

Adrien Coulet

January 28, 2020

Contents

1	Curriculum vitæ	2
2	Publications	7
3	Main results of the five past years	11
3.1	Predicting the need to reduce drug dose, before prescription	11
3.2	The PractiKPharma project, funded by the ANR	11
3.3	The supervision of a pharmacy thesis, but in computer science!	11
3.4	The creation of a major in Data Science at TELECOM Nancy	11

1 Curriculum vitæ

Generalities

First and last names: Adrien Coulet
Date and place of birth: 14/11/1979 in Montpellier, France
Nationality: French

Family status: married, three children

Email: adrien.coulet@loria.fr
Telephone: +33 3 54 95 86 38
Postal address: LORIA – Inria Nancy-Grand Est
Équipe Orpailleur – Bâtiment B
Campus Scientifique, BP 239
54506 Vandoeuvre-lès-Nancy cedex, France

Current position

Since 2010 Associate professor (namely *Maître de conférences*) in Computer Science at the University of Lorraine, Nancy, France

Researcher in the Orpailleur group, of the **Loria** lab (University of Lorraine, CNRS, Inria) and teaching at **TELECOM Nancy**, a graduate school of engineering in computer science, at the University of Lorraine.

I received the *Habilitation à diriger des recherches* (HDR) from the University of Lorraine on Dec. 16, 2019.

Before this

Dec. 2019 HDR in Computer Science from the *University of Lorraine*, Nancy, France mentored by Amedeo Napoli

reviewed by Sarah Cohen-Boulakia, Nathalie Aussenac-Gilles and Olivier Curé, examined by Olivier Dameron and Marie-Dominique Devignes

2017-19 Visiting faculty at Stanford University, California

in *délégation* in the group of Nigam H. Shah, at the **Stanford Center for Biomedical Informatics Research**.

2008-10 Post-doctoral fellow Stanford University, California

on *relation extraction from text and automatic annotation of biomedical databases*, under the co-supervision of Russ B. Altman and Mark A. Musen

2004-08 PhD in Computer Science from the *Université Henri Poincaré*, Nancy, France

untitled *Construction and use of a pharmacogenomic knowledge base for data integration and knowledge discovery*

under the co-supervision of Marie-Dominique Devignes and Malika Smaïl-Tabbone

reviewed by Alain Viari and Mohand-Saïd Hacid, examined by Chantal Raynaud and Nacer Boudjlida

2003-04 Master Healthcare engineering and drug design, *Université Joseph Fourier*, Grenoble, France major on “Spatio-temporal approaches for life sciences”

Internship on EST clustering under the co-supervision of Michel Sève and Jacques Demongeot

2000-03 Engineering diploma (equiv. MSc) in computer science, graduate school of engineering Polytech'Grenoble, *Université Joseph Fourier*, Grenoble, France

major on "IT for healthcare"

Two noteworthy internships: the 1st at Inria Rhône-Alpes on the conception of a proteomic database with Alain Viari and François Rechenmann, the 2nd at the School of Pharmacy of Montpellier on molecular screening for the discovery of novel antibiotics.

2000 DEUG in life science (equiv. BSc), *Université Montpellier 2*, Montpellier, France

PhD supervision

I have had the pleasure of co-supervising 3 PhD students in computer science:
Pierre Monnin, funded on an ANR project, *ongoing*;
Gabin Personeni, funded by doctoral contract, *defended in 2018*;
Mohsen Hassan, funded on an ANR project, *defended in 2017*.

I have also supervised the pharmacy thesis of Kevin Dalleau, *defended in 2017*, awarded by the annual price of the experimental thesis of the School of Pharmacy of Nancy.

Even though I was not officially their supervisor, I had the chance to work with the following PhD students: Stephen Pfohl (*ongoing* at Stanford), Mehwish Alam (*defended in 2014* in Nancy) and Yael Garten (*defended in 2011* at Stanford).

Main research projects

As a Principal Investigator

LUE Future Leader ERC Track (2019-20, 30k€)

I have been granted by the I-SITE project *Lorraine Université d'Excellence* for maturing an ERC research project.

