

Le focus technique

L'ACQUISITION DE DONNÉES ET LE BIG DATA

Les capteurs ont envahi notre quotidien. Ils se multiplient aussi dans l'industrie, grâce notamment aux communications sans fil qui facilitent grandement le développement de nouvelles applications, la surveillance en ligne par exemple. De ce fait, les volumes de données à traiter augmentent de façon exponentielle. Pour les traiter et les exploiter, on entre dans l'univers du Big Data, un concept qui a fait ses preuves ailleurs, pour par exemple étudier les comportements des internautes. Appliqué à l'univers de la mesure, le Big Data ouvre de nouvelles perspectives et peut aller jusqu'à repenser la manière d'exploiter les valeurs mesurées...



Guillaume BOURDON
Cofondateur et dirigeant de Quinten.



Daniel LEROY
Directeur général de Alliantech.



Bruno CATHALA
Responsable test et mesure chez HBM.



Thierry MIALON
Directeur commercial de Alliantech.



Max FEINBERG
Consultant dans l'agroalimentaire, auteur d'ouvrages techniques.



Jean-Michel POU
Président de Delta Mu.



Christian FLACHARD
Directeur commercial de Creative IT.



Emmanuel ROSET
Ingénieur produits acquisition de données chez National Instruments.

L'ACQUISITION DE DONNÉES ET LE BIG DATA 12 QUESTIONS

1. L'ACQUISITION DE DONNÉES, UN DOMAINE EN EXPANSION ? PAGE 42
2. TOUT LE MONDE PEUT-IL RÉALISER UNE ACQUISITION DE DONNÉES ? PAGE 43
3. QUELS SONT LES PIÈGES À ÉVITER OU LES DIFFICULTÉS À SURMONTER ? PAGE 43
4. QUELLES ONT ÉTÉ LES DERNIÈRES ÉVOLUTIONS MARQUANTES ? PAGE 44
5. COMMENT PEUT-ON DÉFINIR LE BIG DATA DANS L'UNIVERS DE LA MESURE ET DU CND ? PAGE 45
6. QUI EST CONCERNÉ PAR LE BIG DATA ? PAGE 46
7. N'ABUSE-T-ON PAS DES POSSIBILITÉS DE L'ÉLECTRONIQUE ? PAGE 46
8. NE FAIT-ON PAS DEPUIS LONGTEMPS DU BIG DATA SANS LE SAVOIR ? PAGE 47
9. LE BIG DATA EST-IL UNIVERSEL ? PAGE 47
10. QU'ELLES SONT LES BASES THÉORIQUES DU BIG DATA ? PAGE 47
11. EXISTE-T-IL DES OUTILS SPÉCIFIQUEMENT DÉVELOPPÉS POUR FAIRE DU BIG DATA ? PAGE 48
12. À QUOI FAUT-IL S'ATTENDRE DEMAIN ? PAGE 48

1.

L'ACQUISITION DE DONNÉES, UN DOMAINE EN EXPANSION ?



Emmanuel Roset : Expansion ?

Le mot est faible. Je dirais que l'acquisition de données est un domaine en explosion, grâce notamment aux capteurs dits "intelligents" et "connectés" (même sans fil). La facilité avec laquelle il est désormais possible de placer des capteurs un peu partout permet de collecter plus d'informations sur les systèmes (y compris les ouvrages, maisons, bâtiments, voitures, corps humain, etc.) que l'on souhaite améliorer ou surveiller, à distance, et ainsi de prendre les bonnes décisions. Cette évolution nécessite des outils de "raffinage" des données brutes pour extraire les informations utiles, dans un bruit de données incompréhensible sans une puissante analyse.



Daniel Leroy : Dans l'industrie, la tendance de fond du marché est de relier la métrologie au système d'information (l'informatique, ndlr).

Les principaux facteurs de cette évolution sont la complexité croissante des produits, la réduction des cycles de développement et le renforcement des exigences réglementaires et juridiques. L'acquisition de données embarquée, en tant que partie intégrante de cette réponse technique, va donc croître durablement. Par exemple, la maintenance industrielle évolue vers la surveillance en ligne, alors que l'on se contentait jusqu'ici d'une inspection périodique.



Christian Flachard : Dans nos applications de suivi de production, nous constatons une forte expansion de l'acquisition de données, que ce soit pour un besoin de suivi du

process ou pour un besoin de contrôle qualité. Le fait que les équipements de production et les équipements de mesure soient de plus en plus communicants facilite bien sûr cette expansion ; de ce point de vue, l'adoption de la norme de communication OPC par un grand nombre de systèmes a été un tournant déterminant.

