des dossiers d'admission à TELECOM Nancy, et aux journées portes ouvertes de l'école.

3 Activités scientifiques

En bref

Mon activité de recherche s'intéresse à l'extraction de connaissances à partir de données, au développement de modèles prédictifs et aux applications biomédicales de ces types d'approches informatiques. J'y étudie en particulier le rôle que peuvent jouer les connaissances de domaines. D'un point de vue applicatif je m'intéresse particulièrement à la médecine de précision et à étudier comment mieux prendre en considération la variabilité inter-individuelle de réponse aux médicaments.

3.1 Publications

Le contexte et contenu de cinq publications sélectionnées [R4, R1, R2, A1, R5] sont présentés à la fin de cette section. Pour une vue générale, le tableau suivant décompte mes publications par type.

Revue internationale :	15
Conférences internationales :	12
Ateliers internationaux :	13
Conférences nationales :	4
Ateliers nationaux :	3
Articles de vulgarisation :	2
Editoriaux :	2
Posters et démonstrations (non listés) :	>20

Revue internationale avec comité de lecture

- [R1] Joël LEGRAND, Romain GOGDEMIR, Cédric BOUSQUET, Kevin DALLEAU, Marie-Dominique DEVIGNES, William DIGAN, Chia-Ju LEE, Ndeye-Coumba NDIAYE, Nadine PETITPAIN, Patrice RINGOT, Malika SMAÏL-TABBONE, Yannick TOUSSAINT et Adrien COULET. "PGxCorpus, a manually annotated corpus for pharmacogenomics". In : *Scientific Data* 7.1 (2020), p. 1-13.
- [R2] Pierre MONNIN, Joël LEGRAND, Graziella HUSSON, Patrice RINGOT, Andon TCHECHMEDJIEV, Clément JONQUET, Amedeo NAPOLI et Adrien COULET. "PGxO and PGxLOD : a reconciliation of pharmacogenomic knowledge of various provenances, enabling further comparison". In : *BMC Bioinformatics* 20.S4 (2019), p. 139.
- [R3] Fatima RODRIGUEZ, Sukyung CHUNG, Manuel R. BLUM, Adrien COULET, Sanjay BASU et Latha P. PALANIAPPAN. "Atherosclerotic Cardiovascular Disease Risk Prediction in Disaggregated Asian and Hispanic Subgroups Using Electronic Health Records". In : *Journal of the American Heart Association* 8.14 (2019), e011874.
- [R4] Adrien COULET, Nigam H. SHAH, Maxime WACK, Mohammad CHAWKI, Nicolas JAY et Michel DUMONTIER. "Predicting the need for a reduced drug dose, at first prescription". In : *Scientific Reports* 8.1 (2018), p. 1-11.
- [R5] Kevin DALLEAU, Yassine MARZOUGUI, Sébastien DA SILVA, Patrice RINGOT, Ndeye Coumba NDIAYE et Adrien COULET. "Learning from biomedical linked data to suggest valid pharmacogenes". In : *Journal of Biomedical Semantics* 8.29 (2017), p. 16.
- [R6] Gabin PERSONENI, Emmanuel BRESSO, Marie-Dominique DEVIGNES, Michel DUMONTIER, Malika SMAÏL-TABBONE et Adrien COULET. "Discovering associations between adverse drug events using pattern structures and ontologies". In : *Journal of Biomedical Semantics* 8.16 (2017), p. 29.
- [R7] Yi LIU, Adrien COULET, Paea LEPENDU et Nigam H. SHAH. "Using ontology-based annotation to profile disease research". In : *Journal of the American Medical Informatics Association* 19.e1 (2012), e177-e186.