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

With Nigam H. Shah and Stefan Pfohl, this one-year project is funded by Stanford University. Its aim is the study of machine learning methods for developing “fair” predictive model in medicine.

LUE Widen Horizon (2017–19, 20k€)

I have been granted by the I-SITE project *Lorraine Université d'Excellence* for my project of a two-year sabbatical at Stanford.

ANR PractiKPharma (2016–19, 675k€)

I am the P.I. of this 4-year project, which aims at extracting biomedical knowledge from domain databases and literature, then checking their validity by mining observational data in Electronic Health Records (EHR) warehouses.

Snowflake/Snowball Stanford–Inria Associate Team (2014–19, 62k€)

This Inria Associate Team involves the *Orpailleur* group of the Loria and the group of Nigam H. Shah at Stanford. It aims at fostering collaborations on knowledge discovery from EHR.

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

This one year project, in collaboration with the group of human genetics of the Nancy University Hospital, focused on mining linked open data for genes responsible of intellectual disabilities. It initiated the PhD project of G. Personeni.

As an investigator

ASSURE (2017-18, \$30k)

This Stanford-funded project was grouping epidemiologists, cardiologists and computer scientist on the use of machine learning methods to evaluate the risk of cardiovascular events.

ANR Hybride (2013–16, 492k€)

I was involved in the Hybride project on a task of extraction from text of disease-symptom relationships.

Teaching and responsibilities

I am or was teaching:

- Semantic Web and logic programming*, 30h, to 3rd year at TELECOM Nancy and M2 in Computer Science,
- Large data set management*, 40h, to 2nd year at TELECOM Nancy and M2 in Data Science
- Initiation to AI, 30h, to 2nd year at TELECOM Nancy,

- Databases*, 35h, to 1st year at TELECOM Nancy.
 - Conception methods, 20h, to 2nd year, at TELECOM Nancy
 - Programming project, 20h, to 2nd year, at TELECOM Nancy
- Teaching units marked with a * are those I am or was responsible for.

From 2011 to 2014, I have been in charge of the coordination of the major on Information Systems at TELECOM Nancy. In 2014, I conducted the creation of a major in Data Science at TELECOM Nancy, which I coordinated from 2014 to 2017.

Other distinctions

Biohackathon 2018 Paris: I have been invited to lead a task during this 5-days event, organised by Elixir and the French Institut of Bioinformatics.

Innovative Teaching Program: The major in Data Science I created in 2014 at TELECOM Nancy has been awarded by Pasc@line, the association of graduate school of computer science as an Innovative Teaching Program.

NCBO Hackathon 2014: The application I prototyped during this event get awarded with the price of the best application.

IMIA Yearbook of biomedical informatics 2011: Our article untitled ‘Using text to build semantic networks for pharmacogenomics’ and published in the *Journal of Biomedical Informatics* has been selected in this annual selection.

Semantic Web Challenge 2010: With colleagues of the National Center for Biomedical Ontologies (NCBO), we won this challenge during the International Semantic Web Conference, ISWC 2010.

Invited talks

Knowledge extraction and comparison for pharmacogenomics. *Lirimm Seminar*, 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*, 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).

Community service

Organisation of the conference *ECCB’14* (European Conference on Computational Biology), about 1,000 participants.

Reviewing and program committees

regularly for the following journals: *Journal of Biomedical Informatics*, *Journal of Biomedical Semantics*, *Journal of the American Medical Informatics Association (JAMIA)* and occasionally for *BMC Bioinformatics*, *BMC Medical Informatics*, *BMC Genomics*, *Technique et Science Informatiques*.

Member of the program committee of following conferences and workshops : *IJCAI 2019*, *ECML-PKDD 2019*, *SWAT4HCLS (Semantic Web Applications and Tools for Healthcare and Life Sciences) 2018*, *Bio-ontologies ISMB SIG (Special Interest Group) 2010-17*, *IA & Santé 2016-19*, *SIIM (Symposium sur l'Ingénierie de l'Information Médicale) 2015-17*, *ECCB (European Conference on Computational Biology) 2016*, *CSHALS (Conference on Semantics in Healthcare and Life Sciences) 2014*, *DILS (Data Integration in Life Sciences) 2013*, *AIMM (Annotation, Interpretation and Management of Mutations) 2012* et relecteur pour les conférences *ISMB (Intelligent Systems for Molecular Biology) 2014 & 17*, *PSB (Pacific Symposium on Biocomputing) 2013 & 16*, *SNP-SIG Meeting 2011*.