Les industriels, mais surtout leurs clients et les organismes chargés de faire des audits, sont rassurés par le fait que les données proviennent directement des équipements plutôt que d'une saisie opératoire. C'est vrai pour la traçabilité matière, pour la traçabilité process et de nombreux autres domaines.



Jean-Michel Pou : Dans le monde actuel, le pilotage des process industriels repose sur des *a priori*.

On estime (ou on a déterminé, mais là aussi avec des croyances ou des idées reçues) qu'il faut que tel ou tel paramètre soit compris entre telles et telles limites pour obtenir la qualité attendue. Le data mining, au contraire, repose sur des recherches *a posteriori*. Il s'agit de retrouver les conditions qui ont permis d'obtenir la qualité attendue. Pour donner à cette nouvelle approche le maximum de chances de trouver le Graal (l'optimal), il faut disposer d'un maximum d'informations, même si elles ne paraissent pas pertinentes *a priori*. Dans ce domaine, tout est bon à prendre et le rêve du *data scientist*, c'est de trouver des explications là où on ne les attendait pas. Cette volonté de sortir des *a priori* conduira donc probablement à des acquisitions "tous azimuts", c'est-à-dire plus de mesures, partout et encore...



Max Feinberg : Je limiterais mon propos au domaine de l'analyse chimique et biochimique. Bien que peu connu du grand public, c'est un secteur où on s'intéresse aux données "massives" (ancêtre du Big Data) depuis plusieurs décennies.

Il y a 20 ans, on trouvait quelques dizaines d'opérateurs dans un laboratoire qui faisaient quelques milliers de

mesures par an. Aujourd'hui, on n'en a plus que deux ou trois qui en font des dizaines de milliers. Cette fantastique augmentation de productivité a correspondu à une informatisation croissante des instruments et des unités d'analyse. Tous les laboratoires modernes possèdent un logiciel dédié, appelé LIMS (*Laboratory Information Management System*). Il gère une base de données adaptée aux fonctions du laboratoire. Par ailleurs, la majorité des instruments de mesure est connectée en réseau et pilotée par des ordinateurs. Dans le même temps, les laboratoires se sont fait accréditer et ont été souvent regroupés au sein de grands groupes. Tous ces facteurs font que l'acquisition des données est en croissance et que tous les outils nécessaires au Big Data sont déjà là.

Reste la question de l'utilisation de ces données. En effet, il est très rare qu'elles fassent l'objet d'une exploitation poussée : on les stocke aussi longtemps que la loi l'oblige, puis on les détruit. Par exemple, en biologie médicale, on pourrait imaginer (rêver à) une mise en commun des mesures qui permettrait d'améliorer les protocoles d'analyses par de simples techniques de machine-learning...



Bruno Cathala : Pour la conception des prototypes, l'avènement des modèles de simulation fiables a permis de diminuer les besoins en acquisition de données.

La simulation nécessite toutefois des essais réels pour la confirmer et en optimiser le processus. Ces dernières années, les attentes en termes d'optimisation (temps ou matière première) a occasionné une demande croissante en matériels d'acquisition de données dans tous les domaines. En effet, les démarches d'optimisation des moyens de R&D, essais, qualité et production ainsi que les actions engagées pour limiter les pertes de matières ont conduit bon nombre de fabricants de machines à s'équiper en système d'acquisition de données afin de pouvoir attester de leur performance ou des actions correctives à apporter.

« DU BON SENS. »

Thierry MIALON

« LES BASES THÉORIQUES DU BIG DATA REPOSENT SUR DES SAVOIRS MATHÉMATIQUE ET INFORMATIQUE. »

Daniel LEROY

De même, la limitation des ressources humaines et l'augmentation des classifications produits (type "ranking") contribuent à alimenter cette tendance. Bon nombre de produits sont à présent qualifiés en fonction de leurs niveaux de performance et de leurs caractéristiques. Chacune de ces qualifications à nécessité des campagnes de mesure et l'utilisation de moyens d'acquisition.



Guillaume Bourdon : Je ne voudrais pas jeter un coup de froid. Mais il y a déjà tant à analyser dans les organisations avant d'investir dans la collecte !

2.

TOUT LE MONDE PEUT-IL RÉALISER UNE ACQUISITION DE DONNÉES ?

Emmanuel Roset : Tout le monde fait désormais de l'acquisition de données, sans même le savoir. De nombreux outils de la vie courante (comme les smartphones ou les bracelets électroniques par exemple) se transforment en puissants systèmes d'acquisition fiables pour mesurer les pressions atmosphériques, les températures, les champs magnétiques et bien d'autres grandeurs.