- [R8] Matthias SAMWALD, Adrien COULET, Iker HUERGA, Robert L POWERS, Joanne S LUCIANO, Robert R FREIMUTH, Frederick WHIPPLE, Elgar PICHLER, Eric PRUD'HOMMEAUX, Michel DUMONTIER et M Scott MARSHALL. "Semantically enabling pharmacogenomic data for the realization of personalized medicine". In : *Pharmacogenomics* 13.2 (2012), p. 201-212.
- [R9] Adrien COULET, Yael GARTEN, Michel DUMONTIER, Russ B ALTMAN, Mark A MUSEN et Nigam H SHAH. "Integration and publication of heterogeneous text-mined relationships on the Semantic Web." In : *Journal of Biomedical Semantics* 2.Suppl 2 (2011), S10.
- [R10] Adrien COULET, Malika SMAÏL-TABBONE, Amedeo NAPOLI et Marie-Dominique DEVIGNES. "Ontology-based knowledge discovery in pharmacogenomics". In : *Advances in Experimental Medicine and Biology, Springer* 696 (2011), p. 357-66.
- [R11] Clement JONQUET, Paea LEPENDU, Sean M. FALCONER, Adrien COULET, Natalya Fridman NOY, Mark A. MUSEN et Nigam H. SHAH. "NCBO Resource Index : Ontology-based search and mining of biomedical resources". In : *Journal of Web Semantics* 9.3 (2011), p. 316-324.
- [R12] Adrien COULET, Nigam H. SHAH, Yael GARTEN, Mark A. MUSEN et Russ B. ALTMAN. "Using text to build semantic networks for pharmacogenomics". In : *Journal of Biomedical Informatics* 43.6 (2010), p. 1009-1019.
- [R13] Yael GARTEN, Adrien COULET et Russ B ALTMAN. "Recent progress in automatically extracting information from the pharmacogenomic literature". In : *Pharmacogenomics* 11.10 (2010), p. 1467-1489.
- [R14] Adrien COULET, Malika SMAÏL-TABBONE, Pascale BENLIAN, Amedeo NAPOLI et Marie-Dominique DEVIGNES. "Ontology-guided data preparation for discovering genotype-phenotype relationships". In : *BMC Bioinformatics* 9.S3 (2008).
- [R15] Benjamin LEBLANC, Adrien COULET, Estelle ANDRÉ, Franck MOLINA et Jean-Paul LEONETTI. "Data processing and exchange tools to facilitate chemical genetic screening processes." In : *BioTechniques* 37.2 (2004), p. 223-5.

Conférences internationales avec comité de lecture

- [C1] Pierre MONNIN, Emmanuel BRESSO, Miguel COUCEIRO, Malika SMAÏL-TABBONE, Amedeo NAPOLI et Adrien COULET. "Tackling scalability issues in mining path patterns from knowledge graphs : a preliminary study". In : *1st international conference "Algebras, graphs and ordered sets" (Algos 2020)*. Nancy, France, août 2020.
- [C2] Pierre MONNIN, Miguel COUCEIRO, Amedeo NAPOLI et Adrien COULET. "Knowledge-Based Matching of n-ary Tuples". In : *25th International Conference on Conceptual Structures, ICCS 2020*. T. Lecture Notes in Computer Science. 12277. Sept. 2020, p. 48-56.
- [C3] Stephen PFOHL, Adrien COULET*, Ben MARAFINO*, Fatima RODRIGUEZ, Latha PALANIAPPAN et Nigam H. SHAH. "Creating Fair Models of Atherosclerotic Cardiovascular Disease Risk". In : *AIES 2019 - 2nd AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*. 2019, *contributed equally.
- [C4] Gabin PERSONENI, Marie-Dominique DEVIGNES, Malika SMAÏL-TABBONE, Philippe JONVEAUX, Céline BONNET et Adrien COULET. "Cooperation of bio-ontologies for the classification of genetic intellectual disabilities : a diseasome approach". In : *SWAT4HCLS 2018 - 11th International Conference on Semantic Web Applications and Tools for Healthcare and Life Sciences*. Antwerp, Belgium, déc. 2018.
- [C5] Pierre MONNIN, Mario LEZOCHE, Amedeo NAPOLI et Adrien COULET. "Using formal concept analysis for checking the structure of an ontology in LOD : the example of DBpedia". In : *ISMIS 2017 - 23rd International Symposium on Methodologies for Intelligent Systems*. LNCS series. Warsaw, Poland : Springer, juin 2017.
- [C6] Kevin DALLEAU, Ndeye COUMBA NDIAYE et Adrien COULET. "Suggesting valid pharmacogenes by mining linked data". In : *SWAT4LS 2015 - International Conference on the Semantic Web Applications and Tools for Life Sciences*. Cambridge, United Kingdom, 2015.

- [C7] Gabin PERSONENI, Simon DAGET, Céline BONNET, Philippe JONVEAUX, Marie-Dominique DEVIGNES, Malika SMAÏL-TABBONE et Adrien COULET. “ILP for Mining Linked Open Data”. In : *ILP 2014 - the 24th International Conference on Inductive Logic Programming*. 2014.
- [C8] Gabin PERSONENI, Simon DAGET, Céline BONNET, Philippe JONVEAUX, Marie-Dominique DEVIGNES, Malika SMAÏL-TABBONE et Adrien COULET. “Mining Linked Open Data : A Case Study with Genes Responsible for Intellectual Disability”. In : *DILS 2014 - 10th International Conference on Data Integration in the Life Sciences*. T. 8574. LNCS series. Springer, juil. 2014, p. 16 -31.
- [C9] Adrien COULET, Florent DOMENACH, Mehdi KAYTOUE et Amedeo NAPOLI. “Using Pattern Structures for Analyzing Ontology-Based Annotations of Biomedical Data”. In : *ICFCA 2013 - International Conference on Formal Concept Analysis*. LNCS/LNAI series. Dresden, Germany : Springer, mai 2013.
- [C10] Mehwish ALAM, Adrien COULET, Amedeo NAPOLI et Malika SMAÏL-TABBONE. “Formal Concept Analysis Applied to Transcriptomic Data”. In : *CLA 2012 - The Ninth International Conference on Concept Lattices and Their Applications*. Fuengirola (Málaga), Spain, oct. 2012.
- [C11] Adrien COULET, Malika SMAÏL-TABBONE, Amedeo NAPOLI et Marie-Dominique DEVIGNES. “Role Assertion Analysis : a proposed method for ontology refinement through assertion learning”. In : *STAIRS 2008 - Proceedings of the Fourth Starting AI Researchers’ Symposium*. T. 179. Frontiers in Artificial Intelligence and Applications. IOS Press, 2008, p. 47-58.
- [C12] Adrien COULET, Malika SMAÏL-TABBONE, Pascale BENLIAN, Amedeo NAPOLI et Marie-Dominique DEVIGNES. “SNP-Converter : an Ontology-Based solution to Reconcile Heterogeneous SNP Descriptions for Pharmacogenomic Studies”. In : *DILS’06 - Data Integration in the Life Sciences*. LNCS series. European Bioinformatics Institute (EBI), Hinxton, UK : Springer, juil. 2006.

