

QUAND **L'INDUSTRIE** EXPLORE LE BIG DATA.

Les grandes entreprises n'ont pas tardé à s'intéresser au *big data*. Celles-ci s'efforcent, aujourd'hui, de découvrir, via une vaste palette d'applications pilotes, quels sont les domaines les plus prometteurs auxquels appliquer cette technologie.

FRANCK BARNU

l y a un buzz phénoménal autour du big data et, cette fois, il est amplement justifié. « Les données sont le nouvel or noir », affirme-t-on ici et là. Ce n'est pas exagéré. Les grandes entreprises industrielles françaises l'ont compris et ne sont pas à la traîne en la matière. « La plupart d'entre elles ont conscience de l'importance du sujet », constate Reda Gomery, associé et responsable data & analytics chez Deloitte. Il tempère toutefois : « Si toutes ont saisi l'importance de la technologie et commencent à s'y lancer, seule une partie d'entre elles ont vraiment perçu qu'au-

delà de la technologie, l'exploitation des données massives était un puissant vecteur de la transformation numérique, et agissent de façon à profiter pleinement de ce potentiel. » L'industrie, il faut le dire, est une adepte de longue date de

l'analyse de données. Alors, que s'est-il passé pour qu'elle juge utile de se lancer à corps perdu dans le big data? Un petit rien qui change tout. Hier, les données utilisables étaient rares et leur exploitation onéreuse avec un coût croissant exponentiellement avec le volume. Aujourd'hui, les données exploitables sont en nombre illimité, leur traitement en volume beaucoup plus facile techniquement et à un coût que le volume n'affecte quasiment plus. Pour résumer abruptement : une denrée rare et chère est devenue abondante et « gratuite », ou du moins nettement moins coûteuse. De quoi changer radicalement la donne. Hier on hésitait à multiplier d'onéreuses analyses. Aujourd'hui, no limit!

Numérisation généralisée des entreprises

Plusieurs facteurs se sont combinés pour favoriser ce changement radical. *Primo*, l'augmentation de puissance des ordinateurs qui accélère notablement le traitement. *Secundo*, la numérisation généralisée des entreprises. Plus important encore, l'arrivée de nouvelles technologies (Hadoop, Map Reduce,

NoSQL,...) qui traitent désormais des volumes phénoménaux d'information en temps réel et sans sourciller. Mieux encore. L'analyse de données ne pouvait s'effectuer que sur des données structurées. Ces outils autorisent désormais le traitement de données non structurées de n'importe quel type, autrement dit de l'immense majorité des données disponibles, dont on ne savait, jusque-là, que faire. Et la corne d'abondance n'est pas près de se tarir, avec notamment le véritable déluge de données que promettent les objets connectés.

C'est une facon

nouvelle de faire

parler les données

Les techniques de traitement de données massives ont une autre grande conséquence : celles-ci changent totalement la pratique de l'analyse des données, et le profit que l'on en tire. Comprendre : le *big data*, c'est une façon radica-

lement nouvelle de « faire parler » les données. Et c'est là que réside toute sa puissance et sa richesse (si vous voulez avoir tout de suite le fin mot de l'histoire, lire l'encadré « Rien de magique », en page 44). En tout cas, tout apparaît désormais possible. En exploitant ce fabuleux trésor, l'entreprise peut inventer de nouveaux services et modèles d'affaires, améliorer son processus de prise de décision, faire des découvertes inouïes... L'industrie dispose ainsi d'un moyen d'engager sa « transformation numérique », en se réorganisant autour de la donnée qui a vocation à devenir centrale dans tous les processus. Tous les services sont impactés depuis la R&D jusqu'au marketing. Et c'est bien là le problème ! On peut tout faire, soit. Mais par quoi commencer? Quelles sont les applications qui vont réellement être le plus profitable? En outre, comme avec toute technologie nouvelle, on n'entrevoit encore qu'une microscopique parcelle de l'univers infini de possibilités qu'elle autorise. Ainsi, lorsque vous posez la question aux vendeurs, ils sont bien en peine de trouver autre chose que la maintenance prédictive - un peu

La démarche à suivre

Reda Gomery, associé et responsable data & analytics chez Deloitte, donne quelques recommandations aux industriels désireux de se lancer dans le *big data*:

- 1. Partir d'un cas d'usage métier pour répondre à une question précise;
- Adapter l'organisation, car le processus implique la collaboration entre les métiers et l'informatique;
- Proscrire l'approche méthodologique du cycle en V et y préférer une approche expérimentale test-évaluation sur des applications pilotes;
- Composer avec l'architecture informatique existante et la faire évoluer, inutile de tout changer;
- 5. Mettre en œuvre des modèles de collaboration avec des spécialistes externes et des partenariats technologiques avec des écoles ou des laboratoires académiques;
- 6. Confier le projet à une personne neutre, c'est-à-dire ni trop proche de l'informatique, ni trop proche des métiers, mais cette personne doit, en revanche, impérativement rendre compte à la direction générale.

récente annonce du groupe Safran de la création

de Safran Analytics, une nouvelle entité dédiée au

big data montre bien à quel point le sujet est pris au

Pierre Naggar, directeur EMEA de Turn, agence de marketing, cite le cas d'un client américain, Kraft Foods, qui a utilisé cette technique pour relancer un produit en déclin. En s'appuyant sur des données riches, le groupe a préparé différents messages concernant cette gamme et les a promus vers ses clients en fonction de leur comportement en ligne. Résultat : « Un taux d'appréciation de la marque et

numérique, s'y intéressent très sérieusement.



« Où en êtes-vous avec votre projet big data, M. Hoskins ? »

la tarte à la crème des applications industrielles du big data - pour illustrer les mérites de la technologie dans ce secteur...