Les outils deviennent plus simples, même dans le domaine industriel. Les besoins en formation sophistiquée sont donc de moins en moins importants. Cependant, quel que soit l'outil, il est toujours nécessaire d'effectuer une formation adaptée afin de tirer parti de l'ensemble des fonctions proposées et d'avoir une vision précise des possibilités. La tendance est dorénavant à l'apprentissage par les outils web comme les vidéos en autoformation, bien plus efficaces et

rapides à appréhender qu'une documentation papier.



Christian Flachard : L'acquisition de données s'est grandement simplifiée depuis plusieurs années, il est de plus en plus aisé de récupérer des valeurs sur des équipements récents pour peu qu'ils soient dotés d'un serveur OPC, d'une base de données ou d'une interface sous forme de fichier. Mais attention ! Récupérer une donnée n'est qu'une partie de l'acquisition. Au-delà de la connexion aux équipements, il faut être vigilant sur l'intérêt de la donnée, son contexte, le bon moment pour la récupérer. Un logiciel de supervision ou une trieuse pondérale pourront donner de multiples informations, encore faut-il choisir les bonnes, au bon moment pour les rattacher au process en cours, à l'ordre de fabrication en cours, au numéro de lot en cours.



Jean-Michel Pou : Le véritable enjeu, c'est de s'assurer de la fiabilité de la donnée qui est transmise puis analysée. Et là, l'expérience compte !

Dans le Big Data, on recherche les spécifications (les règles) à partir des expériences réussies. Mais si la mesure est de mauvaise qualité, "la règle" n'est pas transférable et n'est pas valable dans le temps, elle n'est pas "robuste" (cas des changements de capteurs). Les enjeux de la fiabilité des données mesurées (incertitude) sont donc beaucoup plus importants dans le Big Data qu'aujourd'hui.

d'hui. Le rôle de la métrologie, au sens «science des mesures» (et pas simplement gestion des instruments) pourrait bien devenir enfin prépondérant !



Max Feinberg : Une tendance classique est de penser que des données peuvent être collectées sans arrière-pensée, de façon "naturelle". C'est faux ! Avant de commencer la collecte, il faut toujours avoir en tête le but recherché, ce qui permet de gagner du temps. Quand on acquiert des données et on gagne du temps en réfléchissant au but recherché. Et cela demande une formation minimum. Ne jamais oublier l'adage *Garbage In, Garbage Out* qui peut être traduit par "à partir de la cochonnerie, on obtient de la cochonnerie".



Bruno Cathala : La vraie valeur ajoutée aujourd'hui n'est plus seulement de savoir comment faire une acquisition de données, mais plutôt comment définir le besoin et comment interpréter les résultats obtenus. Il devient alors très important de bien segmenter les intervenants, car définir une chaîne de mesure adaptée à un besoin et en utiliser les données est beaucoup plus complexe que la simple mise en route de celle-ci.

3.

QUELS SONT LES PIÈGES À ÉVITER OU LES DIFFICULTÉS À SURMONTER ?

Thierry Mialon : Les recommandations tombent sous le coup du bon sens. Il faut bien connaître son application et son environnement, savoir adapter son besoin aux solutions techniques existantes et limiter la quantité de données aux paramètres utiles. Il faut aussi tenir compte de la facilité d'apprentissage du

« LA VRAIE VALEUR AJOUTÉE RÉSIDE DANS LA DÉFINITION DU BESOIN ET DANS L'INTERPRÉTATION DES RÉSULTATS OBTENUS. »

Bruno CATHALA

logiciel de paramétrage et enfin avoir un logiciel politique d'archivage.



Guillaume Bourdon : Une bonne acquisition de données commence toujours en se posant la seule question qui vaille : quelle va être son utilité d'un point de vue opérationnel, autrement dit, en ai-je vraiment besoin ? Ensuite, sous quel format dois-je la collecter ? Et enfin, quelle structure de base de données et donc quel système d'information dois-je utiliser pour réaliser simplement mes analyses ? Ce n'est pas la quantité qui compte, mais bien la qualité et ce qu'on veut faire des données. Le piège à éviter est de vouloir tout collecter. La donnée coûte cher, ne collectons pas n'importe quoi, n'importe comment.