Ateliers internationaux avec comité de lecture

- [A1] Pierre MONNIN, Chedy RAÏSSI, Amedeo NAPOLI et Adrien COULET. “Knowledge Reconciliation with Graph Convolutional Networks : Preliminary Results”. In : *DL4KG2019 - Workshop on Deep Learning for Knowledge Graphs*. T. CEUR Workshop Proceedings. Co-located with ESWC 2019 2377. Portoroz, Slovenia, juin 2019.
- [A2] Joël LEGRAND, Yannick TOUSSAINT, Chedy RAÏSSI et Adrien COULET. “Syntax-based Transfer Learning for the Task of Biomedical Relation Extraction”. In : *LOUHI 2018 - The Ninth International Workshop on Health Text Mining and Information Analysis*. Proceedings of LOUHI 2018 : The Ninth International Workshop on Health Text Mining and Information Analysis. Brussels, Belgium, oct. 2018.
- [A3] Pierre MONNIN, Amedeo NAPOLI et Adrien COULET. “Combining Concept Annotation and Pattern Structures for Guiding Ontology Mapping”. In : *FCA4AI 2018 - 6th International Workshop “What can FCA do for Artificial Intelligence ?”* T. CEUR Workshop Proceedings. Co-located with IJCAI/ECAI 2018 2149. Stockholm, Sweden, 2018.
- [A4] Joël LEGRAND, Yannick TOUSSAINT, Chedy RAÏSSI et Adrien COULET. *Tree-LSTM and Cross-Corpus Training for Extracting Biomedical Relationships from Text*. DLPM2017 - 2nd International Workshop on Deep Learning for Precision Medicine. 2017.
- [A5] Pierre MONNIN, Clement JONQUET, Joël LEGRAND, Amedeo NAPOLI et Adrien COULET. “PGxO : A very lite ontology to reconcile pharmacogenomic knowledge units”. In : *NETTAB 2017 - Methods, tools & platforms for Personalized Medicine in the Big Data Era*. NETTAB 2017 Workshop Collection. Palermo, Italy, 2017.
- [A6] Gabin PERSONENI, Marie-Dominique DEVIGNES, Michel DUMONTIER, Malika SMAÏL-TABBONE et Adrien COULET. “Discovering ADE associations from EHRs using pattern structures and ontologies”. In : *Phenotype Day, Bio-Ontologies SIG*. Co-located with ISMB 2016. Orlando, United States, 2016.