Pendant ce temps, les industriels, extrêmement pragmatiques, se sont attelés à l'exploration de cet univers, et à la prise en main de la technologie. Ils commencent par des applications pilotes très concrètes, correspondant à une problématique métier bien identifiée, comme le recommande d'ailleurs Deloitte (lire l'encadré). On n'en est évidemment qu'au tout début mais déjà des cas d'usages variés ont vu le jour dans toutes les industries. Ils ne sont d'ailleurs pas toujours « big data » au sens de la définition stricte du concept qui implique notamment le traitement de très gros volumes, et celui de données non structurées. Parfois, l'application développée ne porte que sur un volume (relativement) restreint de données. Souvent, on se contente de travailler sur les seules données structurées. N'empêche. Cela constitue déjà une véritable incursion dans le domaine. Et la

A lire « Le big data prédictif »: il s'agit du troisième livre blanc publié par G9 +, think tank du numériaue. en partenariat avec le think tank Renaissance numérique

http://bit. ly/1vYj5Pn

Rien de magique

« Le big data n'a rien de magique. C'est une science fondée sur les mathématiques appliquées et l'informatique avancée », précise Bruno Teboul. Cette mise en garde du directeur scientifique, R&D et Innovation de Keyrus, entreprise de services numériques, n'est faite que pour mieux marquer le véritable changement de paradigme qui s'est produit avec le traitement des données statisticiens mettaient en œuvre des outils mathématiques propres à répondre à une question précise à

partir d'hypothèses fondée<u>s sur des</u> données structurées et historicisées. Avec l'arrivée de masses de données non structurées, telles que celles qui peuvent être acquises sur les réseaux sociaux en particulier, il n'est plus possible de procéder de la sorte. Désormais, on travaille à l'envers . on part des données disponibles soigneusement récoltées, nettoyées et normalisées - et on essaie d'élaborer des hypothèses. C'est la méthode dite abductive qui fonctionne par le recours à l'inférence de la meilleure explication à partir des données brutes. Et c'est le

techniques comme le machine learning, de faire émerger du sens à partir des données en appliquant les algorithmes les mieux adaptés à chaque cas. » Ce changement radical de méthode qui ne périme pas les outils classiques - s'applique aux données non structurées, comme celles structurées à traiter. « Ce qui permet parfois de faire sortir des pépites inattendues d'un immense magma de données et de trouver des réponses à des questions qu'on ne se posait pas nécessairement. » Puissant, mais pas magique!

des intentions d'achat en hausse de 11 % », assorti d'un retour sur les dépenses publicitaires « deux ou trois fois supérieur à celui des campagnes précédentes. »

La maintenance préventive, voire prédictive, est bien sûr l'autre appli de choix. Elle se pratiquait déjà chez Snecma, avant même qu'il ne soit question de big data (lire ci-contre). Les cas se multiplient désormais. SAP, par exemple, indique que le fabricant de tracteurs John Deere s'y est mis et l'éditeur allemand propose même une application spécifique de maintenance prédictive autour de son outil SAP Hana. On découvre aussi des applications comme l'optimisation de la gestion des stocks. La start-up française Lokad s'en est fait une spécialité. Elle a commencé dans le domaine du *retail* et a étendu ses prestations au secteur de l'industrie aéronautique. En moulinant des brassées de données issues de multiples sources, elle fournit à ses clients (une demi-douzaine de compagnies aériennes) des informations pour qu'ils disposent des bonnes pièces détachées au bon endroit au bon moment, le tout pour le coût minimal. C'est à la fois un problème très complexe et crucial pour les compagnies. Elles doivent à tout prix éviter d'immobiliser un avion au sol, faute de pièce de rechange, et essaient simultanément de minimiser les stocks très coûteux de pièces.

Le potentiel des objets connectés

Reste que, pour l'industrie, l'application qui recèle probablement le plus de potentiel est liée aux objets connectés. Dès lors qu'un produit est capable de fournir en continu des masses d'information sur son comportement - et celui de ses utilisateurs -, un champ immense s'ouvre à l'imagination. La maintenance prédictive déjà citée est évidemment la première exploitation de ce monitoring permanent – et en temps réel - des produits. Mais il en apparaît déjà d'autres. Chez Snecma comme chez PSA, on se montre très intéressé par les enseignements tirés des données réelles du comportement des produits en fonctionnement. Quel formidable banc d'essais, pour en améliorer la conception. Michelin Solutions, de son côté, a bâti avec CGI une application, Effifuel, qui, à partir de toutes les informations reçues depuis une flotte de camions, permet d'en réduire significativement la consommation. L'entreprise est si sûre de son coup qu'elle ne se rétribue que sur les gains réellement obtenus.

On l'a dit, la véritable puissance du big data tient au changement de paradigme qu'il introduit. Pour faire simple et court : on ne se contente plus de valider des hypothèses comme avec la statistique classique, on découvre, en moulinant des océans de données, des corrélations permettant d'échafauder des hypothèses nouvelles. Exemple typique, la façon dont Snecma envisage d'exploiter les données de ses bancs d'essais moteurs pour

Snecma

De l'analyse de données au big data

e constructeur français de moteurs d'avions, filiale de Safran et alliée à General Electric, met en place un « digital lab », et a, d'ores et déjà, lancé le recrutement de data scientists. Bienvenue dans le monde du big data, puisque c'est la vocation de cette entité. L'entreprise n'est pas novice dans l'analyse de données. « J'ai été recruté, il y a sept ans, pour mettre au point des algorithmes afin d'exploiter les données issues des moteurs d'avions (photo) », indique Jérôme Lacaille,

mathématicien qui a mis en œuvre toute une batterie d'algorithmes pour le motoriste. Application visée: le « health monitoring », à savoir la surveillance de l'état de santé des moteurs afin de détecter au plus tôt l'usure inattendue ou prématurée d'une pièce. Cette application permet d'aller au-delà du programme de maintenance systématique. extrêmement strict, en évaluant en permanence l'état des moteurs,



à partir de données issus de capteurs. C'est également un service fourni aux clients. Aujourd'hui, avec le big data, Snecma veut aller beaucoup plus loin. Recueillir la masse de données disponibles dans toute l'entreprise - en particulier dans les bureaux d'études - et les exploiter plus systématiquement, en particulier comme aide à la conception. Jointe à la connaissance fine du comportement réel de ses moteurs avec l'acquisition de données en vol cela permettra d'en optimiser la conception. Jérôme Lacaille cite plusieurs exemples d'applications. D'abord, la possibilité de corréler le couple usage-usure des moteurs. Ou encore, celle de connaître très précisément les relations entre le dimensionnement des pièces et les performances. Enfin, avec le big data, il devient possible sur les bancs d'essais des moteurs, de profiter des masses considérables de données recueillies pour mettre en évidence, avec les algorithmes ad hoc, des phénomènes non anticipés. Le contraire de ce qui se pratique actuellement où, le recueil de données se fait dans un but précis, permettant avant tout au bureau d'études de valider des hypothèses.