Bruno Cathala : L'adéquation entre le besoin et le matériel mis en œuvre et la prise en considération de l'ensemble de la chaîne de mesure est primordiale et constitue à mon sens l'essentiel des sources d'erreur possible. Il est donc primordial de s'assurer que chaque étape du processus de sélection ou d'utilisation d'un système d'acquisition reste en ligne avec les objectifs fixés, tant en termes "matériel" que "humain". Ceci étant fait, il est évident que la qualité et le suivi du matériel utilisé ou la compétence des opérateurs seront à prendre en compte. De plus en plus de logiciels intègrent une notion d'assistance sur ces points et vous indiquent les actions et performances que vous pouvez attendre, certains d'entre eux permettent même de renseigner l'opérateur sur l'incertitude théorique de la chaîne de mesure en fonction de ses éléments constitutifs.



Jean-Michel Pou : Pour compléter ce qui vient d'être dit, je voudrais insister sur le rôle primordial de

la métrologie, avec deux missions qui me paraissent essentielles. La première est de garantir que le capteur ne dérive pas dans le temps, car chaque mesure, à chaque instant, compte. La deuxième consiste à s'assurer, lors d'un changement de capteur, que les expériences passées restent cohérentes avec les nouvelles observations (robustesse de la mesure aux changements de capteurs).



Emmanuel Roset : Qu'est ce qu'une bonne acquisition de données ? Certaines personnes répondront qu'elles sont très attachées à la précision des mesures avec une grande résolution, alors que d'autres souhaitent simplement une grande vitesse. La réponse se situe donc au niveau de la qualité de l'information attendue par rapport à ce que l'on cherche à savoir sur le système sous test.

Dans le cas d'un système d'acquisition avec des éléments hétérogènes ne venant pas du même constructeur, il est important d'avoir une bonne connaissance de la précision de tous les étages de la chaîne de mesure. Un des pièges à éviter est la méconnaissance de la précision du système et un excès de confiance sur les valeurs annoncées à 6 chiffres alors que le système est à 15 % de la grandeur physique réelle.



Max Feinberg : En chimométrie, qui s'intéresse depuis plus de 20 ans au Big Data, certaines méthodes de mesure dites à haut débit (*high throughput*) sont capables de produire plusieurs gigaoctets de données en quelques minutes. Au début, les analystes passaient plus de temps à interpréter ces données qu'à collecter les mesures. On peut tirer quelques enseignements de l'expérience acquise dans ce domaine. En premier lieu, il ne faut pas oublier qu'une

étape de prétraitement des données est souvent indispensable pour pouvoir appliquer un algorithme. Mais sa durée est imprévisible, elle influence la modélisation ultérieure et elle peut être très gourmande en temps. Ensuite, il faut savoir que les problèmes de modélisation augmentent avec le nombre de variables mesurées. Les statisticiens parlent de "malédiction de la dimensionnalité". Même si cette malédiction, grâce à de nouveaux algorithmes, est moins terrible que ce qu'on pensait autrefois, un excès de données ralentit la vitesse de traitement. Un choix judicieux et a priori des informations fait gagner beaucoup de temps et d'argent. Enfin, il ne faut pas négliger la vérification de la cohérence des données, en termes de valeurs atypiques et de données non renseignées.

4.

QUELLES ONT ÉTÉ LES DERNIÈRES ÉVOLUTIONS MARQUANTES ?



Emmanuel Roset : La technologie évolue énormément en fonction des besoins. Je l'observe en particulier dans l'enregistrement des données, qui est passé du papier au stockage en réseau dans le *cloud*. L'intérêt est d'analyser un plus grand volume d'informations et de tirer de meilleures conclusions sur la décision à prendre pour améliorer ou surveiller le système sous test. Les systèmes autonomes qui échangent des données brutes évoluent donc vers des systèmes capables d'analyser en temps réel et d'aider à une prise de décision rapide. Cette évolution est accompagnée d'une distribution de l'intelligence sur les différents éléments de la chaîne, au niveau des capteurs, des systèmes d'acquisition (processeurs FPGA embarqués) et bien sûr dans le *cloud* et sa grande puissance de calcul.



Guillaume Bourdon : La puissance de calcul a considérablement évolué. Ainsi, des algorithmes d'apprentissage qui nécessi-

« L'ACQUISITION DE DONNÉES EST UN DOMAINE EN EXPLOSION, GRÂCE NOTAMMENT AUX CAPTEURS DITS INTELLIGENTS ET CONNECTÉS. »

Emmanuel ROSET

« LES INDUSTRIELS SONT RASSURÉS PAR LE FAIT QUE LES DONNÉES PROVIENNENT DIRECTEMENT DES ÉQUIPEMENTS PLUTÔT QUE D'UNE SAISIE OPÉRATEUR. »

Christian FLACHARD

taient des supercalculateurs dans les années 1990 et demandaient plusieurs semaines de calculs fonctionnent maintenant sur des petits serveurs en parallèle (*grid computing*) en quelques minutes ou en quelques jours.