Occasional expert for funding agencies (French National Research Agency ANR, the German Research Agency DFG, etc.)

Editorials

I have been invited to write two editorials, the first for a special issue of the *Journal of Biomedical Informatics* on text mining for pharmacogenomics [E1] and the second for presenting and summarising works presented at the PSB 2010 *GPD-Rxn* workshop I organised [E2].

2 Publications

International journals

- [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, and Adrien Coulet. “PGxCorpus, a manually annotated corpus for pharmacogenomics”. In: *Scientific Data* 7.1 (2020), pp. 1–13.
- [R2] Pierre Monnin, Joël Legrand, Graziella Husson, Patrice Ringot, Andon Tchechmedjiev, Clément Jonquet, Amedeo Napoli, and 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, and 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, and Michel Dumontier. “Predicting the need for a reduced drug dose, at first prescription”. In: *Scientific Reports* 8.1 (2018).
- [R5] Kevin Dalleau, Yassine Marzougui, Sébastien Da Silva, Patrice Ringot, Ndeye Coumba Ndiaye, and Adrien Coulet. “Learning from biomedical linked data to suggest valid pharmacogenes”. In: *Journal of Biomedical Semantics* 8.29 (2017).
- [R6] Gabin Personeni, Emmanuel Bresso, Marie-Dominique Devignes, Michel Dumontier, Malika Smaïl-Tabbone, and Adrien Coulet. “Discovering associations between adverse drug events using pattern structures and ontologies”. In: *Journal of Biomedical Semantics* 8.16 (2017).
- [R7] Yi Liu, Adrien Coulet, Paea LePendou, and Nigam H. Shah. “Using ontology-based annotation to profile disease research”. In: *Journal of the American Medical Informatics Association* 19.e1 (2012).
- [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, and M Scott Marshall. “Semantically enabling pharmacogenomic data for the realization of personalized medicine”. In: *Pharmacogenomics* 13.2 (2012), pp. 201–212.
- [R9] Adrien Coulet, Yael Garten, Michel Dumontier, Russ B Altman, Mark A Musen, and 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, and Marie-Dominique Devignes. “Ontology-based knowledge discovery in pharmacogenomics”. In: *Advances in Experimental Medicine and Biology, Springer* 696 (2011), pp. 357–66.
- [R11] Clement Jonquet, Paea LePendou, Sean M. Falconer, Adrien Coulet, Natalya Fridman Noy, Mark A. Musen, and Nigam H. Shah. “NCBO Resource Index: Ontology-based search and mining of biomedical resources”. In: *Journal of Web Semantics* 9.3 (2011), pp. 316–324.
- [R12] Adrien Coulet, Nigam H. Shah, Yael Garten, Mark A. Musen, and Russ B. Altman. “Using text to build semantic networks for pharmacogenomics”. In: *Journal of Biomedical Informatics* 43.6 (2010), pp. 1009–1019.
- [R13] Yael Garten, Adrien Coulet, and Russ B Altman. “Recent progress in automatically extracting information from the pharmacogenomic literature”. In: *Pharmacogenomics* 11.10 (2010), pp. 1467–1489.
- [R14] Adrien Coulet, Malika Smaïl-Tabbone, Pascale Benlian, Amedeo Napoli, and Marie-Dominique Devignes. “Ontology-guided data preparation for discovering genotype-phenotype relationships”. In: *BMC Bioinformatics* 9.S4 (2008).

- [R15] Benjamin Leblanc, Adrien Coulet, Estelle André, Franck Molina, and Jean-Paul Leonetti. “Data processing and exchange tools to facilitate chemical genetic screening processes.” In: *BioTechniques* 37.2 (2004), pp. 223–5.

