BILAN

L'objectif clairement affiché du workshop intitulé « Medical Information, Information Retrieval, and Data Sciences" et qui s'est déroulé le 9 février 2018 était de réunir des chercheurs en biologie, informatique, mathématique et médecine.

Le programme était d'ailleurs conçu de telle sorte que les orateurs soient de ces différentes disciplines. Certains orateurs étaient des séniors au niveau de leur recherche alors qu'une place a également été réservée aux juniors (doctorantes en informatique et en biologie). La parité était également de mise : parmi les 9 orateurs, 4 sont des femmes. D'un point de vue scientifique, une femme et un homme organisaient ces journées.

Recevant des personnalités internationales de renoms, la journée s'est déroulée en anglais ; tant dans sa partie exposé, questions que dans les discussions qui ont suivies.

Concernant les orateurs :

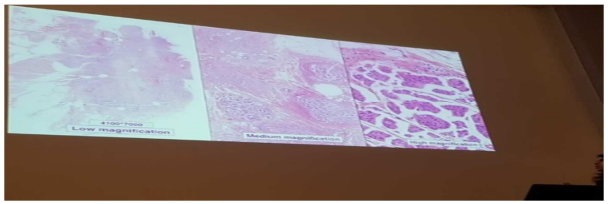
Lena Maier-Hein est professeur et responsable du département santé du centre de recherche Allemand sur le cancer (DKFZ) à Heidelberg en Allemagne ; elle est titulaire d'une bourse ERC et a reçu plusieurs prix pour ses travaux. Son exposé ***Biophotonique computationnelle et science des données chirurgicales pour le traitement du cancer de nouvelle génération*** a particulièrement été apprécié tant il donnait une vue globale de ses recherches pointues tout en les rendant accessibles à tous.

William Hersh, professeur en informatique et responsable du Département d'informatique médicale et d'épidémiologie clinique à l'École de médecine de l'Oregon Health & Science University (OHSU) à Portland, Oregon, États-Unis. est un leader et un innovateur en informatique biomédicale à la fois dans l'éducation et la recherche. Il a présenté les ***Opportunités et challenges de l'utilisation des données cliniques, incluant les données cliniques des patients***.

Lorraine Goeuriot, maitre de conférences à l'Univ. de Grenoble Alpes et chercheure à l'IMAG est responsable depuis plusieurs années du ***Lab e-Health*** à la conférence internationale d'évaluation de l'accès à l'information (CLEF) qu'elle nous a présentée.

Jérémie Pariente, professeur de neurologie au CHU Toulouse Purpan dirige l'unité de neuropsychologie et l'unité des troubles cognitifs, de l'épilepsie et des mouvements du département de neurologie. La recherche de Jérémie Pariente est basée à l'Unité de NeuroImagerie de Toulouse (INSERM) et porte sur la déficience cognitive après un AVC, une lésion cérébrale traumatique ou au cours de maladies neurodégénératives. Il nous a parlé avec un collègue mathématicien et u collègue informaticien du **projet en cours sur le suivi des patients traumatisés crâniens**.

D'autres collègues des universités Toulousaines nous ont fait l'honneur de présenter leurs travaux :



Sébastien Déjean, Ingénieur de recherche en statistiques à l'IMT, qui co-dirige la plateforme de biostatistique au sein du Génopôle Toulouse Midi-Pyrénées a présenté **les enjeux, les difficultés et les bonnes pratiques pour mener des recherches collaboratives entre biologistes et statisticiens. Sonia Mejbri**, étudiante en thèse en informatique co-encadrée par des médecins et des informaticiens nous a présenté ses recherches sur **l'utilisation des réseaux de neurones profonds pour l'analyse d'images de cancers** alors que **Nadia Amara**, également étudiante en thèse en biologie sous la direction de médecins a présenté ses recherches sur le **croisement de sources pour l'analyse des lymphomes**. **Nicolas Savy**, maître de conférences en mathématique à l'IMT a présenté les résultats du **projet MASTODONS BiDaSa** (Big Data en Santé). **Md Zia Ullah**, actuellemen tpostdoc en informatique à l'IRIT a présenté des travaux réalisés durant sa thèse àToyohashi University of Technology (Japon) sur l'utilisation de la théorie des graphes pour mettre en évidence les liens gènes-fonction.

Concernant les participants :

Parmi les 70 participants, 53 ont acceptés de remplir un questionnaire. Ces réponses permettent de voir que la répartition entre les trois disciplines (santé, informatique, mathématique) était respectée. Par ailleurs, la répartition entre séniors et juniors étaient également sensiblement de moitié pour chaque groupe.