Thierry Mialon : Les utilisateurs demandent des solutions de mesures intelligentes, petites, polyvalentes et autonomes, fournissant des informations en temps réel, et ce avec un coût d'usage bien plus faible pour permettre une utilisation massive. Cela a conduit au développement des MEMS, des solutions sans fils et de la récupération d'énergie locale pour l'alimentation.



Bruno Cathala : L'évolution du numérique, grâce notamment à des composants ou techniques issus du grand public, ont considérablement changé le monde de l'acquisition de données et de la mesure industrielle. Les baisses de prix de ces éléments ont permis d'augmenter le nombre de voies, la taille mémoire et la vitesse facilitant l'enregistrement d'un très grand nombre de paramètres pour mieux appréhender un phénomène.



Jean-Michel Pou : Il est assez intéressant de noter que toutes les évolutions qui viennent d'être évoquées, totalement indépendantes, trouvent un point de jonction dans le Big Data, appelé à fortement se développer dans les années qui viennent.



Bruno Cathala : Je pense qu'il est de notre devoir de sensibiliser l'ensemble des intervenants sur l'importance d'une certaine notion d'ex-

pertise et de compétence suffisamment élevée pour permettre à chacun des acteurs de notre marché (vendeur, intégrateur, client final...) de se comprendre et de saisir les nuances techniques qui existent entre deux systèmes ou deux modes de fonctionnement. Si nous nous rangeons trop derrière le seul argument de gain de temps et/ou d'argent, on en perd la notion générale de qualité.

Il me semble que le défi des années à venir sera la transmission des connaissances permettant de choisir ou de qualifier un système en fonction d'un besoin.

5.

COMMENT PEUT-ON DÉFINIR LE BIG DATA DANS L'UNIVERS DE LA MESURE ET DU CND ?



Daniel Leroy : La multiplication exponentielle des sources de mesures rend les outils traditionnels incapables de traiter et d'afficher correctement ce flux massif d'information. On entre alors dans l'univers du Big Data.



Emmanuel Roset : Appliqué au domaine de la mesure, le Big Data porte à mon sens sur la collecte et la gestion de gros volumes de données générés par une variété de systèmes d'acquisition de grandeurs physiques à haute vitesse et de formats différents. La visibilité et les prises de décisions sont effectuées par un accès aux données depuis n'importe quelle position géographique.



Jean-Michel Pou : Je ne sais pas si Big Data a une définition particulière dans l'univers de la



PMX® Amplificateur industriel



Mesure, Essais, Contrôle

Le PMX a été spécialement conçu pour équiper les bancs de contrôle en production et les lignes de fabrication.

- Plus grande productivité avec Ethernet Industriel
- Meilleure performance avec CODESYS
- Reconnaissance et identification automatique des capteurs avec la fonction TEDS
- Possibilité d'extension du système grâce à la flexibilité des cartes enfichables
- Utilisation aisée avec le serveur web utilisant la technologie GWT

PMX l'instrument de mesure industriel dernière génération de HBM.

Surveillance de machines, contrôle de lignes d'assemblage, tests de fonctionnalité, optimisation d'unités de production.

Information sur :
www.hbm.com/pmx



HBM France SAS
info@fr.hbm.com ■ www.hbm.fr

mesure ou du CND. Pour ma part, le Big Data, c'est la possibilité de sortir de croyances historiques pour découvrir de nouvelles règles inédites. Après avoir cru pendant des dizaines d'années qu'il fallait respecter telles consignes pour obtenir telles fonctionnalités, nous allons probablement nous rendre compte qu'il est possible de faire autrement, et probablement de façon plus efficiente (c'est l'objectif !), pour le même résultat. Ainsi, le Big Data, c'est l'espoir de trouver de nouvelles règles plus performantes et moins coûteuses. Mais pour atteindre cet objectif, la confiance dans les informations disponibles, donc dans les mesures, devra être plus grande qu'aujourd'hui. Du coup, le Big Data est un peu, finalement, une renaissance potentielle pour la métrologie.



Bruno Cathala : Le Big Data permet de centraliser l'ensemble des mesures de R&D avec le retour des mesures faites sur terrain ou en provenance des services qualité, voire de production. Il est ainsi possible de corréler au mieux toutes ces données pour toujours mieux appréhender le produit tout au long de son cycle de vie en vue d'une optimisation continue sans jamais oublier le point de vue économique. Il participe aussi à la traçabilité et à la sécurité des mesures en centralisant l'ensemble des données sur des serveurs sécurisés et en offrant l'accès au plus grand nombre de services.