- [A7] Mohsen HASSAN, Olfa MAKKAOUI, Adrien COULET et Yannick TOUSSAINT. “Extracting Disease-Symptom Relationships by Learning Syntactic Patterns from Dependency Graphs”. In : *BioNLP 2015 - Workshop on Biomedical Natural Language Processing*. Co-located with IJCNLP 2015. Beijing, China : Association for Computational Linguistics, juil. 2015, p. 71-80.
- [A8] Mohsen HASSAN, Adrien COULET et Yannick TOUSSAINT. “Learning Subgraph Patterns from text for Extracting Disease - Symptom Relationships”. In : *DMNLP’14 - 1st International Workshop on Interactions between Data Mining and Natural Language Processing*. Co-located with ECML/PKDD 2014. Nancy, France, 2014, p. 81-96.
- [A9] Mehwish ALAM, Melisachew Wudage CHEKOL, Adrien COULET, Amedeo NAPOLI et Malika SMAÏL-TABBONE. “Lattice Based Data Access (LBDA) : An Approach for Organizing and Accessing Linked Open Data in Biology”. In : *DMoLD’13 - Data Mining on Linked Data Workshop*. Co-located with ECML/PKDD 2013. Prague, Czech Republic, 2013.
- [A10] Mehwish ALAM, Adrien COULET, Amedeo NAPOLI et Malika SMAÏL-TABBONE. “Formal Concept Analysis Applied to Transcriptomic Data”. In : *FCA4AI 2012 - International Workshop “What can FCA do for Artificial Intelligence ?”* Co-located with ECAI 2012. Montpellier, France, 2012, p. 7-14.
- [A11] Adrien COULET, Russ B. ALTMAN, Mark A. MUSEN et Nigam SHAH. “Integrating heterogeneous relationships extracted from natural language sentences”. In : *Bio-ontologies SIG*. Co-located with ISMB 2010. 2010.
- [A12] Adrien COULET, Malika SMAÏL-TABBONE, Amedeo NAPOLI et Marie-Dominique DEVIGNES. “Ontology Refinement through Role Assertion Analysis : Example in Pharmacogenomics”. In : *Description Logics*. T. 353. CEUR Workshop Proceedings. CEUR-WS.org, 2008.
- [A13] Adrien COULET, Malika SMAÏL-TABBONE, Amedeo NAPOLI et Marie-Dominique DEVIGNES. “Suggested Ontology for Pharmacogenomics (SO-Pharm) : Modular Construction and Preliminary Testing”. In : *OTM Workshops (1)*. T. 4277. LNCS series. Springer, 2006, p. 648-657.

Conférence nationale avec comité de lecture

- [c1] Walid HAFIANE, Joël LEGRAND, Yannick TOUSSAINT et Adrien COULET. “Expérimentations autour des architectures d’apprentissage par transfert pour l’extraction de relations biomédicales”. In : *Extraction et Gestion des Connaissances (EGC)*. Montpellier, France, jan. 2021.
- [c2] Pierre MONNIN, Amedeo NAPOLI et Adrien COULET. “Discovering Subsumption Axioms with Concept Annotation”. In : *BDA 2017 - Gestion de Données - Principes, Technologies et Applications*. Poster. Nov. 2017.
- [c3] Adrien COULET, Florent DOMENACH, Mehdi KAYTOUE et Amedeo NAPOLI. “Using pattern structures for analyzing ontology-based annotations of biomedical data”. In : *JIAF 2013 - Septièmes Journées d’Intelligence Artificielle Fondamentale*. Aix-en-Provence, France, juin 2013, p. 97-106.
- [c4] Clement JONQUET, Adrien COULET, Nigam H. SHAH et Mark A. MUSEN. “Indexation et intégration de ressources textuelles à l’aide d’ontologies : application au domaine biomédical”. In : *IC’10 - 21èmes Journées Francophones d’Ingénierie des Connaissances*. 2010, p. 271-282.

Ateliers nationaux avec comité de lecture

- [a1] Gabin PERSONENI, Emmanuel BRESSO, Marie-Dominique DEVIGNES, Michel DUMONTIER, Malika SMAÏL-TABBONE et Adrien COULET. “Découverte d’associations entre Evénements Indésirables Médicamenteux par les structures de patrons et les ontologies”. In : *Journée I.A. et Santé*. Nancy, France, juil. 2018.
- [a2] Gabin PERSONENI, Marie-Dominique DEVIGNES, Michel DUMONTIER, Malika SMAÏL-TABBONE et Adrien COULET. “Extraction d’association d’EIM à partir de dossiers patients : expérimentation avec les structures de patrons et les ontologies”. In : *Deuxième Atelier sur l’Intelligence Artificielle et la Santé*. Atelier IA & Santé. Montpellier, France, juin 2016.

- [a3] Adrien COULET, Marie-Dominique DEVIGNES et Malika SMAÏL-TABBONE. “Extraction de connaissances pharmacogénomiques à partir d’études cliniques : problématique”. In : *Deuxième atelier sur la fouille de données complexes, en conjonction avec EGC’05*. 2005.

Editoriaux invités

- [E1] Adrien COULET, K. Bretonnel COHEN et Russ B. ALTMAN. “The state of the art in text mining and natural language processing for pharmacogenomics”. In : *Journal of Biomedical Informatics* 45.5 (2012), p. 825-826.
- [E2] Adrien COULET, Nigam H. SHAH, Lawrence HUNTER, Chitta BARAL et Russ B. ALTMAN. “Extraction of Genotype-Phenotype-Drug Relationships from Text : From Entity Recognition to Bioinformatics Application”. In : *Pacific Symposium on Biocomputing*. World Scientific Publishing, 2010, p. 485-487.

Articles en pré-impression

- [P1] Emmanuel BRESSO, Pierre MONNIN, Cédric BOUSQUET, François-Elie CALVIER, Ndeye-Coumba NDIAYE, Nadine PETITPAIN, Malika SMAÏL-TABBONE et Adrien COULET. *Investigating ADR mechanisms with knowledge graph mining and explainable AI*. 2020.
- [P2] Pierre MONNIN, Chedy RAÏSSI, Amedeo NAPOLI et Adrien COULET. *Rediscovering alignment relations with Graph Convolutional Networks*. 2020. arXiv : 2011.06023 [cs.LG].