découvrir des informations que les connaissances actuelles ne peuvent pas anticiper... Autre exemple, celui fourni par le Français Quinten, spécialisé dans la valorisation des données. Son client (qui ne tient pas à dévoiler son identité) faisait face à un sérieux problème : en quelques mois, deux importants dysfonctionnements étaient apparus sur ses machines tournantes. Impossible d'en découvrir la cause. C'est en analysant, à partir des données issues de six mois de fonctionnement (au rythme de dizaines de milliers de points de mesure

toutes les dix minutes) que le problème a été identifié. Il est apparu dans les deux cas qu'une certaine configuration très spéciale de température et de pression s'était produite juste avant la panne – et dans ces deux cas seulement. Si la cause exacte n'a pas été identifiée, l'entreprise dispose pour le moins d'informations lui permettant d'éviter que la panne ne se reproduise. Quinten, par ailleurs, travaille beaucoup avec l'industrie pharmaceutique

(lire page 48), notamment dans le cadre des analyses cliniques. Une industrie pour laquelle l'apport de traitement de données massives peut être colossal.

La nécessité de nouvelles compétences

Cela dit, attention! Avec ces exemples, il est tentant de déduire – et beaucoup ne font rien pour vous en dissuader – qu'il suffit de réunir des volumes colossaux de données, de secouer le tout avec un peu d'Hadoop, pour récolter des trésors. Autrement dit, le *big data* aiderait à faire de formidables découvertes à partir d'un magma totalement informe de données brutes. Erreur fatale! Le *big data* n'est pas « magique », comme insiste Bruno Teboul. C'est même une technologie difficile à maîtriser, voire très difficile. Au point que pour obtenir des résultats

66 Nous ne pouvons pas laisser d'autres sociétés exploiter les données générées par notre industrie

> Jeff Immelt PDG de General Electric

satisfaisants sont apparus les *data scientists*, espèce encore rare et fort convoitée. Pourquoi recourir à un *data scientist*? Parce que pour profiter des bienfaits du *big data*, il faut commencer par avoir une idée précise de ce que l'on cherche. Il faut ensuite collecter des quantités phénoménales de données, mais pas n'importe lesquelles: celles qui sont le plus susceptibles d'être utiles dans le contexte donné. Puis, il faut les « nettoyer », les organiser, les normaliser... Il faut enfin leur appliquer des algorithmes pertinents, c'est-à-dire à la fois adaptés aux caractéristiques des données à traiter, et aux résultats envisagés.

Pour ce faire, le *data scientist* est épaulé par des outils tels que le *machine learning*. Par auto-apprentissage, cette technique sert à déterminer le type d'algorithmes à mettre en œuvre pour extraire du sens de la masse de données et d'automatiser le process. Mais, en fin de compte, c'est toujours l'opérateur humain qui reste l'acteur principal du travail. Pour ce faire, il doit non seulement être un expert pointu en mathématiques et informatique avancée, mais aussi posséder une bonne connaissance du métier de l'entreprise pour laquelle il travaille, c'est-à-dire être au fait de ses besoins et de ses enjeux. C'est cet ensemble de qualités – et pas le seul *machine learning*, aussi évolué soit-il – qui permet de faire les choix les plus judicieux à tous les stades du processus.

Les *data scientists* sont donc des oiseaux (très) rares. Et comme on n'est jamais si bien servi que par soi-même, l'industrie participe activement à

leur formation. On a ainsi vu Thales, aux côtés d'Orange et de Keyrus, soutenir la chaire de l'Ecole polytechnique créée, cette année, pour former de tels spécialistes. De même, les entreprises s'investissent dans la recherche. Safran, PSA Peugeot-Citroën, aux côtés de Criteo et d'une grande banque française, soutiennent la chaire *Machine learning for big data* créée en septembre 2013 avec la Fon-

dation Télécom au sein de Télécom ParisTech. Le groupe Yves Rocher finance, lui, avec Voyages-SNCF et Deloitte, la chaire Big data & market insights à Télécom ParisTech également. Autant de preuves, de l'intérêt de l'industrie pour le sujet. S'il en fallait d'autres, il suffit de se rappeler l'investissement démesuré - plus d'un milliard de dollars - que General Electric* a consenti pour la R&D sur ce sujet, qu'il appelle « Internet industriel » (saisie et traitement de données issues de ses produits). Son PDG, Jeff Immelt, précise l'enjeu : « Nous ne pouvons pas nous permettre de laisser d'autres sociétés exploiter les données générées par notre industrie », explique-t-il. Il a tout compris. Et c'est ce que répète à l'envi François Bourdoncle, l'un des responsables du plan gouvernemental big data :

« Attention à ne pas vous laisser déposséder du trésor que représentent vos données! », dit-il en substance. Actuellement, le big data est sous le feu des projecteurs. Il a atteint le « peak of inflated expectations » de la fameuse courbe du Gartner Group qui décrit le cycle de « hype », de toute nouvelle technologie. La surexcitation autour du sujet est à son apogée. Chacun sait toutefois qu'après ce sommet vient inévitablement l'étape suivante : le « gouffre de la désillusion ». Le moment où les promesses non tenues et la complexité de la mise en œuvre concrète suscitent doute et déception. Mais on peut prédire sans trop de risque que le potentiel de la technologie est tel qu'il ne se passera guère de temps avant que le big data ne fasse la preuve de sa formidable capacité de transformation et d'innovation. Le temps que les utilisateurs, dans l'industrie comme ailleurs, découvrent ses plus fructueux domaines d'application et que des « killer apps », encore inimaginables, voient le jour. C'est bien parti, même si les grands groupes sont souvent handicapés par la lourdeur des processus de décision. Quoi qu'il en soit, de fantastiques surprises nous attendent.