Les répondants ont indiqué être intéressés par l'interdisciplinarité. Les retours très positifs qu'ont reçu les organisateurs faisaient également état de l'intérêt de telles journées inter ou multi-disciplinaires où chacun fait l'effort pour rendre sa discipline compréhensible sans réduire la portée et l'intérêt de ses recherches.

Plusieurs participants ont émis le souhait de renouveler cette journée. Des étudiants ont également émis le souhait que plus de publicité soit faite dans les formations et une meilleure adéquation avec le calendrier de leur formation pour leur permettre de participer sans manquer des cours par ailleurs.

Vendredi 9 février 2017, IRIT, Univ. P. Sabatier (118 Rte de Narbonne Toulouse)
Inscriptions: http://bit.ly/2mhtzPz

8h30-8h45: Welcome and opening Josiane Mothe, IRIT, Univ. de Toulouse

8h45-9h45: **William Hersh**, Oregon Health & Science University (USA)
Opportunities and challenges of using clinical data, including that from the electronic health record

9h45-10h15: **Jérémie Pariente**, CHU Toulouse Purpan
The OptimiPS-TC Project: Extraction and analysis of health data to optimize the care of cranial trauma patients

10h15-10h30: Coffee break

10h30-11h00: **Sébastien Déjean**, Université de Toulouse, IMT (Toulouse)
Collaborative researches associating biologists and statisticians

11h00-11h30: **Sonia Mejbri**, Université de Toulouse et Oncopole de Toulouse
Deep learning and image analysis

11h30-12h00: **Lena Maier-Hein,** Div CAMI DKFZ German Cancer Research Center
Computational biophotonics and surgical data science for next-generation cancer treatment

12h30-13h30 Lunch break

13h30-14h00: **Lorraine Goeuriot**, Université Grenoble Alpes, IMAG (Grenoble)
CLEF E-Health Lab, Tasks, Evaluation and Lessons learnt

14h30-15h00: **Nadia Amara**, Oncopole de Toulouse
Identification des paramètres morphologiques visant à caractériser le type de Lymphome B Diffus à Grandes Cellules et sa progression par le traitement d'image microscopique numérique

15h00-15h30: **Nicolas** Savy, Institut de Recherche en Informatique de Toulouse
BiDaSa project: From Database Merging to Simulated Clinical Trials

15h30-16h00 **Md Zia Ullah**, Univ. de Toulouse & Toyohashi Univ. of Technology
Bipartite Graph Based Ranking Method for Genetic Disease Prediction

15h30-16h00: **Radu-Tudor Ionescu**, Universitatea Bucuresti (Romania)
Machine learning, image processing and bioinformatics

16h00 Coffee break and closure

Medical Information, Information Retrieval, and Data Sciences

William Hersh is Professor and Chair of the Department of Medical Informatics & Clinical Epidemiology in the School of Medicine at Oregon Health & Science University (OHSU) in Portland, Oregon, USA. Dr. Hersh is a leader and innovator in biomedical informatics both in education and research.

In education, he serves as Director of OHSU's Biomedical Informatics Graduate Program. Dr. Hersh also spearheaded OHSU's efforts in distance learning for biomedical informatics.

Dr. Hersh has made many contributions in research, focused mainly in the area of information retrieval (IR). His recent work has focused on use of IR techniques to identify patient cohorts for clinical research studies. https://dmice.ohsu.edu/hersh/

IR Meets EHR: Retrieving Patient Cohorts for Clinical Research Studies

The growing amount of patient data in electronic health record (EHR) systems provides opportunity for clinical research and other tasks. This talk describes efforts applying techniques of information retrieval (IR) to identify patients who might be candidates for clinical research studies. It also discusses challenges for this sort of task, namely the challenges for using data generated for documenting patient care as well as maintaining its privacy.

Jérémie Pariente,
jeremie.pariente@inserm.fr , MD, PhD

is professor of Neurology, CHU Toulouse Purpan. He is running the Neuropsychology Unit and the Cognitive, Epilepsy and Movement disorder Unit of the Neurology Department. Jérémie Pariente's research is based in the Toulouse NeuroImaging Unit (INSERM) and focused on cognitive impairment after stroke, traumatic brain injury or in the course of neurodegenerative diseases.

The OptimiPS-TC Project: Extraction and analysis of health data to optimize the care of cranial trauma patients

In our clinical practice, we face the problem of clinical outcome after a traumatic brain injury (TBI). Clinical outcome depends on several factors including age, initial severity, primary care, rehabilitation…No protocol or systematic follow up was developed in our institution for patients after a TBI. The consequence is an absence of global vision regarding TBI care in our institution. This includes the absence of information regarding putative modifying factors that may influence recovery. Therefore, we decided to develop a research program to extract and analyze of health data to optimize the care of TBI using software used in the CHU Toulouse. In this talk, the first steps of this program will be explained.