Sélection de cinq publications

[R4] Predicting the need for a reduced drug dose, at first prescription dans *Scientific Reports*

Cet article décrit un modèle capable de prédire, avant la prescription d'un médicament, qu'un patient pourrait en nécessiter une dose réduite, pour éviter les effets indésirables. Ce modèle est entraîné à partir de données classiquement collectées dans les dossiers patients hospitaliers comme les diagnostics et les examens biologiques. Cette contribution démontre pour la première fois dans l'état de l'art qu'il est possible de développer des outils d'aide à la prescription à partir de données collectées en routine à l'hôpital pour éviter les effets indésirables liés au surdosage. Cet article constitue un résultat important de ma collaboration avec l'Université Stanford. J'ai piloté la conception de l'expérience ; j'ai implémenté et réalisé les expérimentations ; et ai piloté la rédaction de l'article.

[R1] PGxCorpus, a manually annotated corpus for pharmacogenomics dans *Scientific Data*

Cet article décrit un corpus de textes manuellement annoté, sa création et la démonstration de son intérêt pour l'extraction automatique de relations *gène-médicament-réponse au médicament* à partir de textes. Il s'agit d'un résultat important du projet ANR PractiKPharma (décrit plus loin), notamment parce qu'une telle ressource manquait au domaine de la pharmacogénomique qui étudie ces relations. Il permettra de progresser d'une part dans la gestion des connaissances associées à ce domaine, et d'autre part à évaluer des outils divers de traitement de la langue. Nous avons démontré dans un travail récent que le modèle BERT, réglé finement à l'aide de PGxCorpus permet de dépasser les meilleures performances de l'état de l'art pour la tâche visée [c1]. J'ai piloté la conception de la méthode de construction et d'évaluation du corpus, le déroulement de la campagne d'annotation par les experts et la rédaction de l'article.

[R2] PGxO and PGxLOD: a reconciliation of pharmacogenomic knowledge of various provenances, enabling further comparison dans *BMC Bioinformatics*

Cet article décrit la constitution et l'utilisation d'un graphe de connaissances pour le domaine de la pharmacogénomique. La particularité de ce graphe est de regrouper des connaissances d'origines différentes : littérature, bases de données expertes, analyses de dossiers patients ; et de les réconcilier. C'est une ressource d'intérêt pour ce domaine car elle permet d'évaluer la complémentarité entre sources ; et au delà, pour l'expérimentation sur les graphes de connaissances et leur réconciliation. C'est une contribution originale en premier lieu car elle propose une méthode pour comparer les connaissances d'un même domaine, mais de provenances diverses. Il s'agit d'un résultat conséquent du projet ANR PractiKPharma. J'ai dirigé la conception de la ressource, de la méthode de comparaison de connaissances et la rédaction de l'article.

[A1] Knowledge Reconciliation with Graph Convolutional Networks: Preliminary Results dans l'atelier *DLAKG: Deep Learning for Knowledge Graphs*

Cet article d'atelier est important et original de mon point de vue car il évalue la capacité d'une méthode d'apprentissage supervisé, un réseau de neurones convolutif, à comparer les connaissances représentées dans un graphe de connaissances. L'intérêt particulier de ce travail est d'utiliser une méthode d'apprentissage profond à partir de graphe pour la réalisation d'une tâche de comparaison de connaissances. Les résultats prometteurs de ce premier article nous ont amené à prolonger ce travail et le soumettre sous la forme d'un article long à un journal international du domaine [P2]. C'est un premier résultat prometteur autour duquel nous continuons de travailler dans le cadre de la thèse de Pierre Monnin. J'ai conçu la méthode avec Pierre Monnin, et dirigé l'interprétation des résultats et la rédaction.

[R5] Learning from biomedical linked data to suggest valid pharmacogenes dans *Journal of Biomedical Semantics*

Cet article porte également sur les graphes de connaissances puisqu'il s'agit d'un travail de prédiction de liens au sein d'un tel graphe. Nous nous sommes intéressés ici à la prédiction de liens *gène-médicament* de sorte à identifier quel gène est le plus susceptible de jouer un rôle dans la variabilité des réponses aux médicaments. Il s'agit d'un travail important car il a été réalisé à partir de données réelles et a donné lieu à une évaluation et une interprétation par des experts du domaine. J'ai conçu l'expérience et dirigé le travail de préparation des données, d'implémentation et d'interprétation des résultats et de rédaction.