*Lire le rapport : « Industrial Internet : Pushing the Boundaries of Minds and Machines », de Peter C. Evans et Marco Annunziata http://invent.ge/1snGZow

PSA Peugeot-Citroën

Une vaste **palette d'applications** en vue

epuis peu, la plate-forme big data de PSA Peugeot-Citroën, conçue avec IBM, est pleinement opérationnelle. C'est le grand départ pour cette technologie avec laquelle, néanmoins. le constructeur a déià commencé à travailler. « Les données structurées représentent quelque 10 % de toutes celles qui sont disponibles dans l'entreprise; le big data va nous aider à exploiter les 90 % restants », explique Jean-Pierre Dumoulin, chief technical officer (CTO) de PSA Peugeot-Citroën et maître expert big data. Des données, il en existe en effet des myriades. Celles recueillies automatiquement par tous les équipements informatiques du groupe. Des informations présentes dans ses bases de données. Des données externes (type réseaux sociaux) et, bientôt, de plus en plus d'informations directement issues de la voiture connectée. Toutes ont vocation à être exploitées. Déjà le groupe a développé une première application, liée à la sécurité, en analysant tous les logs issus des systèmes numériques. Pour lutter contre la fraude, par exemple. Avec les données internes, celles de son réseau, il travaille également à une application permettant de prédire les volumes des ventes par pays et par modèle. Avec celles du Web, il a commencé à analyser ce qui se dit sur les réseaux sociaux à propos de ses véhicules pour mieux cibler ses campagnes marketing. « C'est une méthode qui vient s'ajouter, et non remplacer, les classiques enquêtes marketing », précise Jean-Pierre Dumoulin. Enfin, le constructeur envisage une multitude d'applications liées à la récupération des



Les technologies de la Citroën Cactus proposent notamment un écran tactile de 8 pouces au centre du tableau de bord.

données des véhicules. D'une part, comme le fait l'aéronautique, pour améliorer la conception des voitures, à partir de l'analyse de la masse d'information sur leur comportement réel, obtenue sur un vaste parc. D'autre part, pour améliorer le service client en fournissant à l'utilisateur des informations de maintenance, par exemple, correspondant à son usage exact. Ici, le constructeur a une longueur d'avance. Il a équipé de longue date toutes ses voitures déjà plus d'un million - d'une communication via GSM, initialement conçue pour l'appel d'urgence. « D'ici à trois ans, nous aurons quelque 4 millions de voitures équipées qui pourront chacune transmettre des informations très variées », dit Jean-Pierre Dumoulin. Un vrai filon. ■



A l'instar de nombreux secteurs, l'industrie pharmaceutique est en chasse perpétuelle contre les coûts. Dans la R&D, comme dans la logistique, le big data bouleverse les pratiques face à la complexité grandissante de leurs données..

GUILLAUME MOLLARET, AVEC ALAIN CLAPAUD

ous les industriels de la pharmacie n'en sont pas au même point du développement, ni même de la réflexion quant à la façon d'intégrer le big data dans le processus stratégique de décision. Tour d'horizon dans un secteur avare de confidences... Ainsi, pour réfléchir un cran plus loin sur le sujet, Orange et des partenaires viennent de lancer un think tank dédié, le Healthcare Data Institute. « Son objectif est de réaliser des études, de partager des expériences concrètes mais aussi d'être un lieu d'échanges et de réflexions », avance sa viceprésidente, Isabelle Hilali. Parmi les membres du conseil d'administration, on trouve Sanofi, Vitalia (le numéro 3 français de l'hospitalisation privée), IBM, McKinsey, le cabinet d'avocats spécialisé Desmarais et, bientôt, un assureur. « Beaucoup de laboratoires demandent à adhérer. Certains ont déjà une certaine maturité. A ce titre, les échanges entre grands groupes et start-up promettent d'être intéressants », précise Isabelle Hilali. Pour susciter ces discussions, une conférence-débat, réservée aux membres, est ainsi organisée toutes les six semaines. D'un point de vue purement industriel, la pharma est notamment confrontée à la surveillance de ses approvisionnements. « Quand on fabrique des médicaments sur différents sites avec plusieurs sources d'ap-



l'interview de **Pierre-Henry** Longeray, président Merck Serono France, division pharmaceutique du groupe allemand Merck KgaA, sur **ALLIANCY.FR**

provisionnement, il est essentiel pour nous d'identifier au plus vite d'où vient le potentiel défaut de qualité, explique, sous couvert d'anonymat, le représentant d'un des plus importants groupes du secteur. Le big data peut nous aider à résoudre ce genre de problème. Il implique cependant une remise en question de notre système actuel. C'est un choix lourd qui n'est pas sans conséquence sur nos organisations. C'est pour cela que nous n'avons, à ce jour, rien acté sur la stratégie à opérer. » C'est ainsi que le processus de décision dans le basculement vers une technologie big data met parfois près de trois ans avant d'aboutir à la signature d'un contrat. « Sur ce secteur, nous sommes d'une lenteur qui ne serait, je crois, pas tolérée dans un autre service », poursuit notre interlocuteur.

Démultiplier les puissances de calcul

Comme la R&D engloutit des dizaines - voir des centaines – de millions d'euros dans le développement de molécules, qui ne verront peut-être jamais le jour sous la forme de cachets ou d'injectables, évaluer au plus vite, lors des phases cliniques, si la molécule sur laquelle les chercheurs travaillent, aboutira à la création d'un médicament, relève d'une nécessité économique cruciale. Chez GSK, le numéro 5 mondial, les outils mis en place dans le Data Lab

« permettent l'exploration rapide des hypothèses d'assem-blage de molécules avec une assistance informatique minimale », plaide Brad Donovan, responsable de l'innovation analytique chez GSK. L'analyste réitère ainsi quelques exemples parlants, déjà présentés l'an dernier dans Alliancy, le mag (lire n° 6) : des cycles d'analyses de 130 heures sont ainsi réduits à 5 heures seulement. Pour certains processus de calculs allant de six à dix mois, GSK se pose à présent un objectif à deux mois et demi. Une économie de temps par ailleurs synonyme de moyens...

L'efficacité des médicaments en question

Une fois en vente, se pose la question de l'efficacité du médicament. « Le big data peut clairement aider à justifier le prix d'un traitement* », soutient Joy King, consultant pharma et sciences de la vie chez Teradata. Contrairement à la France, où le patient a le libre choix de sa pharmacie pour acheter ses médicaments ; aux Etats-Unis, les médecins ont la capacité grâce au « e-prescribing » de diriger directement le patient vers une officine. Ces données, anonymisées, sont vendues aux laboratoires pharmaceutiques, qui grâce au big data, peuvent analyser les prescriptions des médecins saisies par cette voie informatique.