International conferences

- [C1] Stephen Pfohl, Adrien Coulet*, Ben Marafino*, Fatima Rodriguez, Latha Palaniappan, and 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.
- [C2] Gabin Personeni, Marie-Dominique Devignes, Malika Smaïl-Tabbone, Philippe Jonveaux, Céline Bonnet, and Adrien Coulet. “Cooperation of bio-ontologies for the classification of genetic intellectual disabilities: a diseasome approach”. In: *SWAT4HCLS 2018 - 11th International SWAT4HCLS Conference Semantic Web Applications and Tools for Healthcare and Life Sciences*. Antwerp, Belgium, Dec. 2018.
- [C3] Pierre Monnin, Mario Lezoche, Amedeo Napoli, and Adrien Coulet. “Using formal concept analysis for checking the structure of an ontology in LOD: the example of DBpedia”. In: *23rd International Symposium on Methodologies for Intelligent Systems, ISMIS 2017*. Warsaw, Poland, June 2017.
- [C4] Kevin Dalleau, Ndeye Coumba Ndiaye, and Adrien Coulet. “Suggesting valid pharmacogenes by mining linked data”. In: *International Conference on the Semantic Web Applications and Tools for Life Sciences (SWAT4LS) 2015*. Cambridge, United Kingdom, 2015.
- [C5] Gabin Personeni, Simon Daget, Céline Bonnet, Philippe Jonveaux, Marie-Dominique Devignes, Malika Smaïl-Tabbone, and Adrien Coulet. “ILP for Mining Linked Open Data”. In: *the 24th International Conference on Inductive Logic Programming (ILP 2014)*. 2014.
- [C6] Gabin Personeni, Simon Daget, Céline Bonnet, Philippe Jonveaux, Marie-Dominique Devignes, Malika Smaïl-Tabbone, and Adrien Coulet. “Mining Linked Open Data: A Case Study with Genes Responsible for Intellectual Disability”. In: *10th International Conference on Data Integration in the Life Sciences, DILS 2014*. Vol. 8574. LNCS series. Springer, July 2014, pp. 16–31.
- [C7] Adrien Coulet, Florent Domenach, Mehdi Kaytoue, and Amedeo Napoli. “Using Pattern Structures for Analyzing Ontology-Based Annotations of Biomedical Data”. In: *International Conference on Formal Concept Analysis*. LNCS/LNAI series. Dresden, Germany: Springer, May 2013.
- [C8] Mehwish Alam, Adrien Coulet, Amedeo Napoli, and Malika Smaïl-Tabbone. “Formal Concept Analysis Applied to Transcriptomic Data”. In: *The Ninth International Conference on Concept Lattices and Their Applications - CLA 2012*. Fuengirola (Málaga), Spain, Oct. 2012.
- [C9] Adrien Coulet, Malika Smaïl-Tabbone, Amedeo Napoli, and Marie-Dominique Devignes. “Role Assertion Analysis: a proposed method for ontology refinement through assertion learning”. In: *STAIRS*. Vol. 179. Frontiers in Artificial Intelligence and Applications. IOS Press, 2008, pp. 47–58.
- [C10] Adrien Coulet, Malika Smaïl-Tabbone, Pascale Benlian, Amedeo Napoli, and Marie-Dominique Devignes. “SNP-Converter: an Ontology-Based solution to Reconcile Heterogeneous SNP Descriptions for Pharmacogenomic Studies”. In: *Data Integration in the Life Sciences 2006, DILS’06*. LNCS series. European Bioinformatics Institute (EBI), Hinxton/UK: Springer, July 2006.

International workshops

- [A1] Pierre Monnin, Chedy Raïssi, Amedeo Napoli, and Adrien Coulet. “Knowledge Reconciliation with Graph Convolutional Networks: Preliminary Results”. In: *DL4KG2019 - Workshop on Deep Learning for Knowledge Graphs*. Vol. CEUR Workshop Proceedings. Proceedings of the Workshop on Deep Learning for Knowledge Graphs (DL4KG2019) Co-located with the 16th Extended Semantic Web Conference 2019 (ESWC 2019) 2377. Portoroz, Slovenia, June 2019.

