

TABLE RONDE :  
“POURQUOI ET COMMENT ENSEIGNER  
L’ANALYSE DE DONNÉES MASSIVES (BIG DATA)”

Organisée par

Chloé Friguet, IUT de Vannes, [chloe.friguet@univ-ubs.fr](mailto:chloe.friguet@univ-ubs.fr)

Frédérique Letué, IUT2/STID UPMF, [frederique.letue@iut2.upmf-grenoble.fr](mailto:frederique.letue@iut2.upmf-grenoble.fr)

Vincent Vandewalle, IUT de Roubaix, [vincent.vandewalle@univ-lille2.fr](mailto:vincent.vandewalle@univ-lille2.fr)

**Résumé.**

Les dernières années ont connu une grande effervescence autour du “big data” ou données massives. Celles-ci soulèvent de nouveaux enjeux scientifiques autour des problèmes de stockage des données (volume des données massives), de leur hétérogénéité (variété) et de leur traitement en temps réel (vélocité). Ces enjeux relèvent de l’informatique, mais aussi de la statistique. Face à ce défi, de nombreux établissements proposent aujourd’hui des modules, voire des formations entières dédiées au “big data”, la demande en spécialistes de ce nouveau domaine étant très forte. La table ronde proposée abordera les enjeux pédagogiques liés à ces nouvelles formations.

**Mots-clés.** Enseignement de la statistique, données massives.

**Abstract.**

The last years have generated a huge scientific excitement around what is called “big data”, or massive data. These data give rise to new scientific issues around data management (due to the data volume), heterogeneity (variety) and real-time treatment (velocity). These concerns fall under computer science, but also under statistics. Therefore, some schools and universities propose now course units or trainings dedicated to big data, the needs for such specialists being very high. The proposed round table will raise some issues linked with these new trainings.

**Keywords.** Statistical education, big data.

## TABLE RONDE :

### “POURQUOI ET COMMENT ENSEIGNER L’ANALYSE DE DONNÉES MASSIVES (BIG DATA)”

Les dernières années ont connu une grande effervescence autour du “big data” ou données massives. Celles-ci se distinguent habituellement des statistiques classiques par les trois V: volume, variété, vélocité. En effet, les données massives se caractérisent par des données de très grande dimension (volume) ce qui engendre des besoins de stockage de données dans des data centers. Ces données sont issues de sources variées et pas toujours structurées : les données sont donc de nature très diverses et doivent être traitées en tenant compte de cette variété. Enfin, les données arrivent à très haute fréquence et doivent donc être traitées quasiment en temps réel, avec des temps d’accès aux données très rapides et des temps de traitement très courts. Les enjeux des données massives semblent donc se placer beaucoup du côté informatique (recueil, stockage, organisation, interrogation de ces données volumineuses), mais également du côté statistique avec la nécessité de mettre en oeuvre de nouveaux algorithmes statistiques mettant plus l’accent sur la rapidité d’accès aux données.

Ces nouveaux enjeux scientifiques et techniques en amènent d’autres sur le plan pédagogique, car il va être nécessaire de former de nouveaux professionnels, capables d’appréhender ces nouveaux problèmes. Devant la demande croissante des entreprises sur ce sujet, de nombreux établissements proposent aujourd’hui des modules, voire des formations entières dédiées au “big data”. Il convient alors de s’interroger sur ces enseignements et leur lien avec les enseignements traditionnels de statistique :

- Que recouvrent exactement les formations intitulées “big data” ?
- Quelles sont les compétences générales nécessaires pour devenir un spécialiste des données massives ?
- Quels sont les sujets communs et ceux qui s’en distinguent entre les formations “big data” et les formations classiques de statistique ?
- Quels sont les intervenants (notamment professionnels) qui sont susceptibles d’intervenir dans de telles formations ?
- Quels sont les logiciels incontournables dans le domaine ?
- Quels sont les métiers visés et à quel niveau (bac+2, +3, +5 ?) se placent-ils ?

Pour répondre à ces questions, la table ronde laissera dans un premier temps la parole à des enseignants impliqués dans des formations “big data” :

- Philippe Besse (Ingénieurs GMM, INSA de Toulouse)
- Stéphane Cléménçon (master de Telecom) *sous réserve de son accord*
- Aurélien Garivier (master SID de Toulouse) *sous réserve de son accord*
- Sophie Lambert-Lacroix (LP Big data Grenoble)
- Jean-Michel Poggi (DU Paris Descartes)
- Nathalie Villa-Vialaneix (Master Economics & Statistics, Toulouse School of Economics).

Dans un second temps, les participants à cette session auront l’occasion d’échanger avec les intervenants pour approfondir le débat.