« Cela permet de recouper les associations de prescriptions, mais également de plaider sa cause auprès des assureurs qui refusent de rembourser un médicament qu'ils estiment trop cher », affirme Joy King, citant le cas d'un client préférant conserver l'anonymat. Il s'agissait d'un médicament injectable. Toute la finesse du laboratoire a été de démontrer que si le traitement était effectivement plus cher à l'unité, son efficience se mesurait sur un temps plus court que celui de ses concurrents... Les assureurs ont donc été amenés à réviser leur jugement puisqu'au final, le traitement s'est révélé moins cher à rembourser. »

En France, Quinten, qui préfère parler de data science que de big data travaille en collaboration avec les laboratoires Roche et les Hôpitaux de Paris sur

l'analyse de données visant à améliorer la recherche sur le cancer colorectal. « Lors d'une précédente étude réalisée sur 121 personnes atteintes d'un cancer du poumon et traitées à l'institut Gustave-Roussy à Villejuif, nous avons pu démontrer grâce à une analyse plus fine qu'il y avait une différence d'efficacité du traitement entre les personnes traitées par chimiothérapie en adjuvant suivant le génome de la tumeur. Or, avec les analyses conventionnelles, aucune différence n'était mise à jour », explique Alexandre Templier, directeur général de la PME francilienne.

Autre possibilité offerte par le big data, la prévention des effets secondaires sur certaines populations, puisque les essais cliniques portent rarement sur une cohorte représentative de la population. HealthCore, un organisme américain de recherche spécialisée pharmaco-épidémiologie, effectue ainsi des tests de données sur certaines catégories - notamment sur critères ethniques, ce qui est possible aux États-Unis, mais impossible en France - après mises sur le marché des médicaments. Le but : surveiller si la prescription d'un ou plusieurs médicaments n'entraîne pas d'effets secondaires indésirables auprès de certaines populations. Rose Cintron-Allen, consultante industrie santé de Teradata, prétend ainsi que si ce type de calcul avait été possible voilà une dizaine d'années, « on aurait pu se rendre compte indépendamment, et beaucoup plus tôt, que l'anti-inflammatoire Vioxx pouvait provoquer des infarctus du myocarde ». Côté rapidité, Joy King ajoute qu'une requête croisée sur l'ensemble de la base de « e-prescribing » pour deux médicaments permet une réponse dans un laps de temps de « trois minutes contre 45 heures auparavant. »

*En France, le prix d'un médicament est le résultat d'une négociation entre l'entreprise exploitant ledit médicament et le comité économique des produits de santé (CEPS), un organisme interministériel. Les prix des médicaments non remboursables sont déterminés librement par les pharmaciens d'officine.

Des supercalculateurs pour créer de nouvelles molécules

En 2011, à l'occasion des premiers essais de Curie, le plus puissant supercalculateur en France, une simulation numérique est exécutée sur les 92000 cœurs de processeurs dans le cadre de la lutte contre la maladie d'Alzheimer. Michel Caffarel, chercheur du CNRS et de l'université Paul Sabatier de Toulouse a pu simuler, au niveau atomique, le rôle du cuivre dans la dégénérescence des neurones. Un calcul qui illustre l'avancée de la recherche médicale dans le domaine de la simulation numérique. La puissance des supercalculateurs permet d'accélérer

la recherche des laboratoires pharmaceutiques avec des tests virtuels qui servent à sélectionner les molécules avant de passer aux tests in vitro et in vivo. La simulation multi-échelles recrée sur ordinateur le comportement d'une molécule sur des cellules au niveau de ses atomes et de leurs électrons, de ses peptides, protéines et supramolécules jusqu'au niveau mésoscopique et macroscopique Une approche qui demande une puissance de calcul colossale. Autre piste extrêmement prometteuse

pour l'avenir, le big data. L'idée est d'analyser toutes les données du patient pour établir un traitement totalement personnalisé, donc plus efficace et avec moins d'effets secondaires. Les traitements de données porteront sur ses analyses en laboratoire, le séquençage de son ADN, mais aussi des informations collectées par les objets connectés, voire même sur les réseaux sociaux. C'est l'approche 4P (personnalisée / prédictive / préventive / participative), fondement de la médecine du futur selon les chercheurs. ■ A. C.



- SÉCURITÉ ÉLECTRONIQUE
- CYBERSÉCURITÉ
- SÉCURITÉ INTÉRIEURE DES ETATS, **DES GRANDS ÉVÈNEMENTS** ET DES FRONTIÈRES
- INGÉNIERIE DE SÉCURITÉ
- SÉCURITÉ MÉCANIQUE ET PHYSIQUE
- SURVEILLANCE HUMAINE
- SERVICES

RÉSERVEZ VOTRE BADGE

dès le 15 janvier avec le code Alliancy sur www.accessecurity.fr

LES PARTENAIRES DU SALON

















































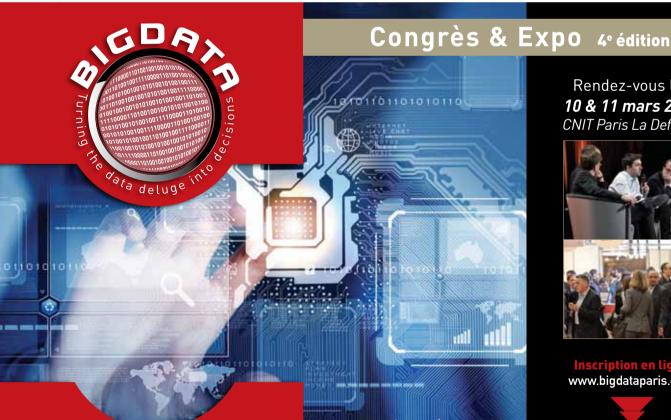












Retrouvez les leaders du secteur sur l'événement majeur du Big Data en France

- 2 jours conférence / exposition
- 150 intervenants experts français et internationaux
- 120 exposants

Rendez-vous les 10 & 11 mars 2015 CNIT Paris La Defense



Inscription en ligne www.bigdataparis.com



www.bigdataparis.com by Color. • 5000 professionnels du Big Data

UNE INFORMATION VITALE.