Max Feinberg : Si le Big Data se caractérise par un grand volume, une diversité des données et une rapidité d'acquisition, tout est déjà là dans les laboratoires. Ainsi, j'ai déjà cité les méthodes à haut débit ou l'imagerie qui cumulent les trois caractéristiques : volume, diversité et vitesse. Mais cette définition me paraît un peu restreinte.

Selon mon expérience dans le domaine de l'analyse, ce qui prévaut dans le Big Data, c'est la question de l'interprétation – ou même de l'interprétabilité – des données. Si on ajoute cette facette, le Big Data démarre alors dès que le nombre de lignes d'observations dépasse une page et/ou que les mesures se répartissent sur plus de trois colonnes.



Guillaume Bourdon : On considère que l'on peut parler de Big Data à partir de 10 téraoctets de données. Jusqu'ici, on avait à traiter de telles masses de données dans le monde du web, de la téléphonie et de la génétique.

6.

QUI EST CONCERNÉ PAR LE BIG DATA ?



Emmanuel Roset : Le Big Data porte sur les applications qui génèrent tellement de données qu'elles ne peuvent pas être traitées par un seul PC local. Je pense notamment aux applications liées aux processus industriels répartis dans différentes usines ou des systèmes à très grand nombre de voies. C'est le cas des réseaux électriques intelligents (Smart Grid) qui génèrent de manière répartie au moins 5 To de données par mois ou alors des mesures sur une turbine d'avion qui peuvent requérir plus de 20 To par heure ! Aujourd'hui, les "petites" applications de surveillance de machines ne sont pas concernées, mais la technologie évoluant, tout le monde y viendra un jour.



Jean-Michel Pou : Dans le Big Data, les données valent de l'or, sous réserve qu'elles ne soient pas "en toc". Tout le monde sait qu'au-

jourd'hui la mesure sert essentiellement à la déclaration de conformité. Tout le monde sait également que les dérogations sont nombreuses, ce qui prouve que les règles actuelles (tolérances/spécifications) ne sont pas optimales. Le *data mining*, c'est au contraire, trouver les règles à partir des expériences réussies. Du coup, chaque mesure devient importante, car elle contribue à l'établissement de la règle. Chaque personne qui produit une mesure devra avoir demain, à l'esprit, son rôle essentiel dans ce nouveau cadre et, par conséquent, la nécessité de produire des données fiables.



Max Feinberg : Tous les utilisateurs d'instruments employés pour la génomique, les diverses spectroscopies multidimensionnelles (2D, 3D) ou la résonance magnétique nucléaire sont directement concernés par le Big Data. Les nouvelles techniques d'imagerie sont très prometteuses et représentent, de par les volumes produits et la vitesse d'acquisition, des exemples typiques de Big Data.

Mais le Big Data n'a pas de sens que dans la mesure où il apporte des outils d'interprétation des données. Appliqué aux médecins, il doit par exemple permettre d'interpréter une radiographie ou fournir un diagnostic à la lecture des résultats d'analyses. Il sera essentiel de proposer des outils d'interprétation suffisamment "transparents" pour qu'ils ne déshumanisent pas la démarche.

7.

N'ABUSE-T-ON PAS DES POSSIBILITÉS DE L'ÉLECTRONIQUE ?



Jean-Michel Pou : Sûrement pas ! Au contraire, seule l'électronique permet cette nouvelle révolution qui sera majeure pour notre avenir. Certains pensent même qu'il s'agit de la réelle troisième révolution de l'humanité, après les révolutions agricole et industrielle. Le Big Data, c'est en quelque sorte

« IL Y A DÉJÀ TANT À ANALYSER DANS LES ORGANISATIONS AVANT D'INVESTIR DANS LA COLLECTE ! »

Guillaume BOURDON

« LE RÔLE DE LA MÉTROLOGIE AU SENS "SCIENCE DES MESURES" POURRAIT BIEN DEVENIR ENFIN PRÉPONDÉRANT ! »

Jean-Michel POU

sortir des croyances et idées reçues pour s'intéresser aux faits ("Opinion driven" versus "Fact driven"). Seules l'accumulation de données et l'analyse de leurs conséquences permettent de sortir de modèles pour s'approcher de la réalité. George Box, statisticien, disait « Tous les modèles sont faux, mais certains sont utiles ». Le Big Data, c'est probablement un pas de plus vers une forme de réalité...⁽¹⁾



Emmanuel Roset : Nous n'abusons pas des possibilités de l'électronique, car il existe un réel besoin d'améliorer l'efficacité des machines qui produisent les biens que nous consommons : par exemple, un véhicule mieux conçu et plus fiable, ou une consommation électrique mieux répartie et plus rentable. La difficulté est d'avoir trop de données, qui paraissent incompréhensibles à première vue. Aujourd'hui, les systèmes d'acquisition sont conçus de manière à soulager les gros systèmes d'analyse par un traitement en local, en restant le moins invasif possible, comme un simple filtrage. Demain, il sera plus intéressant de réaliser l'analyse sur les données brutes, et pour cela la technologie doit encore évoluer afin d'offrir une plus grande puissance d'analyse offerte.