3.2 Autres productions scientifiques

Dans le cadre du projet ANR PractiKPharma, j'ai conduit la constitution de deux ressources d'intérêt : **PGxLOD**, un **graphe de connaissances** pour la pharmacogénomique (domaine qui étudie l'impact des facteurs génétiques sur la réponse aux médicaments) et **PGxCorpus**, un **corpus de textes manuellement annotés** avec les entités et les relations d'intérêt pour ce même domaine biomédical. Ces deux ressources et le code associé sont librement disponibles. Ces deux ressources sont intéressantes pour le domaine de la pharmacogénomique, mais aussi pour l'expérimentation de méthodes informatiques d'analyse de graphes de connaissances (comme la prédiction de liens, la comparaison de connaissances) et d'évaluation d'outils de traitement automatique du langage (comme l'extraction de relations binaires, ternaires, la reconnaissance d'entités discontinues ou imbriquées).

J'ai aussi développé **4 ontologies** dédiées à des domaines et applications biomédicaux : SNP-Ontology, SO-Pharm, PHARE, PGxO. Toutes sont ouvertes et disponibles sur le site web du Bioportal.

3.3 Responsabilités scientifiques

Principaux projets de recherche

Comme porteur

AMI Santé Numérique 2020 Inserm-Inria (projet sélectionné)

Notre proposition de création d'équipe commune Inria-Inserm en réponse à cet appel a été retenue. J'ai co-porté et co-écrit cette réponse avec Sarah Zohar (DR Inserm). L'équipe proposée s'appelle HeKA (*Health data- and model- driven Knowledge Acquisition*) est poursuit actuellement son processus de création.

LUE Future Leader (2019-20, 30k€)

J'ai reçu cette distinction du projet I-SITE *Lorraine Université d'Excellence* pour préparer des projets de recherche. L'Université de Lorraine m'a à ce titre attribué un allègement de service de 64h.

AI in Medicine Seed Grant (2018–19, \$30k)

L'objectif de ce projet joint avec Nigam H. Shah et Stefan Pfohl et financé par l'Université Stanford était l'étude de méthodes d'apprentissage automatique pour le développement de modèles prédictifs équitables en médecine.

LUE Widen Horizon (2017–19, 2 x 10k€)

J'ai obtenu l'aide de l'I-SITE Lorraine Université d'Excellence pour mener un projet de mobilité géographique à Stanford de 2 ans.

ANR PractiKPharma (2016–20, 675k€)

Je suis le coordinateur de ce projet dont l'objectif est l'extraction de connaissances biomédicales à partir de bases de données, puis leur vérification par la fouille de données observationnelles dans un entrepôt de dossiers patients électroniques.

Équipes Associées Inria Snowflake, puis Snowball (2014–19, 62k€)

Cette équipe associée Inria implique les Orpailleurs du Loria et l'équipe de Nigam H. Shah de Stanford. Son objectif est de faciliter nos collaborations autour de la découverte de connaissances à partir de données de dossiers de patients électroniques.

PEPS Mirabelle EXPLOD-BioMed (2013, 11k€)

Ce projet d'un an, en collaboration avec l'équipe de génétique humaine du CHU de Nancy, portait sur la fouille de données ouvertes et liées pour la recherche de gènes responsables des déficiences intellectuelles. Il a été poursuivi par la thèse de G. Personeni.

Comme collaborateur

LOR-AI (2020-23, 600k€)

Ce projet a été retenu par l'ANR pour co-financer 12 thèses en intelligence artificielle à l'Université de Lorraine. Ma proposition de sujet de thèse, en collaboration avec le CHRU de Nancy a été retenue dans le cadre de cet appel.

ASSURE (2017-18, \$30k)

Ce projet financé par Stanford regroupait épidémiologues, cardiologues et informaticiens autour de l'étude de méthodes d'apprentissage sur des données de patients pour l'évaluation du risque de survenue d'accidents cardiovasculaires.

ANR Hybride (2013–16, 492k€)

J'ai participé au projet ANR Hybride sur une tâche d'extraction de relations maladie–symptôme à partir de texte.

Organisation d'événements

J'ai participé à l'organisation et l'animation du *First Inria-DFKI workshop on Artificial Intelligence ID-AI*, environ 80 participants, 20-21 janvier 2020.

J'ai organisé le mini *France-Stanford workshop on EHR mining*, 10 participants 12-13 juillet 2015.

J'ai participé à l'organisation de la conférence internationale *ECCB'14 (European Conference on Computational Biology)*, environ 1000 participants, 7-10 septembre 2014.

J'ai organisé le *GPD-Rxn workshop : Genotype-Phenotype-Drug Relationships Extraction from Text*, en conjonction avec la conférence *PSB 2010*, environ 50 participants, 7 janvier 2010.

