

Bruno Teboul, Keyrus : « Avec l'École Polytechnique, nous voulons former des bataillons de Data Scientists »

Parrainée par **Geneviève Fioraso**, secrétaire d'État à l'Enseignement supérieur et à la Recherche, la toute nouvelle chaire « Data Scientist » de l'École Polytechnique s'inscrit dans le cadre du [plan Big Data visant les 9 milliards d'euros en 2020](#) et la création de 130 000 emplois dans le secteur. **Bruno Teboul**, directeur scientifique, R&D et innovation du Groupe Keyrus, spécialiste de la valorisation de données et mécène de la chaire avec Thales et Orange, détaille les objectifs ambitieux de l'initiative.

Silicon.fr – Quels sont les objectifs de la chaire Data Scientist de l'École Polytechnique ?

Bruno Teboul – Les objectifs de la chaire Data Scientist en tant que chaire d'enseignement et de recherche sont multiples et sa genèse est née du constat suivant : le Big Data nous impose de concevoir de nouvelles méthodes et technologies en matière d'exploitation de ces mégadonnées associant les mathématiques appliquées, l'algorithmique, l'informatique avancée et le management. Et ce afin d'extraire le sens et la valeur des informations pertinentes au service de la performance des acteurs économiques.

Notre premier objectif est de **soutenir l'École Polytechnique et le plan Big Data du gouvernement**, en oeuvrant à la reconstruction de la souveraineté numérique de la France qui doit faire de notre pays un leader mondial dans le domaine du traitement des mégadonnées. Les entreprises doivent désormais faire face à ces nouveaux défis et donc recruter des profils rares et très pointus, dont l'École Polytechnique par son excellence pluridisciplinaire rend légitime et possible.

Notre deuxième objectif est de **former les bataillons de « Data Scientists »** capables de maîtriser ces nouveaux enjeux scientifiques, techniques et économiques. C'est ainsi que nous allons créer toute une filière de formation et d'emploi, une filière d'avenir pour la France, qui permettra de faire rayonner l'enseignement supérieur français dans le monde entier. L'une des activités phares de cette chaire étant son programme de formation des élèves polytechniciens au métier de Data Scientist, complété par le lancement d'un Master 2 en Data Science.

Notre troisième objectif est d'**intégrer ces profils une fois diplômés au sein de nos équipes** chez Keyrus, Thales et Orange en leur proposant, pour commencer, des stages au cours de leur scolarité. Enfin, notre quatrième objectif se focalise sur la recherche fondamentale en « Data Science ». Nous encouragerons et soutiendrons les recherches originales et leurs publications dans des revues scientifiques prestigieuses, tout en innovant par le déploiement de projets disruptifs en traitement des mégadonnées (algorithmique, génie logiciel).

Comment mieux adapter les compétences aux besoins du marché ?

Là encore, notre souhait est de répondre aux besoins précis du marché – pénurique – en matière de formation de Data scientists par des parcours et des programmes d'excellence dispensés par l'École Polytechnique. Depuis la nouvelle spécialisation du cycle d'ingénieur polytechnicien par la création d'un **Master 1 en Data Science**, en passant par l'ouverture d'un **Master 2 délivré conjointement par l'X et Telecom Paristech**, jusqu'aux formations doctorales. Celles-ci correspondent à un besoin chez nombre d'industriels, dont nos propres partenaires Thales et Orange. Ils accueillent déjà ce type de profils, des doctorants sous contrat alliant à la fois intelligence théorique et pratique, ouverture d'esprit et sens de l'innovation.

Néanmoins, le spectre de connaissances et de compétences étant extrêmement large, pour former de véritables Data Scientists il conviendra d'envisager également des enseignements autour du marketing, du management et du consulting. Il s'agit d'acquérir les « soft skills » qui permettront aux virtuoses du traitement des données massives de présenter leurs résultats de manière intelligible et visuelle : **l'art de la datavisualisation** étant crucial, en tant que « last mile » de la data science. Pour ce faire, nous envisageons d'étendre à d'autres départements de l'X leurs participations pédagogiques. Nous pensons à l'économie, l'économétrie, la finance, les sciences de gestion, les sciences humaines et les « humanités numériques ».

Enfin, nous n'avons pas oublié **la formation professionnelle** qui est fondamentale pour accompagner tous les acteurs économiques dans cette mutation sociétale que je qualifie de « Datasphère » et qui implique l'ensemble des collaborateurs concernés par le traitement et la valorisation des données : **de la DSI au Marketing** (data miners, responsables BI et administrateurs BDDS). Par conséquent, il est fondamental de fournir aux collaborateurs impliqués et impactés par la Data un cursus de formation adaptée, gage d'une transition numérique réussie. C'est pourquoi l'X a déjà construit le [Data Science Starter Program](#), une formation certifiante sur 2 mois, que nous soutenons depuis début octobre à travers l'inscription de plusieurs collaborateurs chez Keyrus.

Quelle est votre perception de l'évolution des projets Big Data en France ?

J'aimerais distinguer la vision du marché établie par les cabinets d'études tels que IDC, EMC ou encore Markess, vision fondée sur du déclaratif, et la réalité de notre terrain quotidien. Les instituts tablent tous sur une augmentation très importante du nombre de projets Big Data en France, autour de 40% de progression en moyenne en 2014 par rapport à 2012 ; mais le nombre de projets par entreprise interrogée reste faible (échantillon = 220). Quant à la maturité des acteurs, considérer que 75% ont une bonne, voire une excellente connaissance du sujet, permettez-moi un peu d'en douter...

Nous constatons **une augmentation significative des projets autour de l'analyse prédictive** – mais nous sommes dans la continuité de notre cœur de métier autour du décisionnel et sur des

processus statistiques dit supervisés, fondés sur l'historique –, ainsi que sur l'analyse des données non-structurées (essentiellement web et réseaux sociaux) à des fins de connaissance client et de segmentation marketing plus fine et augmentée. En parallèle, nous recevons beaucoup de consultations et d'appels d'offres autour d'études de cadrage et d'opportunité Big Data ou bien sur des problématiques plus techniques liées aux architectures Big Data et notamment à la clusterisation Hadoop, en environnement Cloud.

Peu d'acteurs comprennent réellement ce que l'**exploitation des données massives en temps réel** peut leur apporter comme effet de levier et/ou comme catalyseur en matière de prise de décisions. Même les acteurs du web et du e-commerce français sont en retard sur le sujet et n'ont toujours pas saisis l'intérêt du machine learning par exemple, de l'algorithmique au service de la prédiction et de la recommandation. Nous sommes dans la deuxième étape de l'acculturation, de l'évangélisation, après la période « buzz word », il y a eu une phase courte de rejet. Désormais, nous mettons en place un grand nombre de projets via une démarche de « maïeutique » qui fait éclore « in vivo », en mode « lab » les projets Big Data portés par les métiers. Le principal frein a souvent été les désaccords, les controverses incessantes entre DSI et Direction Marketing sur les projets Big Data. Le manque d'expertise étant le critère qui désormais revient le plus souvent chez nos clients et qui constitue maintenant une énorme opportunité justifiant notre démarche de création de chaire Data Scientist.

Lire aussi

- [Marketing et DSI s'écharpent autour du Big Data](#)
- [Big Data : beaucoup de projets, peu de déploiements](#)
- [Data scientist et directeur technique : 2 jobs en or \(infographie\)](#)