

TSE FORUM  
DIGITAL  
BIG DATA

LES TEMPS FORTS  
DU FORUM 2018

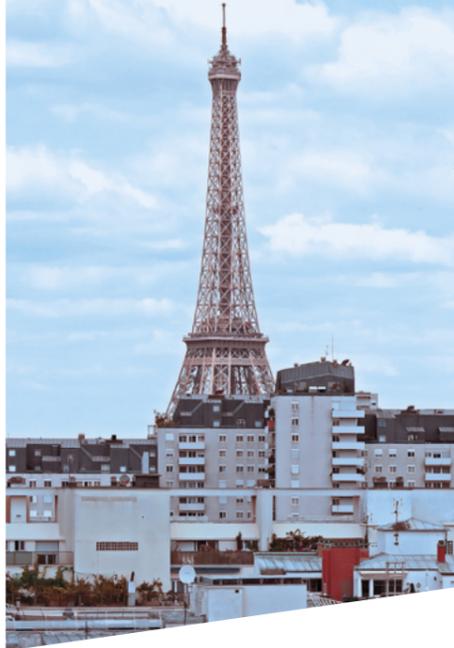
PARIS - Vendredi 22 Juin 2018

8h30 - 13h30

Hôtel NOVOTEL VAUGIRARD

Le temps d'une matinée, économistes, décideurs économiques et régulateurs se sont réunis afin d'échanger sur les grandes questions de l'économie numérique et notamment la place et l'impact des Big Data pour nos sociétés.

Wilfried Sand-Zantman et Eric Gautier ont ainsi présenté leurs récents travaux sur le sujet tandis qu'une table ronde a confronté les points de vue de Jacques Crémer, Amal Taleb et Hubert Tardieu.



## Wilfried Sand-Zantman (TSE)

### Le rôle des données pour les entreprises

Le chercheur de TSE, expert de la concurrence, a tenu à nuancer la révolution amorcée par l'avènement du Big Data, estimant ainsi que *"les données sont un bouleversement, souvent comparé au pétrole, pourtant il y a plusieurs différences majeures. Si elles sont similaires dans les changements qu'elles apportent, contrairement au pétrole, les données sont générées par toute activité. Elles sont en quelque sorte un produit dérivé. D'autre part, les données peuvent s'échanger sans perte. Enfin, elles ne représentent pas une nouveauté, les banques, assurances ou opérateurs téléphoniques les utilisent ainsi depuis longtemps."*

Wilfried Sand-Zantman a ensuite insisté sur l'importance de l'impact des données : *"la valeur d'une donnée dépend des décisions qu'elle permet de prendre."* Et ces décisions ne seront pas les mêmes selon qu'un acteur économique est le seul à disposer d'une donnée ou que celle-ci soit partagée : *"l'information est à double tranchant dans la mesure où les acteurs économiques réagissent à l'information possédée par les autres. Cette réaction peut annuler en grande partie les bénéfices de détenir une information privilégiée"*. Si une information est partagée par tous les concurrents, cela conduit à des offres tarifaires agressives qui érodent les profits au bénéfice des consommateurs.

Enfin l'économiste a détaillé les ciblage publicitaires que permettent les données en rappelant les deux types de ciblage *"un ciblage positif, c'est-à-dire qui consiste à sélectionner les gens que l'on veut toucher, et un ciblage négatif qui permet d'écarter des consommateurs qui ne seraient pas intéressés. Le ciblage optimal dépend du marché ciblé."*



“ La valeur d'une donnée dépend des décisions qu'elle permet de prendre ”

## Eric Gautier (TSE)

### Big Data et prise de décision

Eric Gautier a présenté les enjeux du traitement des données de grande volumétrie : *" Si nous pensons qu'une base de donnée consiste en beaucoup de lignes (les observations) et beaucoup de colonnes (les variables), cela pose deux problèmes de nature différente. Le fait d'avoir beaucoup de lignes est une aubaine car l'incertitude statistique est faible. Cela peut néanmoins poser un problème d'un point de vue computationnel et nécessiter des algorithmes distribués ou un sous-échantillonnage bien pensé. Avoir beaucoup de variables pose un problème très différent. L'économétrie et la statistique classique supposent le nombre de variables fixe et reposent sur des approximations probabilistes. Celles-ci ne sont plus valables lorsque le nombre de lignes est faible par rapport au nombre de colonnes."*

Eric Gautier a ensuite présenté les défis du point de vue de l'optimisation et les solutions basées sur les relaxations convexes et la parcimonie : *"Celle-ci repose sur l'hypothèse que les nombreuses inconnues de la modélisation sont bien approchées par un objet simple, par exemple beaucoup de paramètres sont nuls ou des matrices sont de faible rang."*

Eric Gautier a conclu cette intervention en présentant des axes de recherche actuels issus de l'analyse économique où les outils développés en apprentissage statistique ne sont pas adaptés notamment en analyse causale et pour l'étude de "contrefactuels".



“ Le problème du Big Data c'est qu'à partir d'une certaine taille de données, tous les traitements qui seraient simples deviennent extrêmement complexes ”

## Table ronde

# Optimiser la valeur des données personnelles

**Animée par Jean-Marc Vittori** (Les Échos)

**Jacques Crémer** (TSE)

**Amal Taleb** (SAP)

**Hubert Tardieu** (Atos)

Les échanges ont été vifs et nombreux entre les trois participants. Jacques Crémer a insisté sur le pessimisme et la peur qui entourent le sujet actuellement et a rappelé que les données, si elles sont bien régulées, doivent permettre de meilleurs échanges et donc favoriser la croissance : *“combien de livres avez-vous lus que vous n’auriez jamais connus sans les systèmes de recommandations que permettent les données personnelles ?”*

Amal Taleb a apporté son point de vue de juriste à la question : *“les données personnelles étaient facilement définissables dans le passé. Légalement il s’agissait de toute information qui permettait de cibler un individu au sein d’un ensemble : nom, code client, numéro d’adhérent etc. C’est beaucoup plus complexe aujourd’hui.”*

Hubert Tardieu a rebondi sur les propos de Jacques Crémer et rappelé que les données personnelles ne sont pas perçues comme problématiques dès lors qu’elles permettent un service : *“le grand problème auquel sont confrontés les entreprises c’est qu’elles ne pensent pas au service qu’elles peuvent rendre aux utilisateurs avant de leur demander leurs données. Quand le service passe avant les données, l’exploitation des données personnelles ne pose pas de problèmes.”*

Hubert Tardieu a ensuite expliqué pourquoi selon lui les mélanges entre données personnelles et données d’entreprise seront au cœur des débats, notamment légaux. Amal Taleb a précisé ce point en rappelant que le consentement n’était pas l’alpha et l’oméga de la législation sur le sujet.

Jacques Crémer a conclu en expliquant que de nombreux travaux de recherche seraient nécessaires afin de comprendre les impacts de ces données massives et leurs échanges : *“il faudra beaucoup de travaux de recherche de longue durée sur le sujet et c’est ce que nous essayons de faire à TSE.”*