3.4 Encadrement doctoral et scientifique

Postdoctorat

J'ai encadré Joël Legrand durant ses deux ans de contrat postdoctoral 2016-18, financé par un projet ANR. Joël a travaillé sur le traitement naturel du langage et plus particulièrement sur la construction d'un corpus et l'apprentissage par transfert pour l'extraction de relations biomédicales à partir de textes. Ses publications sont [R1, R2, A2, A4]. Joël est maintenant Maître de Conférences à CentraleSupélec Metz. Nous continuons à collaborer et à publier ensemble [c1].

Doctorats

J'ai le plaisir d'avoir co-encadré 3 doctorants en informatique et de commencer la direction de 3 nouvelles thèses :

- Alice Rogier sur le sujet "Détection and prédiction des toxicités médicamenteuses à partir de données de dossiers patients", financée par un contrat doctoral Inserm-Inria. Co-dirigé avec Bastien Rance.
- Adrien Blanche sur la "Prédiction de la réponse aux traitements cardiovasculaires à partir de données cliniques", co-financé par l'ANR ("contrats doctoraux en IA") et le CHRU de Nancy. Co-dirigé avec Marie-Dominique Devignes et Nicolas Girerd.
- Thibaut Fabacher sur le "Phénotypage profond pour la constitution de cohortes de patients similaires au mien", financé par le CHU de Strasbourg. Co-dirigé avec Erik Sauleau.
- Pierre Monnin, financé par le projet ANR PractiKPharma, dont la thèse *soutenue en 2020* s'intitule "Alignement et fouille de graphes de connaissances du Web des données. Applications en pharmacogénomique". Ses publications sont [C2, C1, R2, C5, A1, A3, A5, c2]. Co-encadré avec Amedeo Napoli. Pierre est maintenant chercheur à Orange Labs, Belfort. Nous continuons à collaborer.
- Gabin Personeni, financé par contrat doctoral, dont la thèse *soutenue en 2018* s'intitule "Apport des ontologies de domaine pour l'extraction de connaissances à partir de données biomédicales". Ses publications sur la période évaluée sont [R6, C4, C7, C8, A6, a1]. Co-encadré avec Marie-Dominique Devignes. Gabin est maintenant ingénieur au CHRU de Nancy.
- Mohsen Hassan, financé par le projet ANR Hybride, *soutenue en 2017* et intitulée "Knowledge Discovery Considering Domain Literature and Ontologies : Application to Rare Diseases". Ses publications sont [A7, A8]. Co-encadré avec Yannick Toussaint. Mohsen est maintenant ingénieur dans l'industrie.

J'ai également dirigé (50% d'encadrement) la thèse d'exercice de pharmacie de Kévin Dalleau, *soutenue en 2017* et distinguée par le prix de la thèse d'exercice à caractère expérimental de la Faculté de Pharmacie de Nancy. Elle s'intitule "Suggestion d'interactions gène-médicament à partir de données ouvertes et liées". Le travail de Kévin a mené à deux publications : la première en 2015 et la seconde en 2016 [R5, C6]. Kévin est actuellement en thèse d'informatique dans l'équipe Orpailleur.

Sans être leur co-directeur, j'ai participé au travail de thèse de : Stephen Pfohl (*en cours* à Stanford), Mehwish Alam (*soutenue en 2014* à Nancy) et Yael Garten (*soutenue en 2011* à Stanford).

Masters

Depuis 2016, j'ai co-encadré 3 étudiants en M2 Informatique (encadrement à 33%) :

- Walid Hafiane en 2020 dont le mémoire s'intitule "apprentissage par transfert pour l'extraction de relations pharmacogénomiques à partir de textes". Son travail de Master sera présenté à la conférence EGC 2021 [c1].
- Gaurav Shajepal en 2019 dont le mémoire s'intitule "Caractérisation linguistique de corpus dans une expérience d'apprentissage",
- Ilias Benjelloun en 2017 dont le mémoire s'intitule "Prises en compte de l'incertitude et des connaissances de domaine dans l'apprentissage à partir de données relationnelles",

3.5 Diffusion des travaux (rayonnement et vulgarisation)

Jurys et suivis de thèses

J'ai rapporté 2 thèses en informatique en 2020, participé au jury de 3 autres en tant que co-encadrant et à un jury de thèse en pharmacie en tant que directeur.

J'ai participé au comité de suivi individuel d'une thèse de bioinformatique (Jonathan Mercier, Orsay) et participe actuellement à 3 autres (Nicholas Halliwell à Sophia-Antipolis, Erwan Schild et Chuyuan Li à Nancy).

Invitations

Séminaires invités J'ai été invité à présenter mes travaux lors des événements suivants :

Découverte de connaissances à partir de données complexes. Applications à la fouille de données en pharmacogénomique. *Colloque Sciences Ouvertes 2020 de l'Université de Lorraine*, avec Amedeo Napoli, Nancy (2020).

Predicting Dose Changes *EIT Health Summit*, Paris (2019).

