



Journées de la Statistique 2018

Défis et enjeux des mégadonnées (« *Big Data* »)

Jeudi 26 avril 2018, 9h00 à 15h30

UQÀM, Pavillon de design (DE)

1440, rue Sanguinet, Montréal, H2X 3X9 Salle 2560

Cette journée est organisée conjointement par la Société Statistique de Montréal (SSM) et le Collectif pour le développement et les applications en mesure et évaluation (Cdame) de l'UQÀM.

Depuis quelques années les possibilités/capacités grandissantes des ordinateurs sont en train de transformer la science et la gestion des organisations.

On a vu apparaître des mégadonnées, au centre de ces nouveaux développements.

Les possibilités de traitement des données, manipulations et visualisations hyper rapides et extrêmement-complexes sont-elles en train de rendre caduque l'approche scientifique traditionnelle?

La statistique y trouve-t-elle encore sa place?

Nos conférenciers, chacun à leur manière, nous exposeront leurs réflexions.

**Chaque conférence dure 50 minutes,
suivies de 10 minutes de questions**

9h00-9h30 Café, brioches.

9h30-10h30 Yves Gingras, professeur titulaire
Département d'histoire et Centre interuniversitaire de
recherche sur la science et la technologie, UQÀM.

Titre. Que faire devant l'avalanche des nombres ?

Résumé. La mode est aux mégadonnées censées révolutionner le monde... La revue branchée *Wired* prédit la fin des théories et les nouveaux gourous de l'informatique pensent que les ordinateurs sont intelligents et vont remplacer les humains... Mon exposé tentera de faire la part des choses dans ce bruit assourdissant en rappelant des tentatives passées de tout fonder sur des « données » et en montrant que les discours sur les « données probantes » se fondent en fait sur une vision simpliste et positiviste de la prise de décision. Je conclurai en notant la responsabilité sociale accrue qui incombe aux statisticiens dans un monde où les chiffres sont partout et sont souvent manipulés au service d'idéologies et d'intérêts privés qui ne servent pas le bien commun.

10h30-11h30 Bernard Clément, Professeur titulaire
Département de Mathématiques et Génie industriel
Polytechnique Montréal.

Titre. Statistique, Sciences des données, Intelligence artificielle :
liens, différences et convergence

Résumé. Nous allons tenter de comprendre les relations et les implications entre la statistique traditionnelle, les mégadonnées, la nouvelle discipline dite de la science des données et l'enthousiasme actuel et délirant pour le domaine de l'intelligence artificielle. Nous tenterons d'élaborer une certaine synthèse entre les domaines respectifs dont les frontières sont encore mal définies et provoquent beaucoup de discussions entre les différentes disciplines scientifiques. Mon objectif est de susciter la réflexion et la discussion.

11h30-12h30 Jean-Herman Guay, professeur titulaire
École de politique appliquée, Université de Sherbrooke.

Titre. Les mégadonnées, une planche de salut pour les sciences sociales ?

Résumé. Il y a quelques années le célèbre sociologue Raymond Boudon écrivait à propos de sa propre discipline : «la sociologie risque de se condamner à devenir une branche mineure de la littérature ou une activité auxiliaire du journalisme ». Il n'avait pas complètement tort. Depuis presque trente ans, les sciences sociales sont considérées comme perpétuellement en crise : conflit paradigmatique, conflit de méthodologies et conflit d'ambitions. Qu'est-ce que les sciences sociales apportent à la société? Quelle plus-value représentent-elles? Or, aujourd'hui, avec toutes les traces laissées par les individus sur leurs comportements, leurs opinions et leurs déplacements, les sciences sociales disposent de données comme jamais auparavant. À la limite, les mégadonnées sont peut-être la planche de salut des sciences sociales. Elles vont peut-être leur permettre de s'affirmer comme de vraies

sciences. Dans cette conférence, j'entends poser le problème, envisager les réponses possibles et apporter une synthèse à ce qui est à la fois un péril et un défi.

12h30 Lunch en commun. (Lieu à prévoir)

13h30-14h30 Jean-François Plante, professeur agrégé, Département des Sciences de la décision, HÉC.

Titre. Défis, outils et utilisations des mégadonnées, Aperçus tirés du *Boot camp of the thematic program on Statistical Inference, Learning and Models for Big Data*, tenu au *Fields Institute* de Toronto en 2015.

Résumé. La Conférence inaugurale du *Boot camp of the thematic program on Statistical Inference, Learning and Models for Big Data*, s'est déroulée au *Fields Institute*, à Toronto du 12 au 23 janvier 2015. Trente-cinq exposés furent présentés qui ont offert une vaste perspective sur le thème. Même si on trouve des mégadonnées dans de nombreuses disciplines et contextes, on a vu apparaître des défis et plusieurs outils communs, de même que plusieurs utilisations de l'inférence dans divers contextes. Dans cet exposé, je vais présenter les défis et outils les plus fréquents, et résumer quelques exemples de réussite. Ce travail est basé sur un papier publié dans la *ISI Review* par les *Fellows* et visiteurs à long terme du *Fields Institute* qui ont participé à ce programme d'études des mégadonnées.

14h30-15h30 Gilles Raïche, professeur titulaire Sciences de l'éducation, UQÀM.

Titre. Un exemple d'applications à l'enseignement de mégadonnées publiques en éducation : le cas du PISA (Programme International de Suivi des Acquis)

Résumé. La disponibilité publique des données des enquêtes à grande échelle en éducation est d'une grande utilité pour soutenir l'enseignement des statistiques. Celles-ci sont disponibles sur plusieurs années et pour un très grand nombre de pays. Il en résulte des échantillons importants et un accès à nombre de variables qui permettent d'élaborer du matériel pédagogique en lien avec l'enseignement de grande quantité de procédures statistiques. À titre d'exemple, les données du PISA seront présentées. Dans cette présentation, on présentera leurs caractéristiques ainsi que comment se les procurer, les manipuler et les apparier entre elles. Il sera aussi question de certains choix qui doivent être faits de par la nature de ces données ainsi que de certaines spécificités de l'utilisation de logiciels statistiques courants.

15h30 Le mot de la fin. Cocktail.