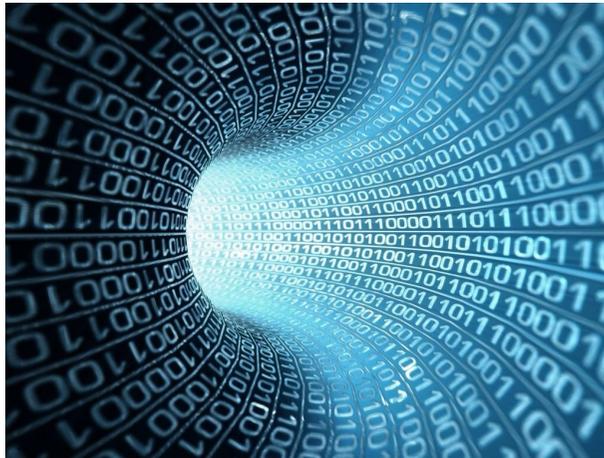


## Big data : données x modèles = Valeurs & Fun, René LEFEBURE, Directeur R&D, CONEXANCE MD

Le développement d'Internet et de la traçabilité des produits se traduit par une croissance de plus en plus forte de la collecte de données. Face à cette croissance exponentielle, les entreprises ont besoin de nouvelles compétences informatiques ou statistiques pour interpréter et utiliser ces données.

Ce phénomène s'annonce comme transverse dans tous les aspects de la vie et bouleversera notre quotidien en tant qu'utilisateur ou consommateur



**Contact :**  
iut.stid@univ-lille2.fr  
Département STID  
25-27, rue du Maréchal Foch  
59100 ROUBAIX  
03 20 65 95 52

## Découverte de la science de données en cette année mondiale de la statistique



**Le jeudi 10 octobre 2013  
de 13h à 17h  
Département STID - IUT C - Université Lille 2**



Université Lille 2  
Droit et Santé



Université Lille 2  
Droit et Santé



## Programme

### Conférences :

- 13h00-13h30 : La biostatistique, **Saïd SERBOUTI**
- 13h35-14h05 : Visualisation des données issues des bases de données ouvertes, **Fatma BOUALI**
- 14h10-14h40 : Big data : données x modèles = Valeurs & Fun, **René LEFEBURE**

### Ateliers en parallèle de 15h à 17h (30 minutes par atelier) :

- L'Analyse de données à l'aide du logiciel R, **Vincent VANDEWALLE**
- Démonstration du logiciel EXOD (Exploration of Open Data) , **Fatma BOUALI**
- Le requêtage dans les bases de données, **Khalid GABER**

---

### « Les données sont le pétrole du XXI<sup>ème</sup> siècle », David McCandless

En cette année mondiale de la statistique, le statisticien armé de ses outils informatiques est plus que jamais au cœur des défis de son époque. Nous en exposerons deux aspects lors de ces conférences. D'une part, le statisticien accompagne les chercheurs dans les progrès de la médecine, en les aidant à interpréter les résultats d'expériences de plus en plus complexes. D'autre part, il analyse la jungle de plus en plus dense des données disponibles, notamment depuis l'avènement d'Internet, pour en extraire les informations essentielles.

Des ateliers permettront d'illustrer quelques éléments de la boîte à outils du statisticien.

### La biostatistique, Saïd SERBOUTI, enseignant en biostatistique, statistique et mathématiques à l'IUT de Lille 2

La médecine et la biologie continuent leur essor à un rythme particulièrement soutenu depuis plusieurs décennies. On assiste aussi à une grande diversification des méthodes expérimentales, des procédures et des concepts. Les systèmes biologiques sont multidimensionnels et les différences entre les individus sont grandes, les méthodes quantitatives dépendent donc du type d'analyse recherché.

Depuis le début du 20<sup>ème</sup> siècle, des techniques puissantes d'analyse ont vu le jour pour permettre aux biologistes et aux médecins de mieux analyser et interpréter les résultats obtenus. Nous verrons que les capacités croissantes des calculateurs continuent d'ouvrir de nouvelles perspectives dans ces domaines.

### Visualisation des données issues des bases de données ouvertes, Fatma BOUALI, Maître de Conférences en Informatique, Université Lille 2. Travail réalisé en collaboration avec Tianyang LIU doctorant en informatique et Gilles Venturini, Professeur des Universités en informatique à l'Université François Rabelais de Tours

Le nombre de données accessibles librement depuis l'avènement d'Internet ne cesse de croître. Cet ensemble forme ce qui est communément appelé l'*Open Data*. La France a commencé à s'engager dans la diffusion de ces données publiques, notamment à travers la plate-forme [data.gouv.fr](http://data.gouv.fr). EXOD (**EX**ploration **of** **O**pen **D**ata) est un outil que nous avons développé pour la visualisation de grandes collections de jeux de données ouvertes. Cet outil, après téléchargement des données, en extrait les caractéristiques essentielles à l'aide d'une analyse de texte. Puis à partir de celles-ci, il établit un graphe qui relie les jeux de données similaires entre eux. Ce graphe peut être visualisé sous forme d'une carte globale ou interactive, qui permet à l'utilisateur d'exploiter plus efficacement l'information disponible. Nous présenterons les résultats obtenus par cet outil sur 293.000 jeux de données issus de l'*Open Data* français.