Pour Franck Von Lennep, la France a la meilleure base de données de santé au monde pour étudier le lien entre prise de médicament et hospitalisation. Explications.

PROPOS RECUEILLIS PAR GUILLAUME MOLLARET

Alliancy, le mag. Quels sont les principaux blocages à l'open data sur les données de santé?

Franck Von Lennep. Tout d'abord, il faut définir de quoi parle-t-on. Il y a, aujourd'hui, un tas de données qui ne sont pas des données sensibles. Ce sont des données de santé qui ne contiennent aucune information sur des personnes que l'on serait susceptible de reconnaître. Pour ces données-là, il n'y a aucun blocage. Il y a déjà beaucoup de données en open data sur ameli.fr, ou sur le site de l'Agence du médicament par exemple. Les données sensibles sont celles qui présentent un risque de réidentification. Il s'agit d'abord des deux grandes bases de données, Sniiram et PMSI, qui sont les données de remboursement de l'Assurance maladie et de séjours hospitaliers. Or, même si les données qu'elles contiennent ne sont pas nominatives - il n'y a, par exemple, pas le numéro de sécurité sociale du patient -, il y a tout de même beaucoup d'autres informations. A ce titre, ces données présentent un risque de réidentification. On est donc là dans le champ de la Loi informatique et libertés.

Quelles sont les limites obligatoires?

Ce que nous avons prôné dans la commission Open data, c'est que davantage d'acteurs, dans des conditions plus simples qu'aujourd'hui, puissent avoir accès à ces données. Il s'agit, en effet, d'une mine d'informations pour faire de la recherche, de la santé publique, apporter de l'information aux Français et in fine améliorer leur état de santé. Comme il s'agit de données sensibles, il faudra encadrer leur accès et le limiter, notamment aux organismes du secteur public du type Agence du médicament, Assurance maladie; aux chercheurs de l'Inserm, de CHU, du CNRS, mais aussi à certains acteurs privés. A condition que leur étude présente un intérêt général en matière de santé publique et de prise en charge sanitaire, et sous réserve naturellement d'autorisation de la Cnil.

Comment se fera l'ouverture de ces données?

Il faudra un guichet unique. Que le demandeur sache où il doit aller, qu'il y dépose son dossier. Ensuite, à l'intérieur de ce guichet, ce sera aux professionnels de les instruire, de s'organiser entre un comité d'experts et une organisation associant des membres de la société civile. Une fois le dossier instruit, il arrivera à la Cnil et, s'il est accepté, la mise en œuvre de l'accès aux données sera du ressort de l'Assurance maladie. Cela relèvera de son métier. Un nouveau métier finalement. Dans tous les cas, même si l'organisation peut paraître compliquée, il faut que les choses soient très simples pour le demandeur.

Quand saura-t-on si vos préconisations seront suivies d'effets positifs?

Il faut attendre l'adoption de la Loi Santé qui sera examinée en première lecture à l'Assemblée nationale au premier semestre 2015. C'est l'article 47 qui porte sur l'open data. On peut espérer une mise en œuvre de ses mesures début 2016.



« Il faut apporter les garanties que ces données ne seront pas ouvertes à tout vent » Franck Von Lennep

Que voyez-vous d'intéressant à l'étranger? Et à l'inverse, les erreurs à ne pas reproduire?

De nombreux pays nordiques ont développé des indicateurs au niveau des soins et de la qualité de l'offre de soins. Ce sont des indicateurs importants pour la démocratie sanitaire, qui aident le citoyen à faire son choix dans le système de santé. En revanche, il faut faire attention à ne pas créer des peurs. Les Anglais communiquent énormément sur l'ouverture de leurs données de santé. Récemment, le projet de création d'une base de données, pourtant pas du tout en open data, simplement ouverte à des chercheurs, a soulevé une inquiétude de la part d'associations de patients. Il faut donc apporter les garanties que ces données ne seront pas ouvertes à tout vent. Un règlement européen devrait être adopté en 2015 à ce sujet.

La France n'a qu'un seul opérateur, la Cnam, ceci constitue-t-il un risque supplémentaire?

Je parlerai plutôt d'une chance! Nous avons la plus grande base de données structurée au monde comportant des données à la fois sur la ville, les médicaments et les séjours hospitaliers. C'est la meilleure base au monde pour étudier le lien entre prise de médicament et hospitalisation... permettant, par exemple, de faire le rapprochement avec d'éventuels effets indésirables.

*Directeur de la recherche, des études, de l'évaluation et des statistiques (Drees).



BATAILLE SUR LES

DONNÉES DE SANTÉ.

Dossiers patients, imagerie médicale et bientôt toutes données générées par les objets de santé connectés, les besoins de stockage font exploser dans les années à venir dans le secteur de la santé. Une aubaine pour les acteurs du cloud.

ALAIN CLAPAUD

vec 38 établissements publics et privés, Région sans film est l'un des projets phare de la dématérialisation des données de santé en Ile-de-France. A terme, 2 millions d'examens médicaux, pour l'essentiel de très volumineux fichiers d'imagerie médicale, seront stockés dans ce service cloud. Orange Healthcare et General Electric (GE) assurent l'hébergement des 64 téraoctets de données. Un exemple des changements qui vont pousser les établissements de santé et les professionnels du secteur à externaliser de plus en plus de données de leurs patients.

Si le volume des données de santé va exploser, les pratiques d'externalisation de ces données patients, très sensibles par nature, sont très encadrées en France. Chaque fournisseur d'applications cloud, chaque datacenter doit décrocher un agrément pour avoir le droit de proposer ses services. L'Agence des systèmes d'information partagés de santé (Asip Santé) a accordé un peu moins de 80 autorisations à ce jour, pour 212 dossiers déposés. 40 sont en cours de traitement, sachant qu'il faut environ huit mois pour décrocher le précieux sésame.

certificat médical
accident du travail
accident du travail
accident du travail
accident du travail

initial de prolongation (*)
de travail
feuille de soins - médecin

Art. R. 161-40 et suivants du Code de la sécurité sociale

Prescription médicale de transport - Notice

Assurance maladie, à l'aide de l'enveloppe M.

date

Prescription médicale de transport - Notice

Assurance maladie

Assurance maladie de transport - Notice

Assurance maladie
DEMANDE D'Acce

Chaque cas est unique, mais il faut faire preuve d'une infrastructure de sécurité à l'état de l'art pour avoir le droit d'héberger des données patients, par nature confidentielle. Parmi les entreprises à avoir décroché l'agrément des opérateurs de télécoms BT, Orange Business en tête, des opérateurs de datacenters comme Interxion, Prosodie Capgemini, Numergy, Cloudwatt, des CHU et groupement sanitaires ou encore des fournisseurs d'applications pour les professionnels de santé comme Pharmagest, Cerner, Docapost BPO, Choregie, etc.