- [A2] Joël Legrand, Yannick Toussaint, Chedy Raïssi, and 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, and Adrien Coulet. “Combining Concept Annotation and Pattern Structures for Guiding Ontology Mapping”. In: *FCA4AI@IJCAI2018 - 6th International Workshop ”What can FCA do for Artificial Intelligence?”* Vol. CEUR Workshop Proceedings. Proceedings of the 6th International Workshop ”What can FCA do for Artificial Intelligence”? co-located with International Joint Conference on Artificial Intelligence and European Conference on Artificial Intelligence (IJCAI/ECAI 2018), Stockholm, Sweden, July 13, 2018 2149. Stockholm, Sweden, July 2018.
- [A4] Joël Legrand, Yannick Toussaint, Chedy Raïssi, and Adrien Coulet. *Tree-LSTM and Cross-Corpus Training for Extracting Biomedical Relationships from Text*. DLPM2017 Workshop - 2nd International Workshop on Deep Learning for Precision Medicine. Held in conjunction with ECML-PKDD 2017. 2017.
- [A5] Pierre Monnin, Clement Jonquet, Joël Legrand, Amedeo Napoli, and Adrien Coulet. “PGxO: A very lite ontology to reconcile pharmacogenomic knowledge units”. In: *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, and Adrien Coulet. “Discovering ADE associations from EHRs using pattern structures and ontologies”. In: *Phenotype Day, Bio-Ontologies SIG, ISMB*. Orlando, United States, 2016.
- [A7] Mohsen Hassan, Olfa Makkaoui, Adrien Coulet, and Yannick Toussaint. “Extracting Disease-Symptom Relationships by Learning Syntactic Patterns from Dependency Graphs”. In: *Proceedings of BioNLP 15*. Beijing, China: Association for Computational Linguistics, 2015, pp. 71–80.
- [A8] Mohsen Hassan, Adrien Coulet, and Yannick Toussaint. “Learning Subgraph Patterns from text for Extracting Disease–Symptom Relationships”. In: *ECML/PKDD 2014 Workshop on Interactions between Data Mining and Natural Language Processing, DMNLP’14*. 2014.
- [A9] Mehwish Alam, Melisachew Wudage Chekol, Adrien Coulet, Amedeo Napoli, and 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*. 2013.
- [A10] Mehwish Alam, Adrien Coulet, Amedeo Napoli, and Malika Smaïl-Tabbone. “Formal Concept Analysis Applied to Transcriptomic Data”. In: *FCA4AI*. 2012, pp. 7–14.
- [A11] Adrien Coulet, Russ B. Altman, Mark A. Musen, and 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, and Marie-Dominique Devignes. “Ontology Refinement through Role Assertion Analysis: Example in Pharmacogenomics”. In: *Description Logics*. Vol. 353. CEUR Workshop Proceedings. CEUR-WS.org, 2008.
- [A13] Adrien Coulet, Malika Smaïl-Tabbone, Amedeo Napoli, and Marie-Dominique Devignes. “Suggested Ontology for Pharmacogenomics (SO-Pharm): Modular Construction and Preliminary Testing”. In: *OTM Workshops (1)*. Vol. 4277. LNCS series. Springer, 2006, pp. 648–657.

National conferences

- [c1] Pierre Monnin, Amedeo Napoli, and Adrien Coulet. “Discovering Subsumption Axioms with Concept Annotation”. In: *Gestion de Données - Principes, Technologies et Applications (BDA 2017)*. Poster. Nov. 2017.
- [c2] Adrien Coulet, Florent Domenach, Mehdi Kaytoue, and Amedeo Napoli. “Using pattern structures for analyzing ontology-based annotations of biomedical data”. In: *Septième Journées d’Intelligence Artificielle Fondamentale*. Aix-en-Provence, France, June 2013, pp. 97–106.