8.

NE FAIT-ON PAS DEPUIS LONGTEMPS DU BIG DATA SANS LE SAVOIR ?



Jean-Michel Pou : Je pense que non. Il existe une différence fondamentale entre la vision actuelle et ce qui se passera à l'avenir. Nos stratégies actuelles s'appuient sur l'inférence statistique, c'est-à-dire la modélisation de la réalité à partir de quelques échantillons (même si le "quelques" peut parfois être important).

Le Big Data, c'est sortir de la modélisation pour aller vers la réalité. Pas de modèles, mais des faits ! C'est la différence essentielle entre demain et aujourd'hui... Le Big Data, ce n'est pas seulement l'accumulation de données, mais une autre façon de voir les données !



Emmanuel Roset : Dans l'univers de la mesure et du CND, le Big Data en est encore à ses balbutiements. Mais il est très développé ailleurs. Les médias sociaux, la collecte web des habitudes de consommation, les commentaires et les avis clients sur internet relèvent tous de la "science" du Big Data. Demain, les balances individuelles et les bracelets connectés vont produire des statistiques sur la santé de catégories d'individus au niveau d'un pays ou au niveau mondial.



Guillaume Bourdon : La grande distribution est un des premiers secteurs à avoir utilisé ses données (via les tickets de caisse) afin de sortir des comportements de consommateurs et des associations d'achats de produits pour optimiser son merchandising, son marketing promotionnel et ses ventes.

9.

LE BIG DATA EST-IL UNIVERSEL ?



Emmanuel Roset : Le Big Data découle de l'évolution des technologies communicantes dans le domaine de la mesure. Il est apparu au début comme une contrainte à gérer, car il y avait soudainement trop de données à exploiter. Puis, avec l'implication de nombreux acteurs industriels qui ont créé des outils logiciels spécifiques pour l'analyse, le trop-plein d'informations est

devenu une source exploitable pour surveiller les systèmes et en déduire des actions, par exemple sur une éolienne afin de mieux répartir la production d'énergie.



Guillaume Bourdon : Le Big Data n'a pas de limite, car il suffit de se poser deux minutes pour réaliser que les données sont présentes partout. Le Big Data est un puits sans fond pour l'innovation, le pilotage et la performance.

10.

QU'ELLES SONT LES BASES THÉORIQUES DU BIG DATA ?



Daniel Leroy : Les bases théoriques du Big Data reposent sur des savoirs mathématique et informatique.



Jean-Michel Pou : Contrairement aux stratégies actuelles (SPC notamment), il n'y a rien de mystérieux dans le data mining... Pour Amazon par exemple, savoir que beaucoup des personnes qui ont acheté le livre A ont aussi acheté le livre B lui permet de proposer aux acheteurs du livre A qui n'ont pas encore acheté le livre B le dit livre B (je simplifie)... Tout cela, c'est juste du bon sens ! Au contraire, comprendre le concept d'écart-type (et sa formule) de notre monde de modélisation actuel me semble bien plus compliqué. Du coup, la compréhension du data mining est bien plus accessible que la modélisation statistique. Le fonctionnement des algorithmes de data mining, quant à lui, est effectivement mystérieux. Mais qu'importe ! Le but n'est pas de comprendre comment marche le moteur, mais de savoir à quoi sert la voiture...



Max Feinberg : Pour moi, le Big Data n'a pas de théorie clairement identifiable, car c'est une

(1) http://fr.wikipedia.org/wiki/Exploration_de_données

discipline d'assemblage qui combine des concepts et des technologies développés par d'autres disciplines (informatique et statistiques).

11.

EXISTE-T-IL DES OUTILS SPÉCIFIQUEMENT DÉVELOPPÉS POUR FAIRE DU BIG DATA ?