Medical Information, Information Retrieval, and Data Sciences

Sébastien Déjean, sebastien.dejean@math.univ-toulouse.fr

is Research engineer in statistics at the Toulouse Mathematics Institute (IMT), Université de Toulouse. He co-leads the biostatistics platform inside Génopôle Toulouse Midi-Pyrénées (a network of technology platforms dedicated to innovation in life science, www.genotoul.fr). His interests are in applied statistics, biostatistics, data integration methods and data-mining. He authored and/or maintains several R packages (www.r-project.org) for exploring and analyzing data (mixOmics, CCA, HAPim).

Collaborative researches associating biologists and statisticians

In this talk, I will discuss about interdisciplinary projects involving biologists and statisticians. From the difficulties to understand each other to the success of fruitful collaborations, I will highlight the role of the biostatistic platform inside Genotoul to act as a centre of expertise to link actors from various domains.

Sonia Mejbri, sonia.mejbri@irit.fr

is a second-year Ph.D. student in Computer Science under the supervision of Professor Josiane Mothe and Dr Emmanuel Faure at Institut de Recherche en Informatique de Toulouse. Her research interests lie at the intersection of Image Processing, Medical Data Segmentation, Machine Learning and Deep Learning. Prior to beginning my doctoral work, she was graduated with a M.S of Information Processing and Life Complexity from The National School of Engineers of Tunis (ENIT), Tunisia (Joint degree with the Univ. de Paris Descartes, Paris V).

Deep learning on multivariate medical data : CNN-based segmentation of breast cancer histology images

I will present my current research which is directed toward the development of histological image analysis approaches that can detect different types of tissue present in the slide, classify them, and provide visual prompts to the pathologist using deep learning models. I will introduce and describe deep learning architectures used in my researches. Then I will discuss some results obtained on the breast cancer dataset provided by the Oncopole of Toulouse. My PhD topic is to apply machine learning techniques in particular deep learning approaches on multivariate medical data to automate segmentation and classification tasks of medical images.

Medical Information, Information Retrieval, and Data Sciences

Prof. Lena Maier-Hein received the doctoral degree (Dr.-Ing.) from Karlsruhe Institute of Technology KIT with distinction in 2009 and conducted her postdoctoral research in the Division of Medical and Biological Informatics at the German Cancer Research Center (DKFZ) in Heidelberg, Germany, and at the Hamlyn Centre for Robotics Surgery at Imperial College London. As a full professor and department head at the DKFZ she is currently leading the Division of Computer-assisted Medical Interventions. Her primary research interests are in multi-modal data fusion in the presence of uncertainties, surgical data science and computational biophotonics with a strong focus on clinical translation. She has received numerous awards, including the Ingrid-zu-Solms Prize 2009/2010, the Heinz Maier-Leibnitz Prize 2013 and the Berlin-Brandenburg Academy Prize 2017. She is/ has further been (co-) principal investigator on a number of national and international grants including a European Research Council (ERC) starting grant 2014.

Computational biophotonics and surgical data science for next-generation cancer treatment

Key challenges in interventional tumor diagnosis and therapy consist of the detection and discrimination of malignant tissue as well as the monitoring of tissue perfusion. While traditional approaches in the field of computer assisted surgery are typically based on multi-modal data fusion of intra-operative data with “offline“ pre-operative images, our concept focusses on optimizing the information acquired during surgery. Based on recent biophotonics techniques, including multispectral optical and optoacoustic imaging, as well as modern machine learning approaches, our methods allow for Augmented Reality visualization of a range of important morphological and functional tissue parameters, such as blood oxygenation, without relying on highly accurate deformable registration algorithms. We further work on surgical data science techniques to provide context-aware assistance to physicians. The work to be presented in the talk has resulted from the European Research Council (ERC) starting grant COMBIOSCOPY – Computational Biophotonics in Endoscopic Cancer Diagnosis and Therapy.

Lorraine Goeuriot lorraine.goeuriot@univ-grenoble-alpes.fr is assistant professor (maître de conférence) in the Université Grenoble Alpes. Her research interests are mainly related to information retrieval and natural language processing centered on the processing of medical information. She had worked or is working on text retrieval, IR evaluation, opinion mining.

Medical Information Retrieval and its evaluation: an overview of CLEF eHealth evaluation task

In this talk, I will introduce my research activities in the field of medical information retrieval, and in particular its evaluation.
The use of the Web as source of health-related information is a wide-spread phenomena, and laypeople often have difficulties finding relevant documents. The goal of the CLEF eHealth evaluation challenge is to provide researchers with datasets to improve consumer health search. I will firstly introduce the task and the datasets built. Then I will describe some experiments and results obtained on this dataset.