- [c3] Clement Jonquet, Adrien Coulet, Nigam H. Shah, and Mark A. Musen. “Indexation et intégration de ressources textuelles à l’aide d’ontologies : application au domaine biomédical”. In: *21èmes Journées Francophones d’Ingénierie des Connaissances, IC’10*. 2010, pp. 271–282.

National workshops

- [a1] Gabin Personeni, Emmanuel Bresso, Marie-Dominique Devignes, Michel Dumontier, Malika Smaïl-Tabbone, and Adrien Coulet. “Découverte d’associations entre Événements Indésirables Médicamenteux par les structures de patrons et les ontologies”. In: *Journée I.A. et Santé*. Nancy, France, July 2018.
- [a2] Gabin Personeni, Marie-Dominique Devignes, Michel Dumontier, Malika Smaïl-Tabbone, and 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, June 2016.
- [a3] Adrien Coulet, Marie-Dominique Devignes, and 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.

Popularisation articles

- [V1] Adrien Coulet and Nicolas Jay. “Contre les effets indésirables, un algorithme pour personnaliser les doses de médicaments”. In: *The Conversation FR* (2019).
- [V2] Adrien Coulet and 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.

Invited editorials

- [E1] Adrien Coulet, K. Bretonnel Cohen, and 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), pp. 825–826.
- [E2] Adrien Coulet, Nigam H. Shah, Lawrence Hunter, Chitta Baral, and 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, pp. 485–487.

3 Main results of the five past years

3.1 Predicting the need to reduce drug dose, before prescription

Within the Snowflake, which became Snowball, Stanford–Inria Associate Team, we developed predictive models from patient Electronic Health Records. In particular we developed models that predict before prescription if a patient will necessitate a drug dose reduction, and by consequence may benefit from a reduced dose from the beginning of the treatment. These results have been published in Scientific Reports [R4], and illustrate the importance of developing a “learning” health system, *i.e.*, a health system that build upon accumulated data to make better decisions. In particular, our work show that such a system may help clinicians prescribing drugs for which dosage is sensitive, and then adverse drug reactions are frequent.

3.2 The PractiKPharma project, funded by the ANR

Since April 2016, I am supervising the PractiKPharma project (*Practice-based evidences for actioning Knowledge in Pharmacogenomics*) funded by the French Research Agency (ANR) for 42 mois and 675 000 €. PractiKPharma is a collaborative project joining 4 academic partners: 2 computer science labs, the LIRMM in Montpellier and the Loria in Nancy ; and two university hospitals, the CHU St-Etienne and the Georges Pompidou European Hospital of the *Assistance Publique des Hôpitaux de Paris*.

PractiKPharma is a computer science project applied to the domain of pharmacogenomics, which studies the impact of genetics on drug response phenotype variability. In this domain, the quality of knowledge is highly variable since its majority still require more validation before to be potentially use in clinical practice. The aim of PractiKPharma is to validate state-of-the-art knowledge units that need more evaluation, by mining EHR warehouses from our partner hospitals. One of the challenge is to compare elements of knowledge that are usually not comparable: results of studies published in the literature, and observations made at the patient level in clinical data. PractiKPharma enabled to publish (as of Nov. 2019) 9 journal, 7 conference and 11 workshop articles. They are referenced on the HAL platform and are accessible from this link.

3.3 The supervision of a pharmacy thesis, but in computer science!

I had the chance to supervise Kevin Dalleau, a student with a highly interdisciplinary profil, since he studied both pharmacy and computer science. His interest for both domain led us to work on link prediction in biomedical linked data and their publication in a conference [C4] and a journal [R5] . His pharmacy thesis has been awarded by the annual price of the experimental thesis of the School of Pharmacy of Nancy.

3.4 The creation of a major in Data Science at TELECOM Nancy

From the teaching point of view, a noteworthy contribution I made in the French panorama of graduate school for engineering is the creation in 2014 of a major in Data Science at TELECOM Nancy. This program was one of the first of this kind in France at the time, and inspired several other graduate schools. This major in Data Science has been enlighten and supported by the Pasc@line association (now *Talents du Numérique*), which groups French graduate schools in Computer Science. Since 2014, over 100 students graduated from this major and most of them pursue their carrier in the domain of Data Science.