Knowledge extraction and comparison for pharmacogenomics. *Séminaire du Lirmm*, Montpellier (2019).

Predicting patient sensitivity to pharmacogenomic drugs, using EHR data. *6^{ème} workshop Inria@SiliconValley*, Paris (2016).

Comment et pourquoi les Big Data sont entrées à TELECOM Nancy? *Conférences Pasc@Line*, avec Olivier Festor, Télécom Paris Tech (2015).

Extraction de relations pharmacogénomiques à partir de la littérature. *Journée Fouille de texte pour la biologie*, organisée par la plateforme de Bioinformatique de Lille, Villeneuve d'Ascq (2011).

Ontology construction in pharmacogenomics. *9^{ème} Journée de la plate-forme Bio-informatique Genouest*, Rennes (2011).

Extraction and integration of heterogeneous PGx relationships. *Seminar of the Computational Biology Research Group, Max Plank Institute for Informatics*, Saarbrücken (2010).

Knowledge Discovery guided by Domain Knowledge in Pharmacogenomics. *Colloquium of the Stanford Research Center for Biomedical Informatics*, Stanford University (2008).

Découverte de connaissances pharmacogénomiques guidée par les connaissances du domaine. *Réunion annuelle du département Alimentation Humaine (AlimH) de l'INRA*, Arèches-Beaufort (2008).

Identification de gènes spécifiques aux îlots de Langerhans par clustering d'EST. *Séminaire du groupe de recherche lorrain en bioinformatique*, Nancy (2004).

Autres invitations

Chercheur invité à l'Université Stanford (2017-19) : J'ai été invité pendant deux ans dans l'équipe de Nigam H. Shah, au *Center for Biomedical Informatics Research* de l'Université Stanford.

Biohackathon 2018 Paris : J'ai été invité à diriger une tâche durant ce hackathon de 5 jours, organisé par Elixir et l'Institut Français de Bioinformatique.

Distinctions

Meilleure application du Hackathon 2014 du NCBO : L'application que j'ai prototypé pendant cet événement a remporté le prix de la meilleure application.

Formation innovante : la formation en science des données dont j'ai piloté la création en 2014 à TELECOM Nancy a été distinguée en 2015 par l'association des écoles d'ingénieurs du numérique Pasc@line, comme formation innovante.

IMIA Yearbook of biomedical informatics 2011 : Notre article intitulé "Using text to build semantic networks for pharmacogenomics" et publié dans le *Journal of Biomedical Informatics* a été sélectionné pour cette compilation annuelle des meilleurs articles du domaine.

Semantic Web Challenge 2010 : Avec les collègues du National Center for Biomedical Ontologies (NCBO), nous avons remporté ce challenge lors de la conférence internationale du web sémantique, ISWC 2010.

Activités communautaires

Relecteur régulier pour les journaux suivants : *Journal of Biomedical Informatics*, *Journal of Biomedical Semantics*, *Journal of the American Medical Informatics Association (JAMIA)* et occasionnellement pour *BMC Bioinformatics*, *BMC Medical Informatics*, *BMC Genomics*, *International Journal of Medical Informatics*.

Membre du comité de programme des conférences et ateliers suivants (pour les 4 dernières années) : *OHDSI Symposium 2020*, *SIMBig 2020*, *IJCAI 2019-20*, *ECML-PKDD 2019*, *SWAT4HCLS (Semantic Web Applications and Tools for Healthcare and Life Sciences) 2018-19*, *Bio-ontologies ISMB SIG (Special Interest Group) 2016-17*, *IA & Santé 2016-20*, *SIIM (Symposium sur l'Ingénierie de l'Information Médicale) 2016-17*, *ECCB (European Conference on Computational Biology) 2016* et relecteur pour les conférences *ISMB (Intelligent Systems for Molecular Biology) 2017*, *PSB (Pacific Symposium on Biocomputing) 2016*.

Expert occasionnel pour les agences de financement (pour l'ANR, pour le DFG allemand, etc.) et comités de sélection.

Vulgarisation

Médiation scientifique En 2019-20, j'ai participé à sept manifestations de médiation avec des publics de jeunes étudiants, d'élèves de classes préparatoires, de collégiens (la fête de la science), de médecins (Forum IA & Santé du CHRU de Nancy) ou d'industriels (Grand Est Numérique, Forum entreprise de la Fédération Charles Hermite).

Articles de vulgarisation J'ai participé à l'écriture des deux articles suivants

[V1] Adrien COULET et Nicolas JAY. "Contre les effets indésirables, un algorithme pour personnaliser les doses de médicaments". In : *The Conversation FR* (2019).

[V2] Adrien COULET et Malika SMAÏL-TABBONE. "Mining Electronic Health Records to Validate Knowledge in Pharmacogenomics". In : *ERCIM News*. ERCIM News 104, Special theme : Tackling Big Data in the Life Sciences 104 (2016), p. 56.