Daniel Leroy : Il existe aujourd'hui un grand nombre de technologies estampillées Big Data : Hadoop et Spark pour le calcul distribué, MongoDB et HBase pour les bases de données distribuées non relationnelles, GlusterFS et Ceph pour adresser les problématiques de stockage distribué haute performance. Les sociétés allergiques à l'externalisation des données peuvent créer leur propre service de cloud computing avec des produits tels que VMware et OpenStack. Enfin, concernant l'analyse des données, on dénombre un grand nombre de logiciels dont Angoss, HP Vertica et RapidMiner.



Jean-Michel Pou : Dans le monde industriel, et pour la gestion de données de production, le marché semble tout juste émergent. Nous comptons, au sein du cluster Auvergne Efficience Industrielle une société spécialisée dans le *data mining* de données industrielles : il s'agit de IP Leanware, dont la suite logicielle Braincube et Braintouch est devenue une référence dans le monde entier, notamment dans les industries du bois et du papier où l'optimisation des consommations énergétiques est un enjeu majeur. À n'en pas douter, d'autres acteurs finiront par entrer sur ce marché et il me semble que les choses vont évoluer très rapidement maintenant...



Max Feinberg : Les outils développés pour le Big Data sont trop nombreux pour en faire ici la

liste. Dans le domaine de la chimio-métrie, un ensemble de sociétés ou de spin-offs se sont créées autour du traitement des données. Par exemple, la méthode de spectrométrie proche infrarouge a permis la création de spin-offs en France qui rencontrent un grand succès. Par ailleurs, on a évoqué les méthodes du machine-learning qui ont aussi suscité la création de nombreuses startups en France et dans le monde.

12.

À QUOI FAUT-IL S'ATTENDRE DEMAIN ?



Emmanuel Roset : Dans le domaine de la mesure et du contrôle, on peut s'attendre à un regroupement mondial par standardisation des plateformes de stockage dans le cloud et à l'exploitation de ces données par des machines intelligentes autonomes réparties dans les usines. Il faudra alors gérer au mieux les sécurités d'accès à ces données collectées et produites par de très nombreuses sources de systèmes d'acquisition différents. Il est certain que d'autres outils logiciels permettront de raffiner de manière autonome les données et de fournir des décisions à d'autres machines en interaction pour atteindre un but commun. L'objectif du Big Data en mesure est de faciliter les prises de décisions par une compréhension plus globale de l'état physique d'un système, en préventif ou pour des essais.



Guillaume Bourdon : Concernant le secteur de prédilection de Quinten (la santé), le Big Data est un bienfait pour l'humanité, car il permet une accélération considérable des innovations et des progrès de la médecine en repoussant encore plus l'espérance de vie de l'individu avec une condition de vie bien meilleure. Le Big Data permet en

effet une médecine personnalisée et plus efficace. Les objets connectés permettent eux, outre des mesures médicales en temps réel, une surveillance permanente et prédictive.



Christian Flachard : L'évolution vers le monde des objets connectés et de façon plus générale vers l'industrie 4.0 va faire exploser le volume de données disponibles. Comment les industriels pourront-ils éviter d'être noyés sous l'information et se concentrer sur l'essentiel, c'est-à-dire les données à forte valeur ajoutée ? C'est sans doute l'un des principaux défis des managers du XXI^e siècle ! Nous pensons qu'il y aura désormais deux niveaux d'informations : un niveau agrégé permettant de prendre les décisions et un niveau de détail pour les experts qui souhaiteront creuser l'information et en effectuer un traitement statistique poussé.



Jean-Michel Pou : Il est probable que le Big Data et le data mining vont profondément bouleverser les pratiques industrielles. La question ne me semble pas être ce savoir si cette révolution aura lieu ou pas, mais de savoir quand le monde industriel sera, en quelque sorte, converti. Il y a 10 ans, les smartphones et leurs innombrables applications nous étaient totalement inconnus. Dans 10 ans, je prends le pari que les pratiques industrielles actuelles nous paraîtront aussi obsolètes que le minitel aujourd'hui...



Max Feinberg : Aujourd'hui, le Big Data est limité par deux choses : d'une part, le coût d'acquisition des données et d'autre part, la capacité de les utiliser pour en tirer un profit. Ainsi, il n'est pas sûr qu'un contrôle plus intensif de la teneur en dioxines des aliments permettrait de mieux assurer la santé des consommateurs à un coût acceptable. Au bout du compte, c'est le rapport entre le coût d'acquisition d'une donnée et le retour sur investissement qu'on peut en attendre (ou la connaissance qu'on va en extraire) qui va amener à collecter du Big Data ou non ●

« CE QUI PRÉVAUT DANS LE BIG DATA, C'EST LA QUESTION DE L'INTERPRÉTATION DES DONNÉES. »

Max FEINBERG