La montée du cloud est inéluctable

Wooxo, un spécialiste du stockage de données sécurisé dans ces boîtiers est en train de faire agréer ces deux datacenters provençaux de Marseille et La Ciotat afin d'orienter son offre vers un stockage hybride local et hébergé. Netplus, qui se pose comme un opérateur de cloud santé, a déjà obtenu son agrément. Le Français propose tout une série de services d'hébergement à destination des professionnels de santé et éditeurs de logiciels du secteur. Netplus s'est récemment allié à Cloudwatt pour accroître ses capacités d'hébergement en vue de l'explosion du marché attendue. « Nous nous sommes alliés à l'un des clouds souverains, car nous pensons que le marché va s'orienter vers une forte externalisation des données de santé. La vocation des hôpitaux n'est pas de faire du stockage et de la gestion de données. Une partie de cette gestion va être externalisée, et cela va arriver relativement vite car le taux d'informatisation des hôpitaux est, aujourd'hui, plutôt élevé et la production de données est en train d'exploser. Or les hôpitaux, l'Etat à moins d'argent à y investir, le modèle Saas avec un paiement à l'usage est une réponse aux barrières économiques. »

Editeur initialement créé par des pharmaciens pour développer des logiciels de gestion d'officines, Pharmagest dispose, aujourd'hui, de plusieurs agréments pour l'hébergement de données de santé, notamment une autorisation pour que les patients puissent consulter leur dossier médical sur le Web. Une révolution dans le secteur médical.

Dominique Pautrat, directeur général de Pharmagest, explique ce qui l'a poussé à investir lourdement dans le cloud : « Dans un premier temps, nous avons proposé nos applications de gestion en local, puis nous sommes devenus fournisseurs d'accès Internet pour connecter les officines, puis opérateur de télécoms. Dès 2008-2009, nos réflexions nous ont conduit à envisager un stockage des données de santé externalisé. Nous avons commencé par stocker des données de gestion pour les officines et, en 2010, nous avions constaté que nos applications en mode hébergé connaissaient le succès. Nous avons alors décidé de construire notre datacenter. » Editeur de logiciel, fournisseur d'accès Internet, opérateur de télécoms, et aujourd'hui opérateur de cloud, Pharmagest aime maîtriser l'ensemble des technologies. « Cela représente énormément d'investissement car la courbe d'apprentissage est longue, notamment au début, mais cela fait partie des gènes de Pharmagest!»

L'écosystème cloud se met donc en place pour accompagner les établissements de santé et les professionnels à basculer au moins en partie leur informatique dans le cloud. L'essor de l'imagerie médicale au format numérique, mais aussi la nécessaire démocratisation du séquençage ADN dans les diagnostics, vont sans nul doute précipiter ce mouvement inéluctable des données vers le cloud.

L'avis de...

Florence Eon

Directrice juridique de l'Agence des systèmes d'information partagés de santé (Asip Santé)



Concernant le processus d'agrément des hébergeurs de données de santé, chaque décision fait l'objet d'une instruction sur la base d'un dossier qui doit décrire très précisément le volet technique, mais aussi économique, juridique et éthique du service. On ne demande pas les mêmes

choses pour l'agrément d'une salle blanche ou à une application liée au dossier patient informatisé... L'article de loi 118 a limité le champ d'application de l'agrément aux données de santé qui ont été collectées et recueillies à l'occasion d'une activité de diagnostic, de prévention ou de soin. Un bracelet pour compter le nombre de pas, sans finalité médicale, ne tombe pas sous la coupe de cet arsenal juridique, mais un tensiomètre connecté, dont les données sont transmises à un professionnel de santé, on peut s'interroger.

se former
rencontrer
échanger
comprendre
anticiper
découvrir

« la modernisation des équipements et des services touristiques est nécessaire pour répondre au plus près aux attentes des clientèles »

> Raphaël Chauvois, Vice-Président à l'économie et au tourisme, Conseil Régional de Basse-Normandie

« Le marketing en temps réel, lié au big data réinvente nos façons de travailler »

Joël Gayet, Chercheur associé Sciences-Po Aix, Directeur de la Chaire « Attractivité et Nouveau Marketing Territorial »



TOURISME3 numerique3 23 - 24 MARS 2015















ros (2 milliards de dollars) en 2014, avec une croissance annuelle approchant les 30 %, qui s'est structuré à toute vitesse... Assurer le patrimoine informationnel de l'entreprise devient une réalité pour bon nombre de sociétés. Et les professionnels du secteur sont au rendez-vous. « En 2013, le sujet était encore en discussion. En 2015, la question est plutôt : qui n'est pas encore positionné sur ce segment chez les assureurs ? », relève Laurent Heslault, directeur des stratégies de sécurité chez l'éditeur américain Symantec.

Fin 2014, c'est, par exemple, Allianz Global Corporate & Specialty (AGCS) qui a doublé la capacité de sa police d'assurance cyber data protect, avec une couverture pouvant atteindre 100 millions d'euros. Lancée durant l'été 2013, cette offre avait à l'origine été prévue avec une capacité de 25 millions, doublée une première fois la même année. « Les assureurs ont évidemment poussé le sujet cyber, mais cette progression vient aussi de très fortes attentes. Le marché français est assez mature, notamment grâce au travail de l'Anssi* ou à la récente LPM** », souligne Iouri Goloubtzoff, responsable produit Cyber chez AGCS.

Se poser les bonnes questions

De fait, le travail de valorisation et de protection de leur système d'information (SI) par le biais d'un contrat d'assurance est le corollaire naturel de la prise de conscience grandissante des professionnels en matière de cybersécurité. Le tout tient à la gestion du risque : mettre

en place des solutions de sécurité est un moyen de réduire ces risques, là où « assurer son SI » revient à les transférer. Dans un cas comme dans l'autre, les entreprises doivent, dès le départ, pouvoir se poser les bonnes questions : à quelles menaces suis-je confronté ? Quels investissements consacrer pour rendre le problème moins épineux ? Concrètement, comment faire ?

Or, aucune réponse juste tant que l'entreprise ne peut définir avec précision la valeur du patrimoine informationnel qu'elle veut défendre. Valeur qui sous-tendra le moindre investissement. « En matière d'assurance, les entreprises avaient l'habitude de protéger le SI matériel, c'est-à-dire le bris de machine et autres problématiques de hardware. Aujourd'hui, le SI est partout, dans toutes les business units, et n'est plus la préoccupation exclusive du département IT. Ce n'est pas un sujet unique, mais une affaire de valeur transversale pour toute l'entreprise », explique Iouri Goloubtzoff. Surtout, cette valeur réside de plus en plus dans les données et de moins en moins dans le matériel luimême. Big data et cloud sont passés par là...

Pour calculer la valeur des données, « il faut rapprocher la réflexion de celle sur l'ensemble du patrimoine immatériel de l'entreprise », conseille Vanessa Bouchara, avocate spécialisée en droit de la propriété intellectuelle. « Au même titre que la marque, le droit

> de création, d'auteur ou les secrets de fabrication, ce patrimoine inclut, aujourd'hui, les données de plus en plus riches et nombreuses, du client jusqu'à la R&D. » Si d'un point de vue légal, rares sont les règles à avoir changé en profondeur sur la dernière

décennie, les entreprises deviennent-elles de plus en plus conscientes des enjeux. Quand bien même les projets complètement aboutis de *big data* restent l'apanage de seulement quelques entreprises en France, le discours est, lui, intégré depuis plusieurs années. La richesse des données, de leur analyse et leur exploitation, est une clé non seulement pour l'activité quotidienne – de la relation client à

L'exemple des Gafa est sur toutes les lèvres



patrimoine informationnel plan large

donnée impacte tout le monde, les réponses ne se trouveront pas (seulement) à la DSI. « Toute une génération de managers voit le SI comme la dernière roue du carrosse, cela doit absolument évoluer », épingle Sabine Bohnké, fondatrice du cabinet de conseil Sapientis et spécialiste de la modernisation des SI (lire ci-contre). « Les métiers ont trop souvent perdu la main sur le sens de la donnée. Or, les SI sont aussi des flux d'information dont chacun doit avoir une utilité précise et connue, qui permette d'en déterminer la valeur ».

Cette connaissance transversale, vitale, a bien sûr une dimension technique, mais elle est surtout le fruit d'un parti pris : « Les entreprises, qui ont une vraie vision de la valeur de leur SI et de leur patrimoine informationnel, sont celles capables d'entremêler le sujet avec leur stratégie métier », explique la directrice de Sapientis. Des secteurs, comme celui de la banque, ont abordé ces approches depuis des années puisque leur activité a très tôt été entièrement liée à leur système d'information. Un destin que connaissent, aujourd'hui, toutes les entreprises, toutes activités confondues. Reste à savoir si l'évolution des mentalités pourra se faire assez rapidement pour que ce patrimoine aux contours souvent mal défini ne soit pas pillé.

* Anssi : Agence nationale de la sécurité des systèmes d'information.

**LPM: La loi de Programmation militaire du 18 décembre 2013 inclut plusieurs articles relatifs à la cybersécurité et à la responsabilité des entreprises vis-à-vis des problèmes de leur SI.

la logistique –, mais aussi un levier d'innovation formidable. L'exemple des Gafa est sur toutes les lèvres (voir l'étude de Fabernovel : « GAFAnomics, new economy, new rules », accessible sur http://bit.ly/1HfsRzo).

Reste à traduire cette conception en actes. « La première problématique est d'avoir une connaissance saine et complète de son patrimoine », poursuit Vanessa Bouchara. La démarche est essentielle pour orchestrer sa cybersécurité, mais également dans une logique assurantielle. « L'assureur doit avoir une idée très précise du fonctionnement du SI, de son niveau de protection et de l'impact sur le business. C'est un sujet qui devient vite très technique et complexe. Chaque SI est unique », précise Iouri Goloubtzoff.

Savoir chiffrer l'usage

Si la connaissance et la description technique de son SI est rendu facilement possible grâce à de l'IT Services Management (ITSM) ou, encore, aux nombreuses solutions qui permettent de surveiller son réseau par exemple, la valeur à accorder à la donnée reste souvent confuse. « Le problème est qu'une donnée seule n'a pas de réelle valeur en soi, c'est l'exploitation qu'on en fait, ce que l'on construit avec, dans une optique business qui importe. Et chiffrer une valeur d'usage est complexe », reconnaît Xavier Cangardel, associé au sein du cabinet Colombus Consulting.

Pour y arriver, l'entreprise doit déjà répondre à un enjeu culturel et organisationnel : puisque le SI est maintenant partout et que l'usage de la

L'avis de... Sabine Bohnké

fondatrice du cabinet de conseil Sapientis, dédié aux systèmes d'information (SI)

Les entreprises, petites et grandes, ne sont majoritairement pas capables en 2015 d'évaluer précisément la valeur de

leur SI. Les PME souffrent d'un manque de temps à consacrer au sujet, trop prises par leurs processus métiers. Le point de départ pour y parvenir serait d'éviter l'entreprise à deux vitesses, avec des décisions stratégiques prises en haut, et l'espoir que le SI s'alignera ensuite de lui-même... La PME doit adhérer à une démarche itérative. Comprendre ce que fait son SI sur un petit périmètre d'abord, puis élargir. Quelles sont les activités impactées ? Avec quelles parties prenantes ? Quels liens avec la stratégie globale ? Refuser de rapprocher systématiquement les aspects techniques du SI et des flux de données, de son quotidien business, c'est admettre que l'entreprise est une boîte noire. Un dirigeant ne doit pas attendre de solutions clé en main pour reprendre le contrôle. L'analyse de l'usage et de la valeur du SI pour les métiers doit être permanente. Etre autonome sur le sujet est le seul moyen pérenne pour éviter que le problème ne grandisse avec la croissance de l'entreprise.



Retrouvez l'intégralité de l'entretien avec Sabine Bohnké sur ALLIANCY.FR