Medical Information, Information Retrieval, and Data Sciences

Nadia Amara, Oncopole de Toulouse, Project manager and PhD Student amara.nadia@iuct-oncopole.fr

DIFFUSE B-CELL LYMPHOMA (DLBCL): Phenomics Data Analysis

My talk will be dealing with the results of the first step of my PhD. My presentation will summarize the preliminary data obtained in the characterization of the phenomics properties of the diffuse large B-cell lymphoma. We anticipate that this will to better diagnosis and treatment the patients.

Nicolas Savy (nicolas.savy@math.univ-toulouse.fr)

Assistant Professor at University of Toulouse III, Toulouse Institute of Mathematics since 2006, Nicolas Savy received a PhD in applied Mathematics in 2003 and a HDR in applied Mathematics in 2014. Nicolas is in charge of CNRS research group « Statistics and Health ». His main skills are Malliavin calculus, Stochastic calculus, Modelling and Simulation in clinical trials, Statistics applied to medical research, Clinical Trials designing.

BiDaSa project: From Database Merging to Simulated Clinical Trials.

BiDaSa (Big Data en Santé) is a project funded by CNRS MASTODONS'17 grant and leaded by Sébastien Déjean, Philippe Saint-Pierre and myself. The starting point of this project comes from the popularity of Big Data in Health which raises many technical issues. Few questions involving "real" Big Data strategy have given relevant results because of the poor quality and the huge heterogeneity of the data collected. An alternative to this strategy is to focus on specific databases more accurate and to merge it. A technical issue raises when the databases are not coded in the same way. An algorithm of recoding (OT-algorithm) has been developed to overpass this problem. Another axis of BiDaSa project is the question of Simulated Clinical Trials which consists in using the data collected from the first clinical trials on a new compound to perform simulation studies. These simulations may allow to optimize the design of a new clinical trial. The main issue consists in generating virtual patients as close as possible to the real ones (in terms of correlations between covariates).

Medical Information, Information Retrieval, and Data Sciences

Md Zia Ullah, mdzia.ullah@irit.fr
is Post-doc researcher at IRIT in the Université Toulouse III - Paul Sabatier. He has received his Ph.D. degree from the Toyohashi University of Technology (TUT), Japan in 2016. His research has focused on query performance prediction, learning to rank, intent mining and diversification, and bipartite graph-based ranking. He is also conducting research in microblog retrieval (e.g. Twitter) and graph-based genetic disease.

Bipartite Graph Based Ranking Method for Genetic Disease Prediction

With vast amounts of medical knowledge available on the Internet, it is becoming increasingly practical to help doctors in clinical diagnostics by suggesting plausible genetic diseases predicted by applying data and text mining technologies. In this talk, firstly I will present a bipartite graph-based technique for ranking genetic diseases for a set of clinical phenotypes. Then, I will describe the dataset, conducted experiments, and results obtained from these experiments.

Language: **mainly English**

Participation is free of charge but you need to register:
http://bit.ly/2mhtzPz

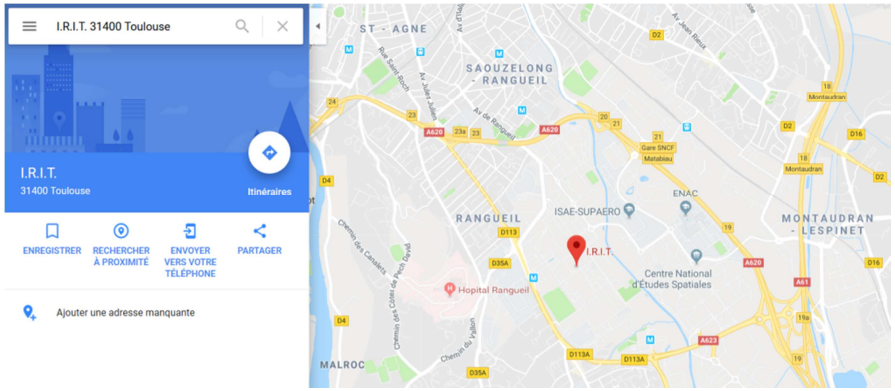
When? **8 février 2018**

Where?

Auditorium J. Herbrand
IRIT - Université Toulouse 3 Paul Sabatier (UT3)

- **Adresse** : 118 Route de Narbonne, F-31062 TOULOUSE CEDEX 9
- **Tél** : (33) 5 61 55 67 65
- **Fax** : (33) 5 61 55 62 58
- **Contact** : info at irit.fr

Métro Ligne B : **Université Toulouse** III Paul Sabatier



http://bit.ly/2mEHlv6

https://www.google.com/maps/place/I.R.I.T.,+31400+Toulouse/@43.561944,1.467911,14z/data=!4m5!3m4!1s0x12aebc3b5e85cf73:0x482eaf46a6227309!8m2!3d43.5617391!4d1.4680172